2015

Remote Sensing for Monitoring the Mountaintop Mining Landscape: Applications for Land Cover Mapping at the Individual Mine Complex Scale

Aaron E. Maxwell

Remote Sensing for Monitoring the Mountaintop Mining Landscape: Applications for Land Cover Mapping at the Individual Mine Complex Scale

Aaron E. Maxwell

Dissertation submitted to the Eberly College of Arts and Sciences
at West Virginia University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Geology

Timothy A. Warner, Ph.D., Chair
Jamison Conley, Ph.D.
J. Steven Kite, Ph.D.
Michael P. Strager, Ph.D.
Nicholas P. Zégre, Ph.D.

Morgantown, West Virginia
2015

Keywords: Remote sensing, land cover mapping, mountaintop mining, machine learning, LiDAR, object-based image analysis, GEOBIA

ABSTRACT

Monitoring the Mountaintop Mining Landscape: Remote Sensing of Land Cover at the Individual Mine Complex Scale

Aaron E. Maxwell

The aim of this dissertation was to investigate the potential for mapping land cover associated with mountaintop mining in Southern West Virginia using high spatial resolution aerial- and satellite-based multispectral imagery, as well as light detection and ranging (LiDAR) elevation data and terrain derivatives. The following research themes were explored: comparing aerial- and satellite-based imagery, combining data sets of multiple dates and types, incorporating measures of texture, using nonparametric, machine learning classification algorithms, and employing a geographical object-based image analysis (GEOBIA) framework. This research is presented as four interrelated manuscripts.

In a comparison of aerial National Agriculture Imagery Program (NAIP) orthophotography and satellite-based RapidEye data, the aerial imagery was found to provide statistically less accurate classifications of land cover. These lower accuracies are most likely due to inconsistent viewing geometry and radiometric normalization associated with the aerial imagery. Nevertheless, NAIP orthophotography has many characteristics that make it useful for surface mine mapping and monitoring, including its availability for multiple years, a general lack of cloud cover, contiguous coverage of large areas, ease of availability, and low cost. The lower accuracies of the NAIP classifications were somewhat remediated by decreasing the spatial resolution and reducing the number of classes mapped.

Combining LiDAR with multispectral imagery statistically improved the classification of mining and mine reclamation land cover in comparison to only using multispectral data for both pixel-based and GEOBIA classification. This suggests that the reduced spectral resolution of high spatial resolution data can be combated by incorporating data from another sensor.

Generally, the support vector machines (SVM) algorithm provided higher classification accuracies in comparison to random forests (RF) and boosted classification and regression trees (CART) for both pixel-based and GEOBIA classification. It also outperformed *k*-nearest neighbor, the algorithm commonly used for GEOBIA classification. However, optimizing user-defined parameters for the SVM algorithm tends to be more complex in comparison to the other algorithms. In particular, RF has fewer parameters, and the program seems robust regarding the parameter settings. RF also offers measures to assess model performance, such as estimates of variable importance and overall accuracy.

Textural measures were found to be of marginal value for pixel-based classification. For GEOBIA, neither measures of texture nor object-specific geometry improved the classification accuracy. Notably, the incorporation of additional information from LiDAR provided a greater improvement in classification accuracy then deriving complex textural and geometric measures.

Pre- and post-mining terrain data classified using GEOBIA and machine learning algorithms resulted in significantly more accurate differentiation of mine-reclaimed and non-mining grasslands than was possible with spectral data. The combination of pre- and post-mining terrain data or just pre-mining data generally outperformed post-mining data. Elevation change

data were shown to be of particular value, as were terrain shape parameters. GEOBIA was a valuable tool for combining data collected using different sensors and gridded at variable cell sizes, and machine learning algorithms were particularly useful for incorporating the ancillary data derived from the digital elevation models (DEMs), since these most likely would not have met the basic assumptions of multivariate normality required for parametric classifiers.

Collectively, this research suggests that high spatial resolution remotely sensed data are valuable for mapping and monitoring surface mining and mine reclamation, especially when elevation and spectral data are combined. Machine learning algorithms and GEOBIA are useful for integrating such diverse data.

TABLE OF CONTENTS

CHAPTER 1

Introduction

1. Motivation and Aim

As the human population has increased in size, and developed tools to alter and shape the landscape on an industrial scale, the role of humans as agents of landscape change is becoming increasingly important. Anthropogenic geomorphic alteration, land use/land cover change, agricultural conversion, urbanization, increases in atmospheric CO_2, and biodiversity reduction have prompted many to argue for the addition of a third epoch to the Quaternary Period, the Anthropocene. The concept of the Anthropocene is based on the suggestion that human activities now rival the forces of nature and are key processes in producing a less biodiverse, less forested, and warmer Earth (Meybeck, 2003; Steffen et al., 2007; Zalasiewicz et al., 2010; Zalasiewicz et al., 2011; Steffen et al. , 2011). Although the concept of an Anthropocene Epoch is currently in debate, the impacts of humans on the landscape should be of upmost concern to the geologist and the landscape scientist. This dissertation therefore addresses the need to monitor and map human-induced landscape change in an accurate and efficient manner.

The overarching aim of this dissertation is to investigate high spatial resolution remotely sensed data (aerial- and satellite-based multispectral imagery, light detection and and ranging (LiDAR), and terrain derivatives) and advanced classification methods (machine learning algorithms and geographical object-based image analysis (GEOBIA) classification) for mapping mine disturbance. It addresses basic research questions that are currently of interest within the fields of remote sensing, land cover mapping, and image analysis. The following research themes were explored: comparing aerial- and satellite-based high spatial

resolution imagery, combining multiple data sets to increase classification accuracy, and increasing classification accuracy by using nonparametric, machine learning algorithms, measures of texture, and object-based classification. Table 1, at the end of this chapter, provides a list of common acronyms used in this dissertation.

This research investigates fundamental and timely questions applied to the problem of mapping land cover in a terrain that is complex as a result of steep topography and spectrally similar pre- and post-mining classes. The applied research component of this work addresses mapping needs within this landscape and explores the utility of techniques and available data for operational mine monitoring. Accurate maps of surface mine disturbance and surface mine reclamation are essential for understanding the long-term impacts of mining (Townsend et al., 2009). For example, land cover data have been shown to be essential in quantifying hydrologic impacts (Negley and Eshleman, 2006; Zégre et al., 2013), modeling water quality impacts (Merriam et al., 2013), and modeling terrestrial habitat impacts (Wickham et al., 2007) of surface mining.

3. Mapping Mountaintop Mining

Surface coal mining, and in particular mountaintop mining, is the dominant agent controlling both land cover change and geomorphic alteration in the southern coalfields of the eastern United States (Hooke, 1999; Saylor, 2008; Townsend et al., 2009; Drummond and Loveland, 2010). Mountaintop coal mining causes more material to be moved, and faster landscape alterations, than more traditional surface coal mining techniques such as auger, contour, and highwall mining (Fritz et al., 2010). The mountaintop mining process also results in the clearing of forests, the removal of top soil, and recontouring of the landscape (Palmer et al., 2010; Bernhardt and Palmer, 2011). It has been estimated that surface mining in the Appalachian

region has resulted in a net loss of 420,000 ha of forest between 1973 and 2000, resulting in the fragmentation of core forest interiors, and this forest loss continues to date due (Wickham et al., 2007; Drummond and Loveland, 2010). Generally, mines in this region are reclaimed to grasslands or shrublands (Simmons et al., 2008; Kazar and Warner, 2013), although more recently there has been interest in reclamation using native forest species (Zipper et al., 2011). Hooke (1994; 1999) estimated that surface mining is responsible for displacing more material in the southern coalfields of West Virginia than river systems and other natural geomorphic processes.

Over the last few decades, surface mining and reclamation have been mapped using moderate resolution satellite data, such as Landsat Multispectral Scanner (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and SPOT data, with varying degrees of success (Anderson and Schubert, 1976; Irons and Kennard, 1986; Parks et. al., 1987; Rathore and Wright, 1993; Anderson et al., 1997; Prakash and Gupta, 1998; Yuill, 2003; Townsend et al., 2009; Sen et al., 2012). Rathore and Wright (1993) noted that active mines can generally be mapped with high accuracy; mine reclamation, however, has proven more difficult to map. Mine reclamation is of specific interest as this imprint of mining generally persists as a legacy landscape alteration (Negley, 2002).

Multi-temporal imagery provides a promising approach for mapping surface mines. For example, Townsend et al. (2009) made use of a Landsat image time series to differentiate active mine disturbance and mine reclamation from similar non-mining cover, such as development and grasslands, achieving accuracies above 85%. Sen et al. (2012) made use of a chronosequence of 23 Landsat TM and ETM+ images to separate mining disturbance from other forest-replacing disturbance using disturbance/recovery trajectories.

This work adds to previous work relating to surface mine mapping by investigating the mapping of surface mine land cover at the individual mine complex scale using high resolution data. Many of the research questions posed here, such as GEOBIA classification, use of textural measures, and exploitation of high resolution data, were suggested as future research needs and knowledge gaps by Townsend et al. (2009) for mapping surface mines and mine reclamation.

3. Research Questions

This work focuses on mine mapping from high resolution data and mapping using advanced image processing techniques, machine learning algorithms and object-based analysis, by addressing the following research questions:

Question 1: Does satellite imagery have inherent benefits (such as radiometric consistency over large areas) compared to aerial imagery that makes satellite imagery preferable for mapping mine-landscapes?

Question 2: What image processing techniques (e.g. textural measures) can be used to increase the classification accuracy for mapping of mining-related land cover from single-date aerial imagery?

Question 3: In comparison to only using multispectral data, does combining LiDAR-derivatives or terrain variables with multispectral data, increase the classification accuracy of mapping mine-landscapes using:

(a) pixel-based classification?

(b) GEOBIA classification?

Question 4: What derived features (such as measures of central tendency, variability, texture, etc.) calculated from multispectral imagery and LiDAR data are most important for classification of mine-related land cover using GEOBIA?

Question 5: How does the classification performance of different machine learning classifiers (random forests (RF), boosted classification and regression trees (CART), and support vector machines (SVM)) compare for mapping of mine-landscapes from high resolution data?

4. Presentation

The research is presented in four research papers, included in this dissertation as the following four chapters.

> **Chapter 2:** Maxwell, A.E., T.A. Warner, M.P. Strager, and M. Pal, 2014. Combining RapidEye satellite imagery and LiDAR for mapping of mining and mine reclamation, *Photogrammetric Engineering & Remote Sensing*, 80(2): 179-189. doi: 10.14358/PERS.80.2.179-189. (Received 10 June 2013, Accepted 25 August 2013)
>
> **Chapter 3:** Maxwell, A.E., M.P. Strager, T.A. Warner, N.P. Zégre, and C.B. Yuill, 2014. Comparison of NAIP orthophotography and RapidEye satellite imagery for mapping of mining and mine reclamation, *GIScience & Remote Sensing*, 51(3): 301-320. doi: 10.1080/15481603.2014.912874. (Received 20 December 2013, Accepted 26 March 2014)
>
> **Chapter 4:** Maxwell, A.E., T.A. Warner, M.P. Stager, J.F. Conley, and A.L. Sharp, 2015. Assessing machine-learning algorithms and image- and lidar-derived variables for GEOBIA classification of mining and mine reclamation, *International Journal of Remote Sensing*, 36(4): 954-978. doi: 10.1080/01431161.2014.1001086. (Received 7 September 2014, Accepted 29 November 2014)
>
> **Chapter 5:** Maxwell, A.E., and T.A. Warner, 2015. Differentiating mine-reclaimed grasslands from spectrally similar land cover using terrain variables and object-based machine learning classification, *International Journal of Remote Sensing* (In Press). doi: 10.1080/01431161.2015.1083632. (Received 6 April 2015, Accepted 11 August 2015)

Each chapter has been formatted in the style of the journal to which it was submitted. The final chapter, Chapter 6, offers a synthesis of the research relative to the aims and questions posed in this dissertation.

5. List of Acronyms

A list of acronyms is provided below to aid in the reading of this dissertation.

Table 1. Acronym list.

Acronym	**Meaning**
CART	classification and regression trees
CTMI	compound topographic moisture index
DEM	digital elevation model
DT	decision tree
ETM+	Enhanced Thematic Mapper Plus
GEOBIA	geographical object-based image analysis
GLCM	grey level co-occurrence matrix
GME	Geospatial Modeling Environment
GPS	global positioning system
GSD	ground sampling distance
k-NN	*k*-nearest neighbor
LiDAR	light detection and ranging
LULC	land use/land cover
MMU	minimal mapping unit
MODIS	Moderate Resolution Imaging Spectrometer
MSS	Multispectral Scanner
NAIP	National Agriculture Imagery Program
nDSM	normalized digital surface model
NIR	near infrared
OOB	out-of-bag
RBF	radial basis function
RF	random forests
SMCRA	Surface Mine Control and Reclamation Act
SPOT	Satellite Pour l'Observation de la Terre
SVM	support vector machines
TM	Thematic Mapper

6. Reference

Anderson, A.T., D. Schultz, N. Buchman, and H.M. Nock, 1997. Landsat imagery for surface-mine inventory, *Photogrammetric Engineering & Remote Sensing*, 43(8): 1027-1036.

Anderson, A.T., and J. Schubert, 1976. ERTS-1 data applied to strip mining, *Photogrammetric Engineering & Remote Sensing*, 42: 211-219.

Bernhardt, E.S., and M.A. Palmer, 2011. The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians, *Annals of the New York Academy of Sciences*, 1223(1): 39-57.

Drummond, M.A. and T.R. Loveland, 2010. Land-use pressure and a transition to forest-cover loss in the eastern United States, *Bioscience*, 60(4): 286-298.

Fritz, K.M., S. Fulton, B.R. Johnson, C.D. Barton, J.D. Jack, D.A. Word, and R.A. Burke, 2010. Structure and functional characteristics of natural and constructed channels draining a reclaimed mountaintop removal and valley fill coal mine, *Journal of the North American Benthological Society*, 29(2): 637-689.

Hooke, R.L., 1994. On the efficacy of humans as geomorphic agents, *GSA Today*, 4(9): 217, 224-225.

Hooke, R.L., 1999. Spatial distribution of human geomorphic activity in the United States: Comparison to rivers, *Earth Surface Processes and Landforms*, 24(8): 687-692.

Irons, J.R., and R.L. Kennard, 1986. The utility of Thematic Mapper sensor characteristics for surface mine monitoring, *Photogrammetric Engineering & Remote Sensing*, 52: 389-396.

Kazar, S.A., and T.A. Warner, 2013. Assessment of carbon storage and biomass on minelands reclaimed to grassland environments using Landsat spectral indices, *Journal of Applied Remote Sensing*, 7(1): 073583. DOI: 10.1117/1.JRS.7.073583.

Merriam, E.R., J.T. Petty, M.P. Strager, A.E. Maxwell, and P.F. Ziemkiewicz, 2013. Scenario analysis predicts context-dependent stream response to landuse change in a heavily mined Central Appalachian watershed, *Freshwater Science*, 32(4): 1246-1259.

Meybeck, M., 2003. Global analysis of river systems: From earth system controls to Anthropocene syndromes, *Philosophical Transactions of The Royal Society*, 358: 1935-1955.

Negley, T.L., 2002. *A Comparative Hydrologic Analysis of Surface-Mined and Forested Watersheds in western Maryland*, M.S. Thesis, University of Maryland, College Park, Maryland, 84 p.

Negley, T.L., and K.N. Eshleman, 2006. Comparison of streamflow responses of surface-mined and forest watersheds in the Appalachian Mountain, USA, *Hydrological Processes*, 20(16): 3467-3483.

Palmer, M.A., E.S. Bernhardt, W.H. Schlesinger, K.N. Eshleman, E. Fourfoula-Georgiou, M.S. Hendryx, A.D. Lemly, G.E. Likens, O.L. Loucks, M.E. Power, P.S. White, and P.R. Wilcock, 2010. Mountaintop mining consequences, *Science*, 327: 148-149.

Parks, N.F., G.W. Petersen, and G.M. Baumer, 1987. High-resolution remote sensing of spatially and spectrally complex coal surface mines of Central Pennsylvania: A comparison between SPOT, MSS, and Landsat-5 Thematic Mapper, *Photogrammetric Engineering & Remote Sensing*, 53(4): 415-420.

Prakash, A., and R.P. Gupta, 1998. Land-use mapping and change in a coal mining area: A case study in the Jharia coalfields, India, *International Journal of Remote Sensing*, 19(3): 391-410.

Rathore, C.S., and R. Wright, 1993. Monitoring environmental impacts of surface coal-mining, *International Journal of Remote Sensing*, 14(6): 1021-1042.

Saylor, K.L., 2008. Land Cover Trend Project: Central Appalachians, U.S. Department of the Interior, U.S. Geological Survey. Washington, DC, URL: http://landcovertrends.

usgs.gov/east/eco69 Report.html (last date accessed: 1 June 2013).

Sen, S., C.E. Zipper, R.H. Wynne, and P.F., Donovan, 2012. Identifying revegetated mines as disturbance/recovery trajectories using an interannual Landsat chronosequence, *Photogrammetric Engineering & Remote Sensing*, 78(3): 223-235.

Simmons, J.A., W.S. Currie, K.N. Eshleman, K. Kuers, S. Monteleone, T.L. Negley, B.R. Pohlad, and C.L. Thomas, 2008. Forest to reclaimed mine land use change leads to altered ecosystem structure and function, *Ecological Applications*, 18(1): 104-118.

Steffen, W., J. Grinevald, P. Crutzen, and J. McNeill, 2011. The Anthropocene: Conceptual and historical perspectives, *Philosophical Transactions of The Royal Society*, 369: 842-867.

Steffen, W., P.J. Crutzen, and J.R. McNeill, 2007. The Anthropocene: Are humans now overwhelming the great forces of nature?, *Ambio*, 36(8): 614-621.

Townsend, P.A., D.P. Helmers, C.C. Kingdon, B.E. McNeil, K.M. de Beurs, and K.N. Eshleman, 2009. Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976-2006 Landsat time series, *Remote Sensing of Environment*, 113: 62-72.

Wickham, J.D., K.H. Riitters, T.G. Wade, M. Coan, and C. Homer, 2007. The effect of Appalachian mountaintop mining on interior forest, *Landscape Ecology*, 22: 179-187.

Yuill, C., 2003. Landscape use assessment: Mountaintop mining and the mountaintop mining region of West Virginia, *Draft Programmatic Environmental Impact Statement on Mountaintop Mining/Valley Fills in Appalachia*, P. III, F-12.

Zalasiewicz, J., M. Williams, A. Haywood, and M. Ellis, 2011. The Anthropocene: A new epoch of geology time?, *Philosophical Transactions of The Royal Society*, 369: 835-841.

Zalasiewicz, J., M. Williams, W. Steffen, and P. Crutzen, 2010. The new world of the Anthropocene, *Environmental Science & Technology Viewpoints*, 44: 2228-2231.

Zégre, N.P., A.E. Maxwell, and S. Lamont, 2013. Characterizing streamflow response of a mountaintop-mined watershed to changing land use, *Applied Geography*, 39: 5-15.

Zipper, C.E., J.A. Burger, J.G. Skousen, P.N. Angle, D. Barton, V. Davis, and J.A. Franklin, 2011. Restoring forests and associated goods and services on Appalachian coal surface mines, *Environmental Management*, 47(5): 751-765.

CHAPTER 2

Combining RapidEye Satellite Imagery and LiDAR for Mapping of Mining and Mine Reclamation[1]

Aaron E. Maxwell, Timothy A. Warner, Michael P. Strager, and Mahesh Pal

Abstract

The combination of RapidEye satellite imagery and light detection and ranging (LiDAR) derivatives was assessed for mapping land cover within a mountaintop coal surface mine complex in the southern coalfields of West Virginia, USA. Support vector machines (SVM), random forests (RF), and boosted classification and regression trees (CART) algorithms were used. Incorporation of the LiDAR-derived data increased map accuracy in comparison to using only the five imagery bands, and SVM generally produced a more accurate classification than the ensemble tree algorithms based on overall map accuracy, Kappa statistics, allocation disagreement, quantity disagreement, and McNemar's test of statistical significance. Based on measures of predictor variable importance within the ensemble tree classifiers, the normalized digital surface model (nDSM) was found to be more useful than first return intensity data for differentiating the classes.

[1] This is an Accepted Manuscript of an article published by the American Society for Photogrammetry and Remote Sensing (ASPRS) in *Photogrammetric Engineering & Remote Sensing* in February 2014, [10.14358/PERS.80.2.179-189]. Maxwell, A.E., T.A. Warner, M.P. Strager, and M. Pal, 2014. Combining RapidEye satellite imagery and LiDAR for mapping of mining and mine reclamation, *Photogrammetric Engineering & Remote Sensing*, 80(2): 179-189. (Received 10 June 2013, Accepted 25 August 2013)

Introduction

Commercial satellite imagery such as IKONOS, GeoEye, and RapidEye provide high spatial resolution but low spectral resolution compared to sensors such as Landsat Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), or Moderate Resolution Imaging Spectrometer (MODIS) (Warner et al., 2009). Although high spatial resolution can yield fine detail for land cover and vegetative mapping, classification is complicated by the increased spatial resolution and decreased spectral resolution. Fine spatial resolution tends to generate high internal variability within land cover classes, which can lead to decreases in classification accuracy (Townshend, 1981; Cushnie, 1987; Townshend, 1992; Baker et al., 2013). This research investigated a potential means to enhance classification accuracy by combining high resolution commercial satellite imagery with light detection and ranging (LiDAR) data.

The analysis focused on mapping land cover classes in a mountaintop coal surface mine complex in the southern coalfields of the eastern United States. Because surface mine complexes experience rapid change due to human disturbance and reclamation, they are particularly good examples of disturbed landscapes. Although this research focuses on mapping land cover within a mountaintop coal mine, the challenges in mapping mining landscapes are typical of other disturbances, such as timber harvesting, urban sprawl, etc.

This work adds to prior remote sensing of surface mine research by investigating information gained by combining LiDAR and commercial satellite data for mapping land cover (Cowen et al., 2000). This research had two components. First, we assessed

LiDAR-derived inputs as predictor variables when combined with commercial satellite imagery to enhance land cover mapping. Second, we compared three machine learning algorithms for the classification: support vector machines (SVM), random forests (RF), and boosted classification and regression trees (CART). The image data consisted of commercial RapidEye imagery. LiDAR-derived predictor variables included the normalized digital surface model (nDSM) generated by subtracting ground return data from the first return data, first return intensity data, and the first return intensity range within raster grid cells.

Background

Machine learning classification

Research has highlighted the improvement in classification accuracy when LiDAR is combined with optical data, suggesting that LiDAR can provide important predictor variables for mapping land cover (Cowen et al., 2000; Brennan and Webster, 2006; Bork and Su, 2007; Chust et al., 2008; Chen et al., 2009; Guo et al., 2011). The combination of imagery and LiDAR data has been investigated in heterogeneous rangeland environments (Chen et al., 2009), urban landscapes (Brennan and Webster, 2006; Guo et al., 2011), and coastal estuary environments (Brennan and Webster, 2006; Chust et al., 2008). Guo et al. (2011) specifically noted the usefulness of nDSM data for mapping urban landscapes.

Combining disparate data such as imagery and LiDAR poses distinct challenges because the combined data set may not meet distribution assumptions required for traditional parametric classifiers. Machine learning algorithms have emerged as an

alternative to parametric classifiers and have been shown to be more accurate and efficient when faced with high dimensional and complex data (Hansen et al., 1996; Huang et al., 2002; Rogan et al., 2003; Pal, 2005; Na et al., 2010; Ghimire et al., 2012). Machine learning algorithms, such as artificial neural networks (Del Frate et al., 2003), SVM (Pal and Mather, 2005; Pal, 2005), and decision trees (Waske and Braun, 2009), do not make assumptions regarding the data distribution (Loosvelt et al., 2012). In summary, in remote sensing, machine learning algorithms are of interest because they offer the potential to handle complex spectral measurement space, multidimensional data, and large volumes of data with reduced processing time compared to traditional classifiers (Hansen and Reed, 2000).

For this study, SVM, RF, and boosted CART algorithms were assessed. SVM and RF have been shown to have comparable accuracies; however, optimizing the RF algorithm is simpler (Pal, 2005). RF and boosted CART have also been shown to provide similar accuracies, though RF provides shorter classification time (Gislason et al., 2006). However, boosted CART has been shown to be more suited for large-area mapping because it is marginally less sensitive to training data size and less sensitive to training data noise (Ghimire et al., 2012).

SVMs make use of statistical learning theory and optimization algorithms to locate decision boundaries between classes using structural risk minimization to find a multi-dimensional plane (hyperplane) that separates two classes with the maximum margin (Vapnik, 1995; Joachims, 1998; Burges, 1998; Pal and Mather, 2005; Pal, 2005; Warner and Nerry, 2009). The points that lie near the hyperplane define the margin and

are therefore termed "support vectors." Because SVM algorithms are designed for two class problems only, strategies have been developed incorporating multiple SVM algorithms to produce a multi-class classification (Vapnik, 1995; Pal and Mather, 2005; Pal, 2005). SVM within the e1071 package, used in this research, uses a "one-against-one" approach for multiclass-classification in which binary classifiers are trained and the appropriate class is found by a voting scheme (Meyer et al., 2012).

RF, introduced by Breiman (2001), uses multiple decision trees to improve upon the accuracy and consistency of single tree classifications. As a result, RF is an ensemble, non-parametric learning algorithm. A random bootstrap sample of the data with replacement (called "bagging" (Breiman, 1996)) is drawn for each tree generated instead of using the entire training dataset. In addition, an out-of-bag (oob) random sample is withheld that can be used for accuracy assessment. To grow a tree, RF uses a random subset of the predictor variables (the number of which is defined by user) while growing each tree of the ensemble. This results in a decrease in the strength of a single tree, however, the correlation between trees is reduced. As a result, the randomized predictor variable selection reduces the generalization error. The Gini index, a measure of impurity of a given class with respect to the rest of the classes, is used to select the best predictor among the randomly selected predictor variables available at each node. RF has been used for classification in remote sensing (Ghimire et al., 2010; Burkholder et al., 2011; Rodríguez-Galiano et al., 2011; Ghimire et al., 2012), and it has many attributes that make it effective for image classification. It can generally model complex interactions between predictor variables, perform supervised and unsupervised classification tasks,

handle data with missing values, and provides high classification accuracies. It also provides measures of predictor variable importance, as will be described below (Steele, 2000; Cutler et al., 2007).

Although boosted CART, like RF, also implements an ensemble classification of decision trees, the algorithm is fundamentally different in that it uses the entire training dataset for classification of each tree as opposed to a subset for each tree as used with RF (Freund and Schapire, 1996; Ghimire et al., 2012). Misclassified samples in prior trees are given higher weights in subsequent trees, which tends to reduce misclassification in subsequent trees. However, it has been shown that this weighting procedure may overfit the training data (Bauer and Kohavi, 1999). Many studies have shown boosted CART to produce more accurate classifications than individual trees (DeFries and Chan, 2000; Muchoney et al., 2000; Friedl et al., 1999; Friedl et al., 2002; McIver and Friedl, 2002; Lawrence et al., 2004; Ghimire et al., 2012). Friedl et al. (1999) found that boosting can reduce misclassification rates by 20% to 50% in comparison to single classification trees, whilst Lawrence et al. (2004) found an improvement of 9%.

LiDAR is an active remote sensing technology that relies on the timing of the two-way travel of laser pulses, differential global positioning system (GPS), and inertial navigation measurements to map the height of the terrain and objects on the ground surface. When LiDAR systems are mounted in aerial platforms, large areas can be surveyed in a relatively short period of time (Lillesand et al., 2008). In addition to elevations, most systems record the return pulse intensity, which is in part a function of the reflectance of the surfaces returning the laser pulse (Brantberg, 2007). However,

return intensity is also influenced by other factors, including footprint size, scan angle, return number, and range distance (Lin and Mills, 2010). Although some studies have demonstrated the benefit of using intensity for land cover classification (Brennan and Webster, 2006), the use of LiDAR return intensity has not been as widely explored as elevation data in land cover mapping due to the difficulty of radiometric calibration of the returned laser intensity (Flood, 2001; Kaasalainene et al., 2005). In this research, both LiDAR elevation and intensity data were used.

Description of Study Area

The study area in in the southern coalfields of the eastern United States, which are found predominantly within the Appalachian Plateau physiogeographic province, a dissected, westward-tilted plateau dominated by Pennsylvanian strata. Pennsylvanian stratigraphy is characterized by cyclic sequences of sandstone, shale, clay, coal, and limestone (Ehlke et al., 1982; WVGES, 2005). The terrain is dissected by a dendritic stream network and show moderate to strong relief with steep slopes. The forests are characterized by mixed mesophytic communities, deciduous species, and high biodiversity (Strausbaugh and Core, 1977).

In the southern coalfields, mountaintop coal mining is the leading cause of land use/land cover (LULC) change (Hooke, 1999; Saylor, 2008; Townsend et al., 2009; Drummond and Loveland, 2010). This mining process causes more material to be moved and faster landscape alterations than more traditional surface coal mining techniques such as auger, contour, and highwall mining (Fritz et al., 2010).The mountaintop mining process results in the clearing of forests, the removal of top soil, and recontouring of the

landscape (Palmer et al., 2010; Bernhardt and Palmer, 2011). It has been estimated that surface mining in the Appalachian region has resulted in a net loss of 420,000 ha of forest between 1973 and 2000, resulting in the fragmentation of core forest interiors (Wickham et al., 2007; Drummond and Loveland, 2010). Generally, mines in this region are reclaimed to grasslands or shrublands (Simmons et al., 2008; Kazar and Warner, 2013), although more recently there has been interest in reclamation using native forest species (Zipper et al., 2011). Hooke (1994; 1999) estimated that surface mining is responsible for displacing more material in the southern coalfields of West Virginia than river systems and natural geomorphic processes.

Over the last few decades, surface mining and reclamation have been mapped from Landsat MSS, TM, ETM+, and SPOT data with varying degrees of success (Anderson and Schubert, 1976; Irons and Kennard, 1986; Parks et. al., 1987; Rathore and Wright, 1993; Anderson et al., 1997; Prakash and Gupta, 1998; Yuill, 2003; Townsend et al., 2009; Sen et al., 2012). Rathore and Wright (1993) found that active mines could generally be mapped with high accuracy; mine reclamation, however, has proven more difficult to map. Mine reclamation is of specific interest as the imprint of mining generally persists as a legacy landscape alteration (Negley, 2002). Multi-temporal imagery provides a promising approach for mapping surface mines. For example, Townsend et al. (2009) made use of a Landsat image time series to differentiate active mine disturbance and mine reclamation from similar non-mining cover, such as development and grasslands, achieving accuracies above 85%. Sen et al. (2012) made use

of a chronosequence of 23 Landsat TM and ETM+ images to separate mining disturbance from other forest-replacing disturbance using disturbance/recovery trajectories.

Study Area

The study area is the Hobet-21 mountaintop mine in Boone and Lincoln counties, West Virginia, USA, shown in Plate 1. This is the largest mine complex in the Appalachian region (Keene and Skousen, 2010). This mine was selected because of the wide variety in age of disturbance, vegetation, and land cover. Historical imagery shows that some of the mine disturbance predates 1987, while portions of the mine were still active at the time of the writing. The mine complex boundary was derived from Surface Mine Control and Reclamation Act (SMCRA) surface mine permit boundaries provided by the West Virginia Department of Environmental Protection (WVDEP) and additional ancillary data. The boundary of the Hobet-21 mine encompasses an area of 5,500 ha of disturbance, reclaimed, and forested (i.e. not disturbed) land.

Methods

Data

The primary optical dataset was a RapidEye multispectral satellite image collected on 1 April 2010, prior to spring leaf out. The scene has an average center image view angle of -2.82˚, azimuth angle of 110.2˚, sun azimuth of 171.2˚, and sun elevation of 56.5˚. The RapidEye system consists of a constellation of 5 satellites that were launched in August 2008. The satellites have sensors with five spectral bands: blue (440-510 nm), green (520-590 nm), red (630-730 nm), red edge (690-730 nm), and near infrared (NIR) (760-850 nm) (Tyc et al., 2005). The ground sampling distance of the

system is 6.5 m. For this project, the 3A product was obtained, which has radiometric, sensor, geometric corrections, and orthorectification applied, including a resampling of the pixels to 5 m. The Hobet-21 mine complex is covered by a single image tile.

LiDAR data were collected for the study area on 12 April 2010, using an aircraft flying height at 1524 m above ground level (AGL) at a speed of 125 knots. The Optech ALTM 3100 C sensor was set to a pulse frequency of 70 kHz, a scan frequency of 35 Hz, and a scan angle of 36˚ (full) to generate points with a nominal pulse spacing of 1 m. A 30% overlap was acquired between swaths. The LiDAR system recorded up to 4 returns per laser pulse. The point data were classified by the vendor as ground, non-ground, or outliers, and delivered in LAS 1.2 format. The point classifications were utilized as they were provided, and no additional point classification or editing was performed.

The imagery and LiDAR datasets were acquired within two weeks of each other. Visual comparison of the RapidEye data and LiDAR-derivatives throughout the mine complex showed little evidence of landscape alterations within this two week period.

Pre-Processing

Trial-and-error experimentation of raster interpolation methods and settings was performed, and an optimal approach was chosen based on a subjective evaluation of the products generated. The point data were first converted to raster data using the LAS Dataset to Raster utility in ArcMap (ESRI, 2012). A digital elevation model (DEM) was produced from the ground return point data on a 5 m grid matching that of the RapidEye imagery. A digital surface model (DSM) was then produced using the first returns. The raster grids were produced using the average value within each 5 m pixel and a linear

interpolation to fill data gaps with no returns within a grid cell. The DEM was subtracted from the DSM to produce an nDSM, a measure of height of objects on the ground, such as vegetation or buildings.

A raster of first return intensity was produced using the ArcMap LAS Dataset to Raster utility. The point values were gridded using the average value within each 5 m grid cell. Data gaps were labeled with a value of zero, because lack of returns could be a result of complete absorption of the laser pulse, for example by water. The range of intensity values within each 5 m pixel was also calculated, with data gaps also labeled as zero. Only first returns were used for the intensity products as intensity values of subsequent returns should not be mixed with first returns (Brantberg, 2007).

All data layers were converted to a data range appropriate for a 16 bit radiometric resolution. An image stack was produced using ERDAS Imagine 9.3 (ERDAS, 2002) comprising eight layers: the five RapidEye bands and the three LiDAR layers, the nDSM, first return intensity, and first return intensity range. This layer stack was used as the predictor variables used to map land cover. Co-registration can be a problem when data sets acquired with different sensors and at different times are integrated in a single analysis. For this study, the LiDAR data have a nominal horizontal error of 0.3 m and vertical error of 0.12 m (Optech, 2008; RAMPP, 2011). The RapidEye data were provided by the supplier as an orthorectified product (termed 3A), which has a nominal error of potentially less than 1 pixel under ideal circumstances (e.g. nadir view and flat terrain) (RapidEye, 2009). The RapidEye data over the study site were acquired at a view angle of -2.82˚(i.e. close to nadir), but the topography of the study site is complex,

therefore the error may be greater than 1 pixel. Nevertheless, a visual inspection of an overlay of the two datasets did not show noticeable differences, suggesting that the registration was adequate for the analysis and not likely to affect the classification.

Land Cover Classes of Interest

Five land cover categories were mapped: forested, reclaimed-herbaceous vegetation, reclaimed-woody vegetation, barren, and water (Table 1). These classes were chosen based on prior experience of typical land cover conditions within surface mines and reclaimed surface mines in the region. It should be noted that the term "forested land cover" is used to indicate forests in the permit area not yet removed for the mountaintop removal mining. The term does not necessarily imply that the area has never been disturbed at all; West Virginia's landscape records extensive historical disturbances ranging from clear-cutting to prior surface coal mining.

Training Data

Training areas were delineated by manual interpretation of 1 m high resolution orthophotography and checked against the RapidEye imagery, LiDAR nDSM, and LiDAR bare earth contour data, which were useful for differentiating the vegetation classes and reclaimed surfaces. The orthophotography was collected as natural color images by Pictometry International Corporation (Rochester, New York) in March of 2010 for the state of West Virginia, approximately one month before the LiDAR and RapidEye data acquisition. The number of training polygons collected is summarized in Table 2.

Using Geospatial Modeling Environment (GME) (Beyer, 2012), 1,000 point examples of each class were randomly selected from the digitized polygons as training

pixels. Due to the small proportion of water within the mine complex, it was not possible to collect 1,000 examples for that class. In total, 4,517 pixels were utilized to train the models as summarized in Table 2.

Classification Methods

Per-pixel classifications were performed using the three machine learning algorithms within the statistical software tool R (R Core Team, 2012). SVM classification was performed using the e1071 package (Meyer et al., 2012), RF using the randomForest package (Liaw and Wiener, 2002), and boosted CART using the adabag package (Alfaro-Cortes et al., 2012).

In order to optimize the parameters required for the SVM and RF algorithms in this study, a grid search of the specified parameters was undertaken in R. Optimal parameters were estimated based on the minimum classification error obtained for the training data using a 10-fold cross validation, in which the data were partitioned into ten unique training sets, using a random assignment. The classifier was then trained ten times, and each time the remaining data, not used for training in that instance, were used for validation. Overall accuracy was calculated as the average across the ten classifications. Optimization was performed for each of the three combinations of predictor variables (bands) used to produce a classification. This ensured that each classification used optimal parameters for the particular set of predictor variables tested. For SVM a radial basis function (RBF) kernel was used and the cost (C) and gamma (γ) parameters were optimized. Once a coarse grid search on a wide range of parameter values was performed, a second grid search of values centered on the optimal settings

predicted was performed in order to further refine the estimate. For RF the number of predictor variables randomly sampled as candidates at each node (*m*) and also the number of trees to grow (*k*) were selected using a grid search procedure.

Classifications were performed for three combinations of the predictor variables for each machine learning algorithm in order to assess their performance in terms of classification accuracy. The combinations included the following:

1. Image Bands (5) = 5 Predictors Variables
2. LiDAR-Derived Data (3) = 3 Predictor Variables
3. Image Bands (5) + LiDAR-Derived Data (3) = 8 Predictor Variables

Error Assessment

Classification accuracy was assessed through an independent dataset, separate from the training data. A total of 1,325 pixels were randomly collected, 265 in each of the five land cover classes. No pixels from within the training areas were used. A stratified sample for the accuracy assessment was chosen because reclaimed-woody vegetation and water classes are not common cover types in the study area and thus would typically be only rarely selected in a purely random approach.

At each of the 1,325, 5 m x 5 m validation pixels, the dominant land cover category was visually interpreted from the Pictometry, RapidEye, and LiDAR data. Each pixel was assessed twice to ensure consistency in the interpretation.

From the randomly selected accuracy assessment data, error matrices and Kappa statistics were produced. As this was a stratified sample, overall accuracy was calculated

by weighting the class accuracies by the proportion of land cover within the study area (Stehman and Foody, 2009). Following Pontius and Millones (2011), quantity disagreement and allocation disagreement were also calculated. These two measures segment disagreement between maps into difference in proportion of the classes and spatial allocation of the classes. Although the value of Kappa is questioned by Pontius and Millones (2011), the statistic was nevertheless provided for potential comparisons to other studies.

In order to evaluate the statistical significance of any differences in the classifications, the results were compared on a pairwise basis using McNemar's test (Dietterich, 1998; Foody, 2004). From the accuracy assessment data, which were used to assess all models, and classification results, 2-by-2 confusion matrices were produced that summarized which pixels were correctly classified by both classifiers, which were incorrectly classified by both, and which were classified correctly by one classifier and not the other. McNemar's test is a non-parametric test of statistical difference that allows for the calculation of a *z*-score from this 2-by-2 matrix. A *z*-score larger than 1.645 indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is better than the other (Bradley, 1968; Dietterich, 1998; Foody, 2004; Agresti, 2007).

The RF algorithm generates an estimate of predictor variable importance during training by excluding each variable sequentially and recording the resulting increased oob error (Breiman, 2001; Rodríguez-Galiano et al., 2012). This approach was used to

determine which predictor variables were most important in the model given all input predictor variables, both the multispectral bands and the LiDAR-derivatives.

Results and Discussion

The optimized algorithm parameter settings for the classifiers for each input variable combination are listed in Table 3. For SVM, the optimal setting for gamma (γ) and cost (C) were different for all band combinations. For RF, the optimal number of input variables sampled at each node (m) varied between 2 and 3 input variables. A total of 500 trees (k) was used for all band combinations for both RF and boosted CART because this was found to be sufficient to stabilize the classification regardless of band combinations used. Breiman (2001) notes that as the number of trees grown increases, generalization error converges and overfitting should not be a problem due to the "strong law of large numbers." Parameter optimization was the most time consuming process in producing the classification models. SVM took the longest time to optimize and was the most complex to optimize because a wide range of values for γ and C had to be tested (Pal, 2005).

The SVM algorithm using the combination of the multispectral and LiDAR-derived predictor variables resulted in the classification with the highest overall accuracy, 86.4% (Table 4). The resulting map is shown in Plate 2 with area and percentage of cover types within the mine permit summarized in Table 5. The map shows that reclaimed-herbaceous vegetation is the most common cover type within the mine permit followed by barren areas. A smaller percentage of the reclaimed area was classified as reclaimed-

woody vegetation, mostly in the eastern portion of the mine, where reclamation is older and more established.

The highest classification accuracies were achieved by the combination of the multispectral and LiDAR-derived data, then the image data alone, and finally the LiDAR-derivatives, regardless of the algorithm used, as shown by Table 4. The McNemar's tests, summarized in Table 6(a), 6(b), and 6(c), indicate the differences between the classification accuracies using the different combinations of predictor variables were generally statistically significant. The only exceptions are the RF and boosted CART algorithms ((Table 6(b) and (c)) for which the classification of the multispectral imagery was not statistically more accurate than classification using the LiDAR-derivatives.

Table 4 indicates that the SVM algorithm outperformed the ensemble methods regardless of the predictor variable combinations used: for all data combinations, the SVM algorithm provided the highest map accuracy. Table 6(d) and 6(f) show that difference was statistically significant between RF and also boosted CART for the classification of the imagery and the combined imagery and LiDAR; however, none of the algorithms were shown to be statistically different for the classification of the LiDAR-derivatives (Table 6(e)).

Allocation disagreement is generally larger than quantity disagreement, and also varies over a wide range of values (Table 4). The SVM classification using the imagery and LiDAR-derived predictor variables has the lowest allocation disagreement, and the second to lowest quantity disagreement. These measures also suggest that there is merit in combining the data sources and using the SVM algorithm.

In summary, for the comparison of different predictor variables (Table 6(a) – (c)), only two combinations were reported as not statistically different out of the possible nine combinations; for the comparison of classification algorithms (Table 6(d) – (f)), five of the possible nine combinations were shown not to be statistically different, and the *z*-scores were often lower than those for the input band combinations. The largest *z*-score, or most significant difference, is for the comparison of the LiDAR data and the combination of imagery and LiDAR, classified using the SVM algorithm. The variables used generally had a larger impact on final map accuracy than the algorithm. Even though the SVM algorithm generated the most accurate output, it was the most difficult to optimize because a wide range of values for γ and C must be tested for optimization, as already noted above. Fewer unique values for m and k must be tested for RF, and only k must be specified for boosted CART. Even though the classifications using the different methods showed statistical differences, the question arises as to whether a 1 to 3% increase in map accuracy merits the increased complexity required for optimizing SVM. Perhaps a simpler method, that has a 1-2% lower accuracy, may be a better choice for routine mapping, for the sake of simplicity.

On the other hand, SVMs have been shown to be relatively robust with respect to parameter settings (Melgani and Bruzzone, 2004), and as parameter optimization was the most time consuming step in the SVM analysis, the question arises as to whether this step is necessary. To address this question, a classification using the default R parameter settings (kernel=RBF, C=1, γ=1/8) (Meyer et al., 2012) using the combined imagery and LiDAR data was generated. The resulting overall accuracy was 86.3%, compared to

86.4% for the classification produced using the optimal parameters, a decrease of only 0.1%. There was no statistical difference between the default and optimal classifications using McNemar's test (z-score=0.254). If, as this result would suggest, optimization does not have to be performed to obtain an accurate classification, this further strengthens the argument for the use of the SVM algorithm.

As the SVM algorithm provided the most accurate classifications, those results will be used to compare error matrices for the three combinations of variables. Tables 7, 8, and 9 show the error matrices for the image bands only, LiDAR-derivatives only, and the imagery bands and the LiDAR-derivatives, respectively. In comparison with image variables alone, the combined LiDAR and image variables show an increase in the number of pixels correctly classified for all land cover classes, other than water. This suggests that the addition of the LiDAR-derived data provided enhanced separation of multiple classes rather than improving only one class. Generally, the LiDAR-derived predictor variables alone provided the lowest accuracy per class. These tables show the benefit for all classes of combining the LiDAR and multispectral image data sources, and they also suggest that LiDAR-derivatives are not a substitute for image data for land cover classification as they do not provide comparable accuracies. Instead, LiDAR data should be used to enhance classification using multispectral data.

The error matrices (Tables 7-9), generally speaking, suggest that reclaimed-herbaceous cover was commonly confused with the reclaimed-woody vegetation, forested, and barren classes. The confusion with barren cover may be attributed to the patchy, heterogeneous nature of some reclaimed areas, which likely resulted in a mixed

pixel problem. Such areas may be in the early stage of reclamation, in which vegetation is just beginning to regrow on the barren landscape. This issue highlights the complexity of land cover mapping of surface mine complexes in which a wide range of vegetation, disturbance, and historical reclamation practices exist on a steep and heterogeneous landscape. Producer's accuracy for water was also generally low, likely a result of the variability in water reflectance on mine sites. Small settlement ponds, water retention ponds, and treatment ponds may have a wide variety of chemical precipitates, dissolved chemicals, sediment levels, and other factors, which can make them spectrally varied and difficult to classify accurately.

Figure 1 shows predictor variable importance as estimated by the oob mean decrease in accuracy for a classification using the combination of LiDAR and multispectral data. The most important variable in the model was RapidEye Band 5 (NIR) followed by the nDSM. The fact that the nDSM was found to be more important in the model for predicting cover than four of the five spectral bands emphasizes the value of the height information. The LiDAR intensity data were found to be less important than the nDSM and all the image bands, with the exception of Band 4 (Red Edge). The first return intensity range showed the lowest importance. Similar results were found by Chust et al. (2008). Although they found improvements in coastal habitat mapping by combining LiDAR-derived data and mutlispectral imagery, the classification accuracy was most increased by the incorporation of a LiDAR-based digital surface model (DSM) as opposed to other LiDAR-derivatives including intensity, slope, and aspect. Guo et al. (2011) also found height differences, or nDSM, to be the most useful LiDAR-derived

dataset, and they attributed this to the ability of the nDSM to help distinguish between above ground and ground features in an urban setting.

For operational mine monitoring at the scale of individual mine permits, the approach suggested here may provide accurate classification (on the order of 85%) of mine disturbance and reclamation. In this research, visually interpreted high resolution orthophotography was used to facilitate the collection of training data, but this may not be necessary as the RapidEye imagery and LiDAR data provide adequate resolution to interpret the land cover class of the training polygons. The proposed mapping approach could be extended from the individual mine to a regional scale, although potential limitations in doing so could include cloud cover, phenological and illumination changes between RapidEye tiles, and the need for LiDAR data over a large area, which would be expensive to collect, if it were not already available. In addition, for regional studies it may be harder to get LiDAR and RapidEye data that are at least approximately temporally coincident.

Conclusions

High spatial resolution can yield fine detail for land cover mapping; however, reduced classification accuracy is expected due to internal variability within classes and decreased spectral resolution. This poses a problem when high resolution land cover data are required, for example mapping of mine permitted lands as investigated here. One means to combat this problem is combining multiple data sources including imagery and LiDAR. Machine learning algorithms, which tend to be robust in dealing with complex predictor variables, are powerful tools for classifying such data. In a landscape where the

land cover changes rapidly, such as the southern coalfields of the eastern United States, planning and coordinating multiple data collections within a short time period may be of particular importance.

The combination of LiDAR-derived data, including a nDSM, first return intensity, and first return intensity range, together with commercial satellite imagery, aided in the classification of accurate land cover within the surface mine permitted area. The input variable combinations used had a larger impact on final classification accuracy than the algorithm used; however, SVM provided a more accurate classification than the ensemble tree algorithms. Finding the optimal parameters for the SVM classification was time consuming. However, the classification accuracy using the default parameters in R was not statistically different from the accuracy of the classification using the optimal parameters. Thus, omitting the optimization step may not have a major effect on the classification. The highest classification accuracy was produced by a SVM algorithm using a combination of multispectral image bands and LiDAR-derivatives. The nDSM proved to be a particularly important predictor variable for the RF algorithm, and was found to be more useful than the intensity data.

Acknowledgments

Funding support for this study was provided by the West Virginia University Experiment Station and the Appalachian Research Initiative for Environmental Science (ARIES) (http://www.energy.vt.edu/ARIES). Additional support was provided by West Virginia View. LiDAR data were provided by the West Virginia Department of Environmental Protection (WVDEP) and the Natural Resource Analysis Center (NRAC)

at West Virginia University. We would specifically like to acknowledge Adam Riley and Paul Kinder for their assistance in obtaining and processing the LiDAR data. The West Virginia GIS Technical Center provided access to the Pictometry orthophotography, and we would specifically like to thank Kevin Kuhn. Lastly, we would like to thank three anonymous reviewers for their helpful comments that greatly improved the manuscript.

References

Agresti, A., 2007. *An Introduction to Categorical Data Analysis*, 2nd Ed., Wiley-Interscience, Hoboken, New Jersey, 400 p.

Alfaro-Cortes, E., M. Gamez-Martinez, and N. Garcia-Rubio, 2012. Adabag: Applies multiclass AdaBoost.M1, AdaBoost-SAMME and bagging, *R Package Version 3.1*, URL: http://CRAN.R-project.org/package=adabag (last date accessed: 1 June 2013)

Anderson, A.T., and J. Schubert, 1976. ERTS-1 data applied to strip mining, *Photogrammetric Engineering & Remote Sensing*, 42: 211-219.

Anderson, A.T., D. Schultz, N. Buchman, and H.M. Nock, 1997. Landsat imagery for surface-mine inventory, *Photogrammetric Engineering & Remote Sensing*, 43: 1027-1036.

Baker, B., T.A. Warner, J. Conley, and B. McNeil, 2013. Does spatial scale matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations, *International Journal of Remote Sensing*, 34(5): 1633-1651.

Bauer, E., and R. Kohavi, 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants, *Machine Learning*, 36(1): 105-139.

Bernhardt, E.S., and M.A. Palmer, 2011. The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians, *Annals of the New York Academy of Sciences*, 1223(1): 39-57.

Beyer, H.L., 2012. *Geospatial Modeling Environment (Version 0.6.0.0)*, http://www.spatialecology.com/gme (last date accessed: 15 May 2013).

Bork, E.W. and J.G. Su, 2007. Integrating LiDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis, *Remote Sensing of Environment*, 111(1): 11-24.

Bradley, J.V., 1968. *Distribution-Free Statistical Tests*, Prentice Hall, Englewood Cliffs, New Jersey, 388 p.

Brantberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar, *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5): 325-340.

Breiman, L., 1996. Bagging predictors, *Machine Learning*, 24(2): 123-140.

Breiman, L., 2001. Random forests, *Machine Learning*, 54(1): 5-32.

Brennan, R. and T.L. Webster, 2006. Object-oriented land cover classification of lidar-derived surfaces, *Canadian Journal of Remote Sensing*, 32(2): 162-172.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2(2): 121-167.

Burkholder, A., T.A. Warner, M. Culp, and R.E. Landenberger, 2011. Seasonal trends in separability of leaf reflectance spectra for *Ailanthus altissima* and four other tree species, *Photogrammetric Engineering and Remote Sensing*, 77(8): 793-804.

Chen, Y., W. Su, J. Li, and Z. Sun, 2009. Hierarchical object oriented classification using very high resolution imagery and LiDAR data over urban areas, *Advances in Space Research*, 43(7): 1101-1110.

Chust, G., I. Galparsoro, Á Borja, J. Franco, and A. Uriarte, 2008. Coastal and estuarine habitat mapping, using LiDAR height and intensity and multi-spectral imagery, *Estuarine, Coastal and Shelf Science*, 78(4): 633-643.

Cowen, D.C., J.R. Jensen, C.J. Hendrix, M.E. Hodgson, S.R. Schill, 2000. A GIS-assisted rail construction econometric model that incorporates LiDAR data, *Photogrammetric Engineering and Remote Sensing*, 66(11): 1323-1328.

Cushnie, J.L., 1987. The interactive effort of spatial resolution and degree of internal variability within land-cover types on classification accuracies, *International Journal of Remote Sensing*, 8(1): 15-29.

Cutler, D.R., T.C. Edwards, Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler, 2007. Random Forests for classification in ecology, *Ecology*, 88(11): 2783-2792.

DeFries, R.S., and J.C.W. Chan, 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data, *Remote Sensing of Environment*, 74(3): 503-515.

Del Frate, F., G. Schiavon, D. Solimini, M. Borgeaud, D.H. Hoekman, and M.A.Vissers, 2003. Crop classification using multiconfiguration C-band SAR data, *IEEE Transactions on Geoscience and Remote Sensing*, 41(7): 1611-1619.

Dietterich, T.G., 1998. Approximate statistical test for comparing supervised classification learning algorithms, *Neural Computation*, 10(7): 1895-1923.

Drummond, M.A. and T.R. Loveland, 2010. Land-use pressure and a transition to forest-cover loss in the eastern United States, *Bioscience*, 60(4): 286-298.

Ehlke, T.A., G.S. Runner, and S.C. Downs, 1982. *Hydrology of Area 9, Eastern Coal Province, West Virginia*, United States Geological Survey Water-Resources Investigation, Open File Report 81-803, 63 p.

ERDAS Imagine, 2002. *ERDAS Field Guide*, 6th edition, 2002. Leica Geosystems, Atlanta, Georgia, 658 p.

ESRI, 2012. *ArcGIS Desktop: Release 10.1*. Environmental Systems Research Institute, Redlands, California.

Flood, M., 2001. Laser altimetry: From science to commercial lidar mapping, *Photogrammetric Engineering & Remote Sensing*, 67(11): 1209-1217.

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

Friedl, M.A., C.E. Brodley, and A.H. Strahler, 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales, *IEEE Transactions on Geoscience and Remote Sensing*, 37(2): 969-977.

Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf, 2002. Global land cover mapping from MODIS: Algorithms and early results, *Remote Sensing of Environment*, 83(1-2): 287-302.

Fritz, K.M., S. Fulton, B.R. Johnson, C.D. Barton, J.D. Jack, D.A. Word, and R.A. Burke, 2010. Structure and functional characteristics of natural and constructed channels draining a reclaimed mountaintop removal and valley fill coal mine, *Journal of the North American Benthological Society*, 29(2): 637-689.

Freund, Y. and R.E. Schapire, 1996. Experiments with the new boosting algorithm, *Machine Learning: Proceeding of the Thirteenth International Conference*, San Francisco, California, pp. 148-156.

Ghimire, B., J. Rogan, and J. Miller, 2010. Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic, *Remote Sensing Letters*, 1(1): 45-54.

Ghimire, B., J. Rogan, V. Rodríguez-Galiano, P. Panday, and N. Neeti, 2012. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA, *GIScience & Remote Sensing*, 49(5): 623-643.

Gislason, P.O., J.A. Benediktsson, and J.R. Sveinsson, 2006. Random Forests for land cover classification," *Pattern Recognition Letters*, 27: 294-300.

Guo, L., N. Chehata, C. Mallet, and S. Boukir, 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests, *ISPRS Journal of Photogrammetry and Remote Sensing*, 66: 56-66.

Hansen, M.C., and B. Reed, 2000. A comparison of the IGBP DISCover and University of Maryland 1 km global land-cover products, *International Journal of Remote Sensing*, 21(6-7): 1365-1373.

Hansen, M., R. Dubayah, and R. DeFries, 1996. Classification trees: an alternative to traditional land cover classifiers, *International Journal of Remote Sensing*, 17(5): 1075-1081.

Hooke, R.L., 1994. On the efficacy of humans as geomorphic agents, *GSA Today*, 4(9): 217, 224-225.

Hooke, R.L., 1999. Spatial distribution of human geomorphic activity in the United States: Comparison to Rivers. *Earth Surface Processes and Landforms*. 24(8): 687-692.

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

Irons, J.R., and R.L. Kennard, 1986. The utility of Thematic Mapper sensor characteristics for surface mine monitoring, *Photogrammetric Engineering & Remote Sensing*, 52: 389-396.

Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features, *Proceedings of European Conference on Machine Learning*, 21-23 April 1998, Chemnitz, Germany, pp. 137-142.

Kaasalainen, S., E. Ahoka, E. Hyyppa, and J. Suomalainen, 2005. Study of surface brightness from backscattering laser intensity: Calibration of laser data. *IEEE Geoscience and Remote Sensing Letters*. 2(3): 255-259.

Kazar, S.A., and T.A. Warner, 2013. Assessment of carbon storage and biomass on minelands reclaimed to grassland environments using Landsat spectral indices, *Journal of Applied Remote Sensing*, 7(1): 073583. DOI: 10.1117/1.JRS.7.073583.

Keene, T. and J. Skousen, 2010. Mine spoil reclamation with switchgrass for biofuel production, *2010 National Meeting of the American Society for Mining and Reclamation*, 5-11 June 2010, Pittsburgh, Pennsylvania: pp. 1-16.

Lawrence, R., A. Bunn, S. Powell, and M. Zambon, 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, *Remote Sensing of Environment*, 90(3): 331-336.

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

Lillesand, T.M., R.W. Kiefer, and J.W. Chipman, 2008. *Remote Sensing and Image Interpretation*, 6th Ed., Wiley & Sons, Hoboken, New Jersey,756 p.

Lin, Y., J.P. Mills, 2010. Factors influencing pulse width of small footprint, full waveform airborne laser scanning data, *Photogrammetric Engineering and Remote Sensing*, 76: 49-59.

Loosvelt, L., J. Peters, H. Skriver, H. Lievens, F.M.B. Van Coillie, B. De Baets, and N.E.C. Verhoest, 2012. Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification, *International Journal of Applied Earth Observation and Geoinformation*, 19: 173-184.

McIver, D.K. and M.A. Friedl, 2002. Using prior probabilities in decision-tree classification of remotely sensed data, *Remote Sensing of Environment*," 81(2-3): 235-261.

Melgani, F. and L. Bruzzone, 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42: 1778–1790.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, 2012. e1071: Misc functions of the department of statistics (e1071), *R Package Version 1.6-1*, URL: http://CRAN.R-project.org/package=e1071 (last date accessed: 1 June 2013).

Muchoney, D., J. Borak, H. Chi, M. Friedl, , S. Gopal, J. Hodges, N. Morrow, and A. Strahler, 2000. Applications of the MODIS global supervised classification model to vegetation and land cover mapping of Central America, *International Journal of Remote Sensing*, 21(6-7): 1115-1138.

Na, X., S. Zhang, X. Li, , H. Yu, and C. Liu, 2010. Improved land cover mapping using random forests combined with Landsat Thematic Mapper imagery and ancillary geographic data, *Photogrammetric Engineering & Remote Sensing*, 76(7): 833-840.

Negley, T.L., 2002. *A Comparative Hydrologic Analysis of Surface-Mined and Forested Watersheds in western Maryland*, M.S. Thesis, University of Maryland, College Park, Maryland, 84 p.

Optech, 2008. *3100 EA: Enhanced Accuracy*. URL: http://www.optech.ca/pdf/Brochures/ALTM3100EAwspecsfnl.pdf (last date accessed: 28 August 2013).

Pal, M., 2005. Random forest classifiers for remote sensing classification, *International Journal of Remote Sensing*, 26(1): 217-222.

Pal, M, and P.M. Mather, 2005. Support vector machines for classification in remote sensing, *International Journal of Remote Sensing*, 5(10): 1007-1011.

Palmer, M.A., E.S. Bernhardt, W.H. Schlesinger, K.N. Eshleman, E. Fourfoula-Georgiou, M.S. Hendryx, A.D. Lemly, G.E. Likens, O.L. Loucks, M.E. Power, P.S. White, and P.R. Wilcock, 2010. Mountaintop mining consequences, *Science*, 327: 148-149.

Parks, N.F., G.W. Petersen, and G.M. Baumer, 1987. High-resolution remote sensing of spatially and spectrally complex coal surface mines of Central Pennsylvania: A comparison between SPOT, MSS, and Landsat-5 Thematic Mapper, *Photogrammetric Engineering & Remote Sensing*, 53: 415-420.

Pontius, R.G., Jr., and M. Millones, 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment, *International Journal of Remote Sensing*, 32(15): 4407-4429.

Prakash, A., and R.P. Gupta, 1998. Land-use mapping and change in a coal mining area: A case study in the Jharia coalfields, India, *International Journal of Remote Sensing*, 19: 391-410.

R Core Team, 2012. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL: http://www.R-project.org/ (last date accessed: 1 June 2013)

RAMPP, 2011. *RAMPP West Virginia LiDAR QA/QC: Coal River Delivery*. URL: http://www.wvview.org/data/lidar/SouthernWV/Metadata%20and%20Tiling%20Scheme/RAMPP_WV_Memo_20110722_Final.pdf (last date accessed: 28 August 2013)

RapidEye, 2009. *RapidEye Standard Image Product Specifications*, RapidEye AG, Berlin, Germany, 47 p.

Rathore, C.S., and R. Wright, 1993. Monitoring environmental impacts of surface coal-mining, *International Journal of Remote Sensing*, 14: 1021-1042.

Rodríguez-Galiano, V.F., F. Abarca-Hernández, B. Ghimire, M. Chica-Olmo, P.M. Atkinson, and C. Jeganathan, 2011. Incorporating spatial variability measures in land-cover classification using random forests, *Procedia Environmental Sciences*, 3: 44-49.

Rodríguez-Galiano, V.F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J.P. Rigol-Sanchez, 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification, *ISPRS Journal of Photogrammetry and Remote Sensing*, 67: 93-104.

Rogan, J., J. Miller, D. Stow, J. Franklin, L. Levien, and C. Fischer, 2003. Land-cover change monitoring with classification trees using Landsat TM and ancillary data, *Photogrammetric Engineering & Remote Sensing*, 69(7): 793-804.

Saylor, K.L., 2008. Land Cover Trend Project: Central Appalachians, U.S. Department of the Interior, U.S. Geological Survey. Washington, DC, URL: http://landcovertrends.usgs.gov/east/eco69 Report.html (last date accessed: 1 June 2013).

Sen, S., C.E. Zipper, R.H. Wynne, and P.F., Donovan, 2012. Identifying revegetated mines as disturbance/recovery trajectories using an interannual Landsat chronosequence, *Photogrammetric Engineering & Remote Sensing*, 78(3): 223-235.

Simmons, J.A., W.S. Currie, K.N. Eshleman, K. Kuers, S. Monteleone, T.L. Negley, B.R. Pohlad, and C.L. Thomas, 2008. Forest to reclaimed mine land use change leads to altered ecosystem structure and function, *Ecological Applications*, 18(1): 104-118.

Steele, B.M., 2000. Combining multiple classifiers: An application using spatial and remotely sensed information for land cover mapping, *Remote Sensing of Environment*, 74(3): 545-556.

Stehman, S.V., and G.M. Foody, 2009. Accuracy assessment, *The SAGE Handbook of Remote Sensing* (T.A. Warner, M.D., Nellis, and G.M. Foody, editors), Sage Publications Ltd., London, United Kingdom, pp. 297-309.

Strausbaugh, P.D. and E.L. Core, 1997. *Flora of West Virginia*. Seneca Books, Morgantown, West Virginia, 1079 p.

Townsend, P.A., D.P. Helmers, C.C. Kingdon, B.E. McNeil, K.M. de Beurs, and K.N. Eshleman, 2009. Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976-2006 Landsat time series, *Remote Sensing of Environment*, 113: 62-72.

Townshend, J.R.G., 1981. The spatial resolving power of Earth resources satellites, *Progress in Physical Geography*, 5: 32-55.

Townshend, J.R.G., 1992. Land cover, *International Journal of Remote Sensing*, 13: 1319-1328.

Tyc, G., J. Tulip, D. Schulten, M. Krischke, and M. Oxfort, 2005. The RapidEye mission design. *Acta Astronautica*, 56: 213-219.

Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, New York,188p.

Warner, T.A., and F. Nerry, 2009. Does a single broadband or multispectral thermal data add information for classification of visible, near- and shortwave infrared imagery of urban areas?, *International Journal of Remote Sensing*, 30(9): 2155-2171.

Warner, T.A., M.D. Nellis, and G.M. Foody, 2009. Remote sensing data selection issues, *The SAGE Handbook of Remote Sensing* (T.A. Warner, M.D., Nellis, and G.M. Foody, editors), Sage Publications Ltd., London, United Kingdom, pp. 4-17.

Waske, B., and M. Braun, 2009. Classifier ensembles for land cover mapping using multiemporal SAR imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, 64: 450-457.

Wickham, J.D., K.H. Riitters, T.G. Wade, M. Coan, and C. Homer, 2007. The effect of Appalachian mountaintop mining on interior forest, *Landscape Ecology*, 22: 179-187.

WVGES (West Virginia Geologic and Economic Survey), 2005. Physiographic provinces of West Virginia, WVGES, URL: htt://www.wvgs.wvnet.edu/www/maps/pprovinces.htm (last date accessed: 1 June 2013).

Yuill, C., 2003. Landscape use assessment: Mountaintop mining and the mountaintop mining region of West Virginia, *Draft Programmatic Environmental Impact Statement on Mountaintop Mining/Valley Fills in Appalachia*, P. III, F-12.

Zipper, C.E., J.A. Burger, J.G. Skousen, P.N. Angle, D. Barton, V. Davis, and J.A. Franklin, 2011. Restoring forests and associated goods and services on Appalachian coal surface mines, *Environmental Management*, 47: 751-765.

Table 1: Land cover class definitions.

Class	Description
Forested	Land dominated by mature, woody vegetation that has not been directly disturbed by surface mining; mature forest that generally represents pre-mining conditions of the slopes
Reclaimed-herbaceous vegetation	Reclaimed areas dominated by herbaceous/non-woody vegetation
Reclaimed-woody vegetation	Reclaimed areas dominated by clumped or clustered woody plants that include shrubs and immature trees
Barren	Barren land lacking vegetation; manmade structures; haul roads; active quarries; lands disturbed by mining
Water	Water, including retention ponds, streams, and standing water

Table 2: Number of training polygons, area (ha) of training polygons, and number of randomly selected training pixels for each land cover class based on manual photograph interpretation of high resolution aerial orthophotography.

Land Cover Class	**Number of Training Polygons**	**Area of Training Polygons (ha)**	**Number of Training Pixels Randomly Selected From Within The Training Polygons**
Forested	81	77.0	1000
Reclaimed-herbaceous vegetation	121	37.9	1000
Reclaimed-woody vegetation	59	9.0	1000
Barren	106	44.7	1000
Water	22	1.3	517

Table 3: Optimized parameter settings for SVM, RF, and boosted CART.

Parameters Selected			
Band Combinations	**SVM (γ, C)**	**RF (m, k)**	**Boosted CART (k)**
Imagery	1.5, 4	3, 500	500
Imagery + LiDAR	0.14, 35	2, 500	500
LiDAR	1.2, 105	2, 500	500

Table 4: Performance of the different classifiers and variable combinations; overall accuracy (OA), Kappa (K), allocation disagreement (AD), and quantity disagreement (QD), respectively.

Band Combinations (Number of Bands)	**Measures**	**SVM**	**RF**	**Boosted CART**
Imagery (5)	OA (%)	80.6	77.4	77.6
	K (%)	73.4	69.2	79.6
	AD (%)	11.4	13.0	12.8
	QD (%)	8.0	9.6	9.6
Imagery + LiDAR (8)	OA (%)	86.4	84.1	83.5
	K (%)	81.2	78.1	77.2
	AD (%)	7.6	8.8	9.2
	QD (%)	6.0	7.1	7.3
LiDAR (3)	OA (%)	76.1	75.9	75.6
	K (%)	67.3	67.0	66.8
	AD (%)	18.2	18.0	17.7
	QD (%)	5.7	6.1	6.7
OA = Overall Accuracy K = Kappa Statistic AD = Allocation Disagreement QD = Quantity Disagreement				

Table 5: Area (ha) and percentage of land cover classes for the surface mine complex shown in Figure 2.

Land Cover Class	**Area (ha)**	**Percentage of Permitted Areas**
Forested	1,132	20.5%
Reclaimed-herbaceous vegetation	2,315	41.7%
Reclaimed-woody vegetation	854	15.4%
Barren	1,201	21.6%
Water	48	0.9%

Table 6: McNemar's test to assess statistical differences between classifications produced given different input predictor variables and classification algorithms. A *z*-score larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is better than the other.

a) Comparison of Input Variables Classified Using SVM		
	LiDAR	Imagery + LiDAR
Imagery	2.105*	4.969*
LiDAR		6.074*
b) Comparison of Input Variables Classified Using RF		
	LiDAR	Imagery + LiDAR
Imagery	0.810	5.802*
LiDAR		5.800*
c) Comparison of Input Variables Classified Using Boosted CART		
	LiDAR	Imagery + LiDAR
Imagery	1.535	3.467*
LiDAR		4.811*
d) Comparison of Classification Method Using Image Data		
	RF	Boosted CART
SVM	4.303*	2.466*
RF		1.539
e) Comparison of Classification Method Using LiDAR Data		
	RF	Boosted CART
SVM	1.627	1.360
RF		0.098
f) Comparison of Classification Method Using Image and LiDAR Data		
	RF	Boosted CART
SVM	2.621*	3.464*
RF		1.287

Table 7: Error matrix for SVM with imagery bands only. Overall accuracy is 80.6%.

		Reference Data (Pixels)						
		Forested	Reclaimed herbaceous vegetation	Reclaimed woody vegetation	Barren	Water	Totals	User's Accuracy
Classified Data (Pixels)	Forested	238	35	42	2	21	338	70.4%
	Reclaimed herbaceous vegetation	14	193	20	6	0	233	82.8%
	Reclaimed woody vegetation	13	20	203	0	0	236	86.0%
	Barren	0	16	0	247	40	303	81.5%
	Water	0	1	0	10	204	215	94.9%
	Totals	265	265	265	265	265		
	Producer's Accuracy	89.8%	72.8%	76.6%	93.2%	77.0%		

Table 8: Error matrix for SVM with LiDAR-derivatives only. Overall accuracy is 76.1%.

		Reference Data (Pixels)						
		Forested	Reclaimed herbaceous vegetation	Reclaimed woody vegetation	Barren	Water	Totals	User's Accuracy
Classified Data (Pixels)	Forested	217	10	39	2	6	274	79.2%
	Reclaimed herbaceous vegetation	8	192	19	34	1	254	75.6%
	Reclaimed woody vegetation	38	21	204	1	3	267	76.4%
	Barren	2	41	3	210	31	287	73.2%
	Water	0	1	0	18	224	243	92.2%
	Totals	265	265	265	265	265		
	Producer's Accuracy	81.9%	72.5%	77.0%	79.2%	84.5%		

Table 9: Error matrix for SVM with imagery bands and LiDAR-derivatives. Overall accuracy is 86.4%.

		Reference Data (Pixels)						
		Forested	Reclaimed herbaceous vegetation	Reclaimed woody vegetation	Barren	Water	Totals	User's Accuracy
Classified Data (Pixels)	Forested	253	11	24	1	7	296	85.5%
	Reclaimed herbaceous vegetation	5	210	11	5	0	231	90.9%
	Reclaimed woody vegetation	6	23	230	0	0	259	88.8%
	Barren	1	21	0	248	55	325	76.3%
	Water	0	0	0	11	203	214	94.9%
	Totals	265	265	265	265	265		
	Producer's Accuracy	95.5%	79.2%	86.8%	93.6%	76.6%		

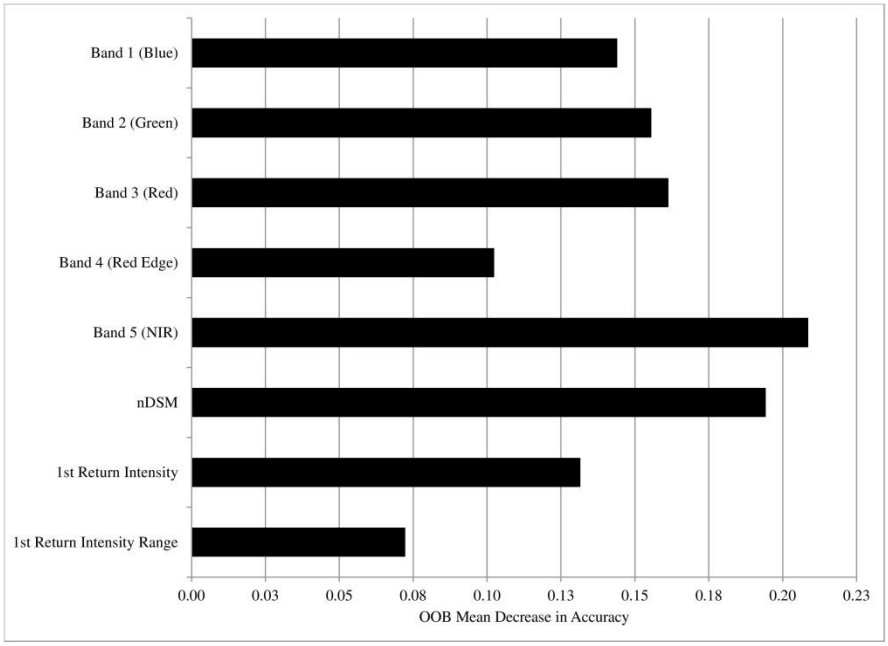

Figure 1: Relative importance of predictor variables as estimated by the out-of-bag (oob) mean decrease in accuracy by RF. This analysis suggests that the most important predictor variables in the classification were RapidEye Band 5 (NIR), followed by the LiDAR-derived nDSM.

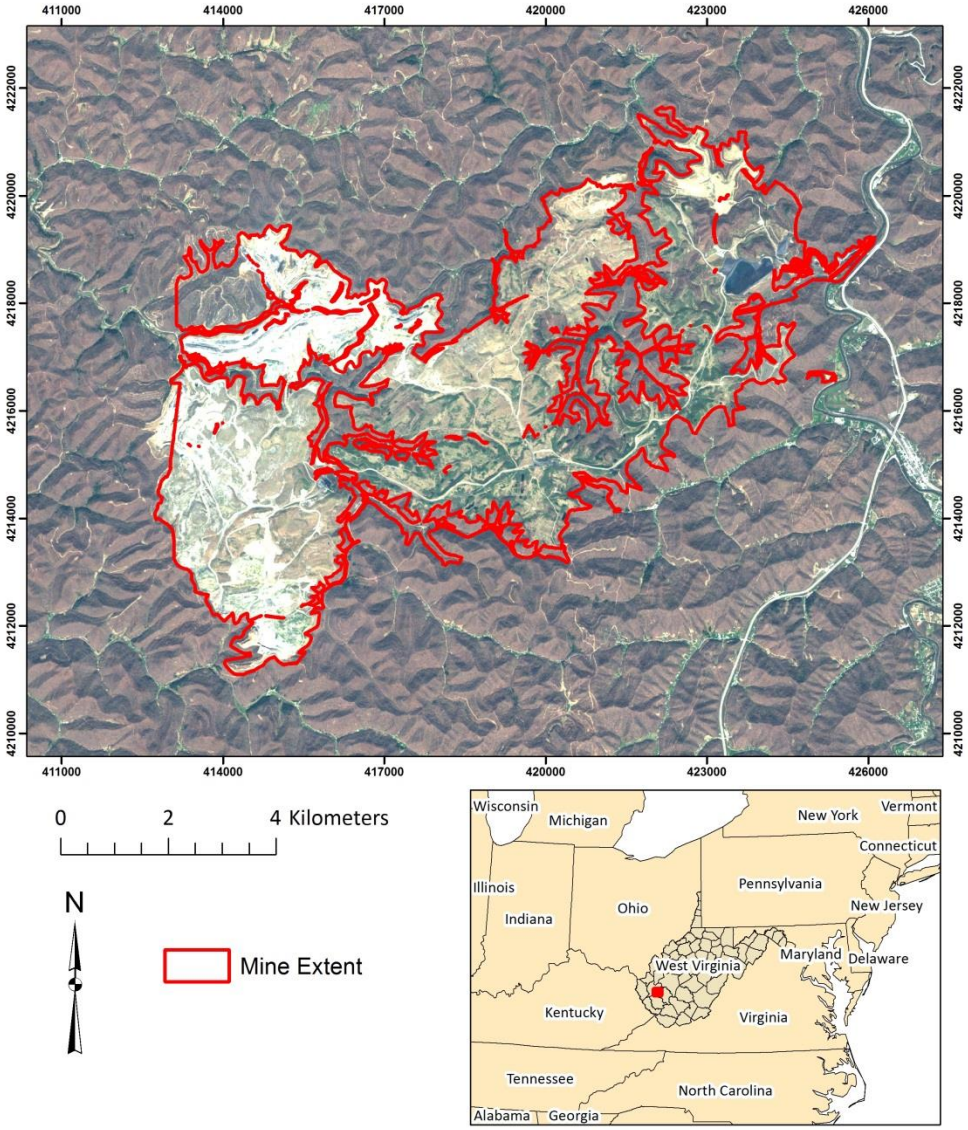

Plate 1: Hobet-21 surface mine complex. Base image is the RapidEye scene acquired on April 1, 2010 and displayed in simulated natural color (Bands 3, 2, 1 as RGB). The depicted mine extent is based on the surface mining permit obtained from WVDEP. The map is projected in NAD83 UTM Zone 17 N.

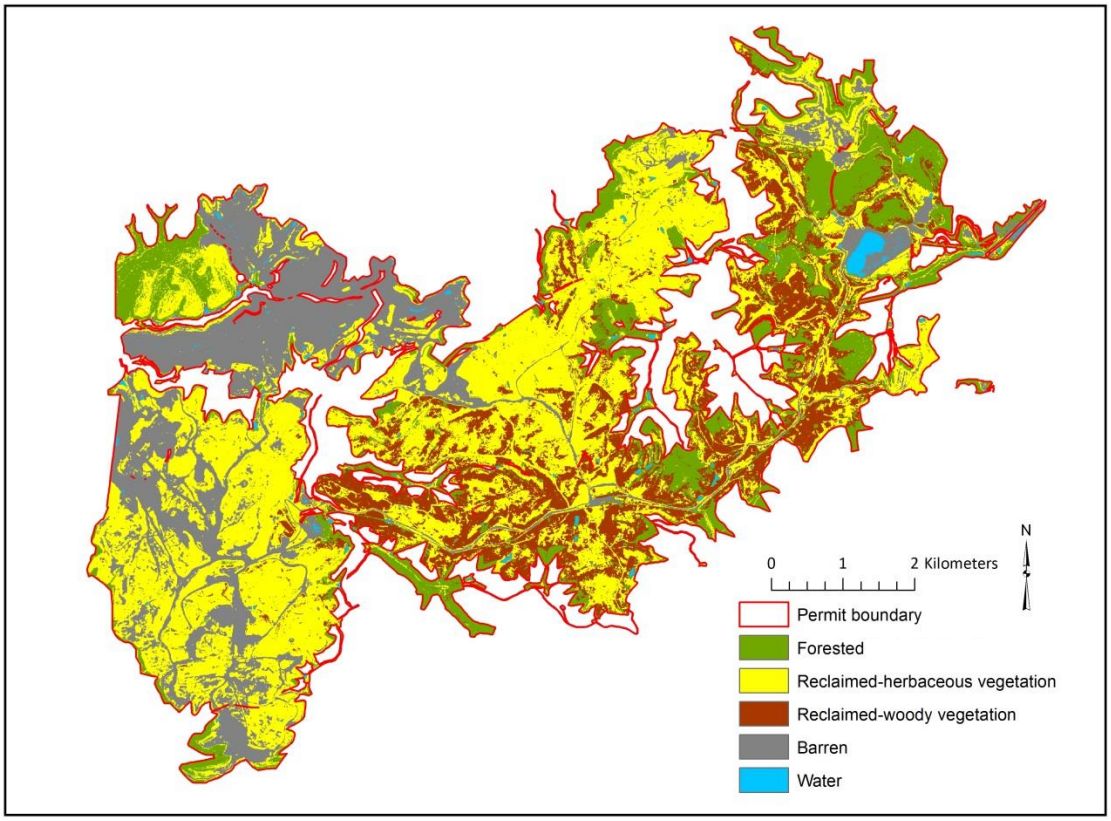

Plate 2: Land cover classification of the Hobet-21 mine complex using all RapidEye imagery bands and LiDAR-derivatives. SVM was used with a radial basis function (RBF) kernel, a gamma value (γ) of 0.14, and a cost value (C) of 35.

CHAPTER 3

Comparison of NAIP orthophotography and RapidEye satellite imagery for mapping of mining and mine reclamation[1]

Aaron E. Maxwell, Michael P. Strager, Timothy A. Warner, Nicholas P. Zégre, and Charlie Yuill

Abstract

National Agriculture Imagery Program (NAIP) orthophotography is a potentially useful data source for land cover classification in the United States due to its nation-wide and generally cloud-free coverage, low cost to the public, frequent update interval, and high spatial resolution. Nevertheless, there are challenges with working with NAIP imagery, especially regarding varying viewing geometry, radiometric normalization, and calibration. In this paper we compare NAIP orthophotography and RapidEye satellite imagery for high resolution mapping of mining and mine reclamation within a mountaintop coal surface mine in the southern coalfields of West Virginia, USA. Two classification algorithms, support vector machines (SVM) and random forests (RF), were used to classify both data sets. Compared to the RapidEye classification, the NAIP classification resulted in lower overall accuracy and Kappa, and higher allocation disagreement and quantity disagreement. However, accuracy of the NAIP classification was improved by reducing the number of classes mapped, using the near infrared (NIR) band, using textural measures and feature selection, and reducing the spatial

[1] This is an Accepted Manuscript of an article published by Taylor & Francis in *GIScience and Remote Sensing* on 09 May 2014, available online: http://wwww.tandfonline.com/doi/abs/10.1080/15481603.2014.912874. Maxwell, A.E., M.P. Strager, T.A. Warner, N.P. Zégre, and C.B. Yuill, 2014. Comparison of NAIP orthophotography and RapidEye satellite imagery for mapping of mining and mine reclamation, *GIScience & Remote Sensing*, 51(3): 301-320. (Received 20 December 2013, Accepted 26 March 2014)

resolution slightly by pixel aggregation or by applying a Gaussian low pass filter. With such strategies, NAIP data can be a potential alternative to RapidEye satellite data for classification of surface mine land cover.

Keywords: National Agriculture Imagery Program; NAIP; RapidEye; Mountaintop removal coal mining; surface coal mining; land cover classification; machine learning

Introduction

The United States Department of Agriculture (USDA) National Agriculture Imagery Program (NAIP) orthophotography offers high spatial resolution (1 m ground sampling distance (GSD)), up to four spectral bands, nearly cloud-free coverage of large areas (e.g. entire states) acquired during a single growing season, and low cost to the user. For the state of West Virginia it is available for the growing seasons of 2007, 2009, and 2011 (Aerial Photography Field Office, 2012), potentially allowing multi-temporal analysis. Because of these characteristics, it is compelling to consider these data for land cover mapping and monitoring. However, prior research has found that high spatial resolution, variable viewing geometry and illumination, and in some cases limited spectral resolution, can complicate mapping tasks from aerial data (Cushnie, 1987; Myeong et al., 2001; Bozheva et al., 2005; Guo et al., 2007; Baker et al., 2013). In comparison, high-spatial resolution commercial satellite images can potentially offer large scenes with relatively consistent viewing and illumination geometry and four to eight well-characterized spectral bands (Toutin, 2009).

This research compares the use of NAIP orthophotography and RapidEye satellite imagery for the mapping of mining and mine reclamation within an individual mine complex in the southern coalfields of the eastern United States. RapidEye satellite data offers five spectral bands and a 5 m cell size (Tyc et al., 2005). RapidEye data were used, as opposed to IKONOS or other high resolution satellite data, because cloud free coverage was available that was acquired at a time nearly coincident with the NAIP collection, because the RapidEye constellation of five satellites offers a frequent return time (5.5 days at nadir) (Tyc et al., 2005), and because the researchers have previously used this imagery for mapping surface mine land cover (Maxwell et al., 2014) with promising results. The following factors were investigated:

1. Comparative classification accuracy for five land cover classes under various processing scenarios (at the original cell size of each data set, at a common cell size, at a common radiometric resolution, and at a common cell size and radiometric resolution).
2. The relative importance of different image bands in the random forests (RF) classification.
3. The relative importance of multi-seasonal data.
4. The relative importance of the near-infrared (NIR) band for NAIP classification.
5. The relative improvement in classification accuracy as the number of classes is reduced.
6. Improvement in classification accuracy of NAIP with the incorporation of texture measures from the gray level co-occurrence matrix (GLCM).
7. Improvement in classification accuracy of NAIP resulting from image smoothing using a Gaussian low pass filter.

Background

In recent years, data selection for fine-scale mapping has become more complex due to the wide range of choices of both aerial- and satellite-based high spatial resolution data. Determining which sensors offers the appropriate spatial, spectral, temporal, and radiometric scales to meet specific mapping needs may not be a simple task as multiple criteria must be considered and evaluated (Phinn, 1998). Variations in the cost of data further complicate data selection (Tarnavsky et al., 2004; Warner et al., 2009).

The complexity of land cover mapping from high resolution aerial imagery has been previously explored. For example, Myeong et al. (2001) noted that shadows and similarity in spectral response between mapping classes complicated the mapping of urban cover from high

resolution (0.61 m pixel size) color-infrared imagery. There is also a growing literature on the use of NAIP orthophotography for land cover mapping. For example, Davies et al. (2010) used NAIP to map juniper cover in Idaho, USA, while Meneguzzo et al. (2013) attempted to classify isolated trees in Steele County, Minnesota, USA. Baker et al. (2013) compared pixel-based and object-based classification methods for mapping forest clearings associated with natural gas drilling operations in Greene County, Pennsylvania, USA. They noted the challenges of using NAIP data for image classification, including issues related to the range in time of day and season of acquisition, which resulted in differences in digital number (DN) values between tiles, differences in shadow length and direction, and phenological differences. A search of the current literature suggests that there is little work published on comparison of NAIP and high resolution satellite data or NAIP for the classification of mining and mine reclamation.

Mountaintop mining occurs in southern West Virginia, eastern Kentucky, eastern Tennessee, and southwestern Virginia, a region known as the southern coalfields of the eastern United States (USEPA, 2005). In this region, mountaintop coal mining is the leading cause of land use/land cover change (Saylor, 2008; Townsend et al., 2009; Drummond and Loveland, 2010). Coal extraction results in forest clearing (for example, 420,000 ha in Appalachia between 1973 and 2000 (USEPA, 2005)) and fragmentation (Wickham et al., 2007), removal of top soil, and a recontouring of the landscape (Palmer et al., 2010; Bernhardt and Palmer, 2011). Furthermore, mountaintop mining causes faster landscape alteration than more traditional mining methods such as auger, contour, and highwall mining (Fritz et al., 2010). After mining is complete, the disturbed areas are reclaimed, commonly to grasslands or shrublands (Simmons et al., 2008; Kazar and Warner, 2013). More recently, there has been increased interest in reclamation to native forest species on the topographically altered terrain (Zipper et al., 2011).

Remote sensing has been investigated as a means to map landscape change due to mining and mine reclamation using data from the Landsat Multispectral Scanner (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and SPOT (Rathore and Wright, 1993; Townsend et al., 2009; Sen et al., 2012; Zégre et al., 2013). Mapping mine reclamation is of specific concern as this is commonly the legacy imprint of mining on the landscape (Negley, 2002), although Rathore and Wright (1993) found that mine reclamation was mapped with lower accuracy than active mining. Recently, there has been an interest in multi-temporal imagery for separating mining and non-mining as a way to overcome the spectral similarities of reclaimed and undisturbed landscapes such as pastureland or herbaceous cover (Townsend et al., 2009; Sen et al., 2012). Textural measures have been investigated as a means to increase classification accuracy using single-date data (Warner, 2011), and such variables have been shown to improve land cover classification in multiple studies (for example, Chica-Olmo and Abarca-Hernández, 2000; Agüera et al., 2008; Ghimire et al., 2010; Rodríguez-Galiano et al., 2012b).

Machine learning algorithms such as support vector machines (SVM) and RF offer particular advantages for classifying high spatial resolution imagery of mined landscapes. Machine learning algorithms have been shown generally to be more accurate and efficient than parametric classifiers, which make assumptions regarding the data distribution (Hansen et al., 1996; Huang et al., 2002; Rogan et al., 2003; Pal, 2005; Ghimire et al., 2012; Loosvelt et al., 2012). Pal (2005) suggests that SVM and RF can provide comparable map accuracies, though RF is simpler to apply. Previous work by the authors (Maxwell et al., 2014) suggests that SVM provided a more accurate classification than RF for classification of surface mine land cover from high resolution satellite imagery. In addition, optimizing the SVM algorithm, often the most time consuming step in executing the algorithm, did not statistically increase the accuracy

of the classification, suggesting that the method is relatively robust (Maxwell et al., 2014), although other studies have suggested that optimization may be of importance (Pal, 2005).

SVMs make use of structural risk minimization to find the hyperplane that separates two classes with the maximum margin; it is an implementation of statistical learning theory and optimization algorithms to locate decision boundaries between data classes (Vapnik, 1995; Burges, 1998; Pal and Mather, 2005; Pal, 2005; Su and Huang, 2009). The points that lie closet to the hyperplane, termed "support vectors," define the margin. Because SVM algorithms are designed for two-class problems only, strategies have been developed incorporating multiple SVM algorithms to generate a multi-class classification (Vapnik, 1995; Pal and Mather, 2005; Pal, 2005). The e1071 package in R, which is the implementation of SVM used in this research, uses a "one-against-one" approach for multiclass-classification in which binary classifiers are trained and the appropriate class is found by a voting scheme (Meyer et al., 2012).

RF, introduced by Breiman (2001), uses multiple decision trees as a means to improve the accuracy and consistency of single tree classifications. RF functions as an ensemble of decision trees and is a non-parametric learning algorithm. Bagging, which comprises bootstrap sampling with replacement (Breiman 1996), is used to generate a subset of the data to train each tree instead of using the entire training data set. The remaining data, termed out-of-bag (oob) data, are withheld and can be used for accuracy assessment. Instead of using all predictor variables in each tree, RF uses a random subset of the predictor variables (the number of which is defined by the user) to grow each tree of the ensemble. Although the use of a subset of variables decreases the classification accuracy of any one tree, correlation between trees is also reduced, resulting in a reduction in generalization error, which is the strength of RF. The Gini index, which provides a measure of impurity of a given class with respect to the rest of the classes, is

used to select the best predictor among the randomly selected predictor variables available at each node (Breiman, 2001; Ghimire et al., 2010; Rodríguez-Galiano et al., 2011; Ghimire et al., 2012; Rodríguez-Galiano et al., 2012a; Rodríguez-Galiano et al., 2012b).

Study area

The study area is a single large mine complex, the Hobet-21 mountaintop mine in Boone and Lincoln counties, West Virginia, USA (Figure 1). It is the largest surface mine complex in the Appalachian region (Keene and Skousen, 2010), with a wide variety in age of disturbance, vegetation, and land cover. Historical imagery shows that some of the mining disturbance predates 1987, while portions of the mine were still active at the time of this study. Multiple remotely sensed data sets are available for this site, including NAIP orthophotography, light detection and ranging (LiDAR) data, multiple RapidEye scenes, and high resolution imagery from Pictometry International Corporation (Rochester, New York). Multiple data sources allowed the collection of accurate training and assessment data, making this an excellent location to study surface mine mapping. A total area of 5,500 ha was classified within the extent of the Hobet-21 mine.

Methods

Data

The characteristics of the NAIP and RapidEye data used are described in Table 1. The RapidEye data were provided by the supplier as an orthorectified product (termed 3A), which has a nominal error of potentially less than 1 pixel under ideal circumstances (e.g. nadir view and flat terrain) (RapidEye, 2009). The primary RapidEye data over the study site were acquired at a view angle of 6.72˚ (i.e. close to nadir), but the topography of the study site is complex, therefore the error may be greater than 1 pixel. Nevertheless, a visual inspection of an overlay of the two image data sets did not show noticeable differences, suggesting that the registration was adequate

for the analysis and not likely to affect the classification. However, as a precaution in case of co-registration problems or potential changes in land cover between the data set acquisitions, each pixel in the training and assessment data sets was checked to ensure that the assigned land cover class for that location was the same in each data set. Any samples that failed this test (less than 5% of all samples) were excluded from the analysis. In this study, single-date imagery in which the training, assessment, and image data were on the same scale were used; as a result, as suggested by Song et al. (2001), DN values were used as delivered, not converted to reflectance or with atmospheric corrections applied.

The NAIP orthophotography was collected with an Integraph Z/I Imaging Digital Mapping Camera (DMC) on 14 July, 2011, and thus leaf-on conditions during the growing season were captured. This imagery has a 1 m GSD and four spectral bands (blue, green, red, and NIR) as described in Table 1. The data were provided by the Aerial Photograph Field Office (APFO) of the USDA Farm Service Agency (FSA) as uncompressed quarter quadrangles. These tiles were mosaicked to produce a single image covering the mine.

The RapidEye data were collected on 1 August, 2011. The scene has an average center azimuth angle of 279.6˚, sun azimuth of 165.1˚, and sun elevation of 69.6˚. The satellites have sensors with five spectral bands, as described in Table 1 (Tyc et al., 2005). The GSD of the system is 6.5 m. The 3A product used in this study has radiometric, sensor, and geometric corrections applied, and is orthorectified to a 5 m grid. The Hobet-21 mine complex is covered by a single image tile so there was no need to mosaic multiple tiles.

A second RapidEye image was also used in the analysis in order to assess the benefit of multi-season data. The image was collected on 1 April 2010, prior to spring leaf out. The scene has an average center image view angle of -2.82˚, azimuth angle of 110.2˚, sun azimuth of

171.2˚, and sun elevation of 56.5˚. Although the multi-season data were combined to assess the potential benefits of multi-season data, which is not possible with NAIP, the primary objective of this study was to compare leaf-on, single-date imagery.

Image processing

Resampling was necessary in order to compare the images at a common cell size. As the RapidEye data are collected at a 6.5 m resolution then subsequently resampled to 5 m, the coarsest resolution of the two data sets, 6.5 m, was used as the common cell size for analysis. Due to the much smaller original cell size of the NAIP imagery (1 m), it was resampled using pixel aggregation to 6.5 m, while the RapidEye data were resampled using cubic convolution. As cubic convolution resampling may smooth the image, nearest neighbor resampling was also tested, and no statistical difference in classification was noted between the resulting classifications using the two resampling methods. As a result, cubic convolution was used in the subsequent analyses. Also, each image was converted to an 8-bit stretch using a linear rescaling in order to allow for comparison of the images at a common radiometric scale.

GLCM texture measures (Haralick et al., 1973) were calculated from the NAIP orthophotography. To facilitate comparison between the data sets, the texture calculations were applied to the imagery at the common 6.5 m pixel size. The following textural bands were calculated: variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation. The mean was not included as the impact of smoothing the data was assessed separately. The textural measures were calculated from the red and green bands and also the first and second principle component bands. In order to assess texture at different spatial scales, three different offsets, 3, 2, and 1 pixels, were tested. The kernel size, which determines the amount of smoothing (Warner, 2011), was set to be close to the minimum possible given the offset distance.

The kernel size was set at 11 x 11 for the 3 pixel offset, 7 x 7 for the 2 pixel offset, and 5 x 5 for the offset of 1. Texture was calculated in the vertical, horizontal, and diagonal directions and then averaged in order to produce a composite, non-directionally-sensitive value (Warner, 2011). A total of 84 textural measures were produced.

A visual comparison of the 6.5 m NAIP and RapidEye data suggested that the latter image has a smoother texture (i.e. more autocorrelation), even though both images had the same nominal pixel size. Therefore, as an experiment, the 6.5 m NAIP was smoothed with a Gaussian low pass filter with kernel sizes of 3 x 3, 5 x 5, and 7 x 7 pixels. Although applying a simple Gaussian low pass filter does not replicate the modulation transfer function (MTF) (Myneni et al., 1995) of the RapidEye sensor, this method provides a simple means to assess whether or not blurring the image (i.e. reducing the effective spatial resolution) would improve the classification accuracy of the NAIP data.

In summary, the NAIP imagery provided four bands, or predictor variables, for the classification. These four bands were used in the classification algorithms at the original cell size (1 m), resampled to 6.5 m, and also smoothed using the Gaussian filter. After calculating the textural measures, 88 predictor variables were available from NAIP (the four image bands and 84 textural measures). RapidEye provided five image bands, which were used in the classification algorithms at the original cell size, resampled to 6.5 m, and reduced to an 8-bit radiometric range. Combining the primary, leaf-on RapidEye data with the leaf-off scene resulted in a total of ten predictor variables available to classify land cover.

Land cover classes

Five land cover classes were mapped within the surface mine: forested, reclaimed-herbaceous vegetation, reclaimed-woody vegetation, barren, and water (Table 2). It should be

noted that the term "forested land cover" is used to indicate forests in the permit area not yet removed for the mountaintop removal mining. The term does not necessarily imply that the area has never been disturbed at all; West Virginia's landscape records extensive historical disturbances ranging from clear-cutting to prior surface coal mining. In order to assess potential increases in classification accuracy when the number of classes is reduced, the data were also classified with the forested and reclaimed-woody vegetation classes combined, resulting in a four-class problem: woody vegetation, herbaceous vegetation, barren, and water. In addition, the data were also classified with the vegetation classes combined, resulting in just three classes: vegetated, barren, and water.

Training data

Training data polygons were delineated by manual interpretation of multiple sources including the following: NAIP orthophotography, RapidEye imagery (both leaf-on and leaf-off), 1 m resolution Pictometry orthphotography, and LiDAR-derived topographic slope, first return intensity, and normalized digital surface model (nDSM) raster grids. As this data spans the period from March 2010 (Pictometry orthphotography) to August 2011 (most recent RapidEye image), there was potential for landscape change during the data collection period due to mine expansion or reclamation activities. Therefore, the training data were screened for potential change, and only samples where the land cover appeared to be the same in all data sets were used. The number of training polygons collected is summarized in Table 3.

The software tool Geospatial Modeling Environment (GME) (Beyer, 2012) was used to draw a random sample of training pixels from the polygons; 1,000 pixel examples of each class were randomly selected. Because water is not a common cover type in the mine complex, it was

not possible to collected 1,000 examples for that class. In total, 4,386 pixels were utilized to train the models as summarized in Table 3.

Classification methods

In order to generalize the results of this study, two different per-pixel classifiers were utilized, SVM and RF, as implemented within the statistical software tool R (R Development Core Team, 2012), using the e1071 package (Meyer et al., 2012) and the randomForest package (Liaw and Wiener, 2002), respectively.

Optimization of the parameters for the SVM and RF algorithms was carried out using a ten-fold cross validation grid search approach using the e1017 package in R (Meyer et al., 2012). This tuning function partitions the data into ten disjoint training sets, using a random assignment. The classifier is then trained ten times, and each time the remaining 10% of the data not used for training in that instance is used for validation to test for optimal parameter settings. Overall accuracy is calculated as the average across the ten cross validations. This function offers a means to test parameter combinations relative to how well withheld training data are classified. Once a coarse grid search using a wide range of parameter values was performed, a second grid search of values centered on the initial optimal value, and with a narrow range, was performed in order to further refine the estimate. Optimization was performed for each classification separately to ensure optimal parameters for the particular set of predictor variables tested. For SVM a radial basis function (RBF) kernel was used, and it was necessary to optimize the cost (C) and gamma (γ) parameters using the tuning function. For RF it was necessary to optimize the number of predictor variables randomly sampled as candidates at each node (m) and also the number of trees to grow (k).

Error assessment

Classification accuracy was assessed through an independent data set, which did not use any pixels from the training data or from within the training polygons from which they were selected. A stratified sample for the accuracy assessment was chosen because reclaimed-woody vegetation and water classes are not common cover types in the study area and thus would typically be only rarely selected in a purely random approach. A total of 1,300 pixels were randomly selected, 260 in each of the five land cover classes.

For each of the validation pixels, the dominant land cover category over the 6.5 x 6.5 m nominal area of the pixel was visually interpreted from the Pictometry, RapidEye, NAIP, and LiDAR data. As one goal of this work was to compare the statistical difference in map output, it was necessary to use the same validation data for all predictor variable combinations. Also, only validation points that were interpreted as the same cover type in all of the reference data sets were used in the analysis. This may have resulted in a slight inflation of map accuracy as mixed pixels may have been excluded.

From the randomly selected accuracy assessment data, error matrices and Kappa statistics were produced. Since stratified random sampling was implemented, it was necessary to calculate overall accuracy by weighting the class accuracies by the proportion of land cover within the study area (Stehman and Foody, 2009). Following Pontius and Millones (2011), quantity disagreement and allocation disagreement were also calculated. Quantity disagreement provides a measure of error in the proportions of the categories, while allocation disagreement provides a measure of error in the spatial allocation of the categories (Pontius and Millones, 2011). Allocation and quantity disagreement sum to overall error. Although the value of Kappa is questioned (Pontius and Millones, 2011), the statistic was nevertheless provided for potential comparisons to other studies.

In order to evaluate the statistical significance of any differences in the classifications, the results were compared on a pairwise basis using McNemar's test (Dietterich, 1998; Foody, 2004). First, 2-by-2 confusion matrices were produced that summarized the number of pixels correctly classified by both of the methods being compared, the number incorrectly classified by both methods, and the number classified correctly by one attempt but not the other. McNemar's test is a test of statistical difference that generates a *z*-score from this 2-by-2 matrix and is based on a chi-square test that compares the distribution of error expected under the null hypothesis (that the classifications have the same error rates) and the observed error. A *z*-score larger than 1.645 indicates a 95% confidence of statistical significance for the one-directional test of whether one classification is more accurate than the other (Bradley, 1968; Dietterich, 1998; Foody, 2004; Agresti, 2007).

Importance in random forest model

The RF algorithm generates an estimate of predictor variable importance during training by excluding each variable sequentially and recording the resulting increased oob error (Breiman, 2001; Rodríguez-Galiano et al., 2012a; Rodríguez-Galiano et al., 2012b). This feature of RF was used as a means to assess the importance of predictor variables, including raw image bands and textural measures. For example, the importance of the NAIP and RapidEye spectral bands were assessed in a model using both image data sets.

Results and discussion

Table 4 compares the NAIP and RapidEye data at the original cell sizes of each data set (i.e. 1 m for NAIP, 5 m for RapidEye), a common cell size (6.5 m), a common radiometric resolution (8-bit) and the original cell sizes, and a common cell size and radiometric resolution (6.5 m, 8-bit). At the original cell size, the RapidEye imagery was found to provide an overall accuracy for mapping the five classes 9.6% greater than the NAIP imagery using the SVM

algorithm and 10.6% greater using the RF algorithm, and both these differences were found to be statistically significant using McNemar's test (z-score (SVM) = 8.062 and z-score (RF) = 8.322). This higher accuracy for RapidEye data compared to NAIP is also reflected in Kappa, allocation disagreement, and quantity disagreement; the classification of the RapidEye data had lower allocation and quantity disagreement and a higher Kappa.

With a coarsening of the NAIP data to a 6.5 m pixel, accuracy increased by 2.3% for the SVM algorithm and by 3.4% for the RF algorithm. Both increases were shown to be statistically significant (z-score (SVM) = 2.668 and z-score (RF) = 3.621). Reduced spatial resolution associated with an increase in the cell size or with smoothing may result in a less complex and heterogeneous signatures for the map classes, producing increased classification accuracy (Latty et al., 1985; Warner et al., 2009). However, the NAIP 6.5 m data classification was still found to be less accurate than the RapidEye 6.5 m classification (Table 4) (z-score (SVM) = 5.852 and z-score (RF) = 6.041), though the z-scores were smaller than those for the comparison at the original cell sizes. Allocation and quantity disagreement were also reduced. This suggests that coarsening the NAIP data increased classification accuracy, but not to the level obtained from the RapidEye data. Baker et al. (2013) also found that coarsening NAIP improved classification accuracy for a pixel-based classification.

Table 4 suggests that radiometric resolution had little impact on the classification accuracy. For example, reducing the RapidEye data to an 8-bit radiometric range only decreased the accuracy by 0.3% using the SVM algorithm, which was not a statistically significant difference (z-score (SVM) = 1.347 and z-score (RF) = 0.714). Also, the RapidEye data on an 8-bit scale were still found to be statistically more accurate than the NAIP data at the original cell size and at a common cell size. This finding suggests that decreasing the radiometric resolution,

perhaps to reduce the file size, may not have a large impact on classification accuracy for this specific classification task.

Figure 2 shows the relative importance of predictor variables as estimated by the oob mean decrease in accuracy for a RF model using both the RapidEye and NAIP spectral bands at the original cell sizes. The analysis suggests that all of the RapidEye bands were more useful in the model than the NAIP bands. The greater importance of the RapidEye bands is likely due to a more consistent viewing geometry and illumination conditions for satellite data in comparison to an aerial image mosaic, perhaps resulting in a more consistent class signature throughout the extent of the mapped area. Figure 3 provides the same comparison of relative importance, except for the data at a common cell size (6.5 m). Once the NAIP data were coarsened, the NAIP bands take on a greater importance in the model, particularly the NAIP blue and NAIP NIR bands. This supports the finding discussed above: coarsening the data improves classification accuracy of NAIP orthophotography for this mapping task. As an improvement in classification was found when the NAIP data were coarsened to 6.5 m, the 6.5 m data were used for the rest of the analysis.

NAIP orthophotography is not always provided with a NIR band. For example, the 2009 collection for the state of West Virginia was only a true color collection. Therefore, the usefulness of the NIR band was tested here. Removing the NIR band from the analysis resulted in a 5.7% decrease in classification accuracy for the SVM algorithm and a 6.0% decrease in accuracy for the RF algorithm. The overall accuracy dropped below 80% for both the SVM and RF models, and both allocation and quantity disagreement increased. Removing the NIR band resulted in a statistically significant decrease in accuracy for both classifiers (z-score (SVM) =

7.993 and *z*-score (RF) = 8.459). This confirms the expected finding that there is merit in using the NIR band if NAIP data are to be used for classification tasks and NIR data are available.

As NAIP is acquired during the growing season, leaf-off data are not available, unlike satellite data where phenology can potentially be captured through multi-temporal imagery. To assess the merit of multi-season data, leaf-on and leaf-off RapidEye data were compared, and then combined for a classification using leaf-on and leaf-off bands. Leaf-on data provided a statistically more accurate classification than the leaf-off data using both the SVM and RF algorithms (*z*-score (SVM) = 3.273 and *z*-score (RF) = 4.020), however, the accuracy difference is much less than the difference between classification of NAIP and leaf-on RapidEye data. In comparison to the leaf-off data, classification of the leaf-on data resulted in only a 0.9% increase in accuracy using SVM and a 3.2% increase using RF, however, the differences were in both cases statistically significant (*z*-score (SVM) = 4.020 and *z*-score (RF) = 3.273). This suggests that leaf-on data are marginally more useful for this classification task. When the data were combined in a multi-season layer stack, the accuracy statistically improved compared to the leaf-on and leaf-off models, resulting in a 3.7% increase over the leaf-on data using the SVM algorithm and a 3.3% increase using the RF algorithm. Figure 4 indicates that both leaf-on and leaf-off bands were important to the RF model. In summary, multi-season data increased map accuracy for this classification task. This is of concern for NAIP, as multi-season data are not available.

The error matrix for the NAIP (6.5 m) classification using RF is shown in Table 5, and the error matrix for the RapidEye classification using RF is shown in Table 6. Four classification results are shown in Figure 5, using both the RF and SVM classifiers. The error trends are very similar for both the SVM and RF models, with the reclaimed-woody vegetation class having the

lowest producer's accuracy for both the NAIP and RapidEye classifications. This can be attributed to confusion with forest cover and somewhat to confusion with reclaimed-herbaceous cover. This is expected, as reclaimed-woody vegetation has similar spectral characteristics to both of these classes. For both image data sets, reclaimed-herbaceous vegetation was confused with barren cover; this confusion may be attributed to the patchy, heterogeneous nature of the vegetation in some reclaimed areas, which likely resulted in a mixed pixel problem. In the early stage of reclamation, in which vegetation is just beginning to regrow on the barren landscape, herbaceous cover is often present but can be sparse. In active surface mines where a wide range of disturbance and reclamation classes exist on a steep and heterogeneous landscape, the separation of these two classes is complex, and the boundary between them may be gradational as barren cover transitions to herbaceous cover.

For the NAIP classification, producer's accuracy for water was also generally low, likely a result of the variability in water reflectance on mine sites. Spectral reflectance may vary between small settlement ponds, water retention ponds, and treatment ponds that may have a wide variety of chemical precipitates, dissolved chemicals, and sediment levels. This complicates the classification of water in surface mines.

In comparison to the NAIP classification, producer's accuracy was greater for all classes for the RapidEye classification using either the SVM or RF algorithm, and the user's accuracy was higher in comparison to NAIP for all classes except barren for the RF model. This generally suggests that the increase in accuracy associated with RapidEye cannot be attributed to alleviating a single misclassification problem; RapidEye allows for enhanced separation of most classes in comparison to NAIP.

Since the reclaimed-woody vegetation class showed confusion with forest cover, it was combined with the forest class to produce a woody vegetation class and, in doing so, simplifying the classification to a four-class problem. This increased the classification accuracy of the NAIP (6.5 m) data by 5.4% using the SVM algorithm and 5.2% using the RF algorithm. Although the reduction in the number of classes resulted in a substantial improvement in classification accuracy for NAIP, these results were still significantly less accurate than the RapidEye classification with four classes (z-score (SVM) = 4.469 and z-score (RF) = 4.833). Combining all vegetation classes to create a three-class problem increased the accuracy by 10.7% in comparison to the five class classification for the NAIP classification using the SVM algorithm and 10.7% using the RF algorithm. Nevertheless, the RapidEye results with just three classes were still more accurate than the NAIP classification (z-score (SVM) = 4.181 and z-score (RF) = 3.579). Even though RapidEye provided a higher classification accuracy, the NAIP classification accuracy was nevertheless high (SVM = 94.3% and RF = 96.7%), suggesting that NAIP may be appropriate if a relatively small number of classes are to be separated. Although the simplification of the classification scheme is generally likely to improve the overall classification accuracy by reducing the chance for confusion, this research suggests that if the number of classes being separated is less important than the overall classification accuracy, it may be of value to combine classes.

Although many studies suggest an improvement in classification with the inclusion of textural measures (for example, Ghimire et al., 2010; Rodríguez-Galiano et al., 2012b), in this study, incorporating all 84 textural measures derived from the NAIP data did not increase accuracy to that of the RapidEye spectral data alone. Improvements were observed for the NAIP spectral and texture bands compared to classification using only the image bands from NAIP, but

only in some cases was this improvement statistically significant. On the other hand, in some cases, the addition of the textural bands decreased the overall accuracy, especially for the SVM algorithm. For all bands used (red, green, first principle component, and second principle component), the textural measures were found to be much less important than the spectral bands, suggesting they provide little additional information for the classification.

It has been suggested that increasing the number of predictor variables can cause a decrease in classification accuracy, contrary to what might be expected, a phenomenon known as the Hughes effect or "curse of dimensionality" (Hughes, 1968; Rodríguez-Galiano et al., 2012b). For the RF algorithm, this decrease in performance with additional predictor variables was noted by Evans et al. (2011) and Hastie et al. (2009). As a result, feature selection was tested as a means to increase the classification accuracy when using measures of texture. The feature selection method implemented is after Murphy et al. (2010) and uses a model selection approach that uses variable importance measures to select a model with reduced dimensionality and potentially increased classification accuracy. Models using the top 10%, 20%, 50%, 70%, 90%, and all the variables were tested. The greatest accuracy was found using the top 10% of the variables, which included all the image bands and a subset of the spectral bands. Even though an increase in accuracy was observed using the top 10% of the predictor variables, all of the image bands were nevertheless found to be of greater importance than the selected textural measures. Using the top 10% of the bands resulted in a statistically significant increase in accuracy relative to using all bands and textural measures (z-score = 2.090) and a statistically significant increase relative to using just the spectral bands (z-score = 4.016) for the RF algorithm. However, it was still less accurate than the RapidEye classification (z-score = 3.491).

Given that producing these texture bands is time-consuming, deciding upon which measures to use, appropriate kernel sizes and offset distances, and which spectral bands to calculate them from, can be subjective, and feature selection may be required to reap the full benefit of these measures, it is questionable whether there is merit in calculating these measures as the potentially slight classification improvement is outweighed by complexity and the time-consuming nature of the task.

The NAIP data were also smoothed using a Gaussian low pass filter with kernel sizes of 3 x 3, 5 x 5, and 7 x 7 pixels. Smoothing was found to produce a greater increase in accuracy than adding texture bands. Smoothing with a 3 x 3 pixel kernel did not statistically increase the classification accuracy with the SVM algorithm (z-score = 0.686); however, it did for RF (z-score = 2.546). At kernel sizes of 5 x 5 and 7 x 7 pixels, a statistically significant increase was observed, and the highest classification accuracy for the five class-problem using the NAIP imagery was obtained using the RF algorithm with the image resampled to 6.5 m and a Gaussian low pass filter with a 7 x 7 pixel kernel size applied. An overall accuracy of 88.1% was obtained (a 4.7% increase in accuracy relative to the NAIP 6.5 m classification without smoothing and a 8.1% increase in accuracy relative to the NAIP 1 m classification without smoothing), however this classification was still found to be statistically less accurate than the classification of the RapidEye data without any smoothing (z-score = 2.965). The model using the top 10% of the image bands and textural measures was found to be statistically comparable to this output (z-score = 0.825).

Comparing the results using the textural measures and the Gaussian low pass filter indicates that smoothing the data yielded a higher increase in classification accuracy than utilizing texture unless feature selection was implemented. Perhaps this is a result of reduced

heterogeneity within class signatures. Minimizing this heterogeneity using smoothing may be an alternative to trying to characterize the heterogeneity using textural calculations.

Conclusions

In this study, the accuracy of classification with NAIP was statistically less accurate than that of RapidEye satellite imagery for classification of five classes within a surface mine permit. This suggests that RapidEye satellite data are more suited for this mapping task. Nevertheless, NAIP orthophotography has many characteristics that make it useful for surface mine mapping and monitoring, including its availability for multiple years, a general lack of cloud cover, contiguous coverage of large areas, availability, and low cost to the user.

Although the accuracy of classification with NAIP was not found to be comparable to that of RapidEye, the NAIP imagery provided a high classification accuracy when the number of classes was reduced to four or fewer classes. If the aim of a mapping project is only to differentiate vegetated and non-vegetated cover, or to differentiate woody vegetation from non-woody vegetation, NAIP may be adequate. Reducing the spatial detail, either by simply increasing the pixel size or by smoothing the data using a Gaussian low pass filter, also increased classification accuracy. Textural measures from the GLCM were of value, but feature selection was necessary to select amongst the large number of derived texture variables. If NAIP data are to be used for classification, the NIR band is of value and should be acquired and used, if available.

In summary, NAIP orthphotography offers a wealth of data and should not be ignored as a potential data set for classification. Although perhaps not as useful as satellite imagery for image classification, NAIP does offer many favorable characteristics for mapping changing terrains, such as the mountaintop mining region of the eastern United States.

Acknowledgments

Funding support for this study was provided by West Virginia View, Alderson Broaddus University, and the Appalachian Research Initiative for Environmental Science (ARIES). The views, opinions and recommendations expressed herein are solely those of the authors and do not imply any endorsement by ARIES employees, other ARIES-affiliated researchers or industrial members. The project described in this publication was also supported in part by Grant/Cooperative Agreement Number 08HQGR0157 from the United States Geological Survey via a subaward from AmericaView. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS. Lastly, we would like to thank two anonymous reviewers for their helpful comments that greatly improved the manuscript.

References

Aerial Photography Field Office (APFO), 2012, "Imagery Programs: National Agriculture Imagery Program (NAIP) [available online at http://fsa.usda.gov/FSA/apfoapp?area=home&subject=prog&topic=naip].

Agresti, A., 2007, *An Introduction to Categorical Data Analysis*, 2nd Ed., Hoboken, New Jersey: Wiley-Interscience.

Agüera, F., Aguilar, F. J., and M. A. Aguilar, 2008, "Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses," *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(6): 635-646.

Baker, B. A., Warner, T. A., Conley, J. F., and B. E. McNeil, 2013, "Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations," *International Journal of Remote Sensing*, 34(5): 1633-1651.

Beyer, H. L., 2012, *Geospatial Modeling Environment (Version 0.6.0.0)* [available online at http://www.spatialecology.com/gme].

Bernhardt, E. S., and M. A. Palmer, 2011, "The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians," *Annals of the New York Academy of Sciences*, 1223(1): 39-57.

Bozheva, A. M., Petrov, A. N., R. Sugumaran, 2005, "The effect of spatial resolution of remotely sensed data in dasymetric mapping of residential areas," *GIScience & Remote Sensing*, 42(2): 113-130.

Bradley, J. V., 1968, *Distribution-Free Statistical Tests*, Englewood Cliffs, New Jersey: Prentice Hall.

Breiman, L., 1996, "Bagging predictors," *Machine Learning*, 24(2): 123-140.

Breiman, L., 2001, "Random forests," *Machine Learning*, 54(1): 5-32.

Burges, C. J. C., 1998, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discover*, 2(2): 121-167.

Chan, J. C. W., Laporte, N., and R. S. Defries, 2003, "Textural classification of logged forests in tropical Africa using machine-learning algorithms," *International Journal of Remote Sensing*, 24(6): 1401-1407.

Chica-Olmo, M., and F. Abarca-Hernández, 2000, "Computing geostatistical image texture for remotely sensed data classification," *Computers & Geoscience*, 26(4): 373-383.

Cushnie, J. L., 1987, "The interactive effort of spatial resolution and degree of internal variability within land-cover types on classification accuracies," *International Journal of Remote Sensing*, 8(1): 15-29.

Davies, K. W., Petersen, S. L., Johnson, D. D., Davis, D. B., Madsen, M. D., Zvirzdin, D. L., and J. D. Bates, 2010, "Estimating juniper cover from National Agriculture Imagery Program (NAIP) imagery and evaluating relationships between potential cover and environmental variables," *Rangeland Ecology & Management*, 63(6): 630-637.

Dietterich, T. G., 1998, "Approximate statistical test for comparing supervised classification learning algorithms," *Neural Computation*, 10(7): 1895-1923.

Drummond, M. A. and T. R. Loveland, 2010, "Land-use pressure and a transition to forest-cover loss in the eastern United States," *BioScience*, 60(4): 286-298.

Evans, J. S. Murphy M. A., Holden, Z. A., and S. A. Cushman, 2011, "Modeling species distribution and change using Random Forest," in *Landscape Ecology: Concepts and Application* Drew, C. A., Wiersma, Y. F., and F. Huettman (Eds.), New York, NY: Springer.

Foody, G. M., 2004, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering & Remote Sensing*, 70(5): 627-633.

Fritz, K. M., Fulton, S., Johnson, B. R., Barton, C. D., Jack, J. D., Word, D. A., and R. A. Burke, 2010, "Structural and functional characteristics of natural and constructed channels draining a reclaimed mountaintop removal and valley fill coal mine," *Journal of the North American Benthological Society*, 29(2): 637-689.

Ghimire, B., Rogan, J., and J. Miller, 2010, "Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic," *Remote Sensing Letters*, 1(1): 45-54.

Ghimire, B., Rogan, J., Rodríguez-Galiano, V., Panday, P., and N. Neeti, 2012, "An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA," *GIScience & Remote Sensing*, 49(5): 623-643.

Guo, Q., Kelly, M., Gong, P., and D. Liu, 2007, An object-based classification approach in mapping tree mortality using high spatial resolution imagery," *GIScience & Remote Sensing*, 44(1): 24-47.

Hansen, M., Dubayah, R., and R. DeFries, 1996, "Classification trees: an alternative to traditional land cover classifiers," *International Journal of Remote Sensing*, 17(5): 1075-1081.

Haralick, R. M., Shanmugam, K., and I. H. Dinstein, 1973, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, 3: 610-621.

Hastie, T. Tibshirani R., and J. Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Ed., New York, NY: Springer.Hooke, R. L., 1999, "Spatial distribution of human geomorphic activity in the United States: Comparison to rivers," *Earth Surface Processes and Landforms*, 24(8): 687-692.

Hughes, G., 1968, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, 14(1): 55-63.

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

Kazar, S. A., and T. A. Warner, 2013, "Assessment of carbon storage and biomass on minelands reclaimed to grassland environments using Landsat spectral indices," *Journal of Applied Remote Sensing*, 7(1): 073583. DOI: 10.1117/1.JRS.7.073583.

Keene, T. and J. Skousen, 2010, "Mine spoil reclamation with switchgrass for biofuel production," *2010 National Meeting of the American Society for Mining and Reclamation*, June 5-11, 2010, Pittsburgh, Pennsylvania: 1-16.

Latty, R. S., Nelson, R., Markham, B., Williams, D., Toll, D., and J. Irons, 1985, "Performance comparison between information extraction techniques using variable spatial resolution data," *Photogrammetric Engineering & Remote Sensing*, 51(9): 1459-1470.

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

Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F. M. B., De Baets, B., and N.E.C. Verhoest, 2012, "Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification," *International Journal of Applied Earth Observation and Geoinformation*, 19: 173-184.

Maxwell, A. E., Warner, T. A., Strager, M. P., and M. Pal, 2014, "Combining RapidEye satellite imagery and Lidar for mapping of mining and mine reclamation," *Photogrammetric Engineering & Remote Sensing*, 80(2): 179-189.

Meneguzzo, D. M., Liknes, G. C., and M. D. Nelson, 2013, "Mapping trees outside of forests using high-resolution aerial imagery: a comparison of pixel- and object-based classification approaches," *Environmental Monitoring and Assessment*, 185: 6261-6275.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and F. Leisch, 2012, "e1071: Misc functions of the department of statistics (e1071)," *R Package Version 1.6-1*, [available online at http://CRAN.R-project.org/package=e1071.

Murphy, M. A., Evans, J. S., and A. S. Storfer, 2010, "Quantifying *Bufo boreas* connectivity in Yellowston National Park with landscape genetics," *Ecology*, 91: 252-261.

Myeong, S., Nowak, D. J., Hopkins, P. F., and R. H. Brock, 2001, "Urban cover mapping using digital, high-spatial resolution aerial imagery," *Urban Ecosystems*, 5: 243-256.

Myneni, R. B., Maggion, S., Iaquinta, J., Privette, J. L., Gobron, N., Pinty, B., Kimes, D. S., Verstraete, M. M., and D. L. Williams, 1995, "Optical remote sensing of vegetation: Models, caveats, and algorithms," *Remote Sensing of Environment*, 51: 169-188.

Negley, T. L., 2002, *A Comparative Hydrologic Analysis of Surface-Mined and Forested Watersheds in western Maryland*, M.S. Thesis, College Park, Maryland: University of Maryland.

Pal, M., 2005, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, 26(1): 217-222.

Pal, M., and P. M. Mather, 2005, "Support vector machines for classification in remote sensing," *International Journal of Remote Sensing*, 5(10): 1007-1011.

Palmer, M. A., Bernhardt, E. S., Schlesinger, W. H., Eshleman, K. N., Foufoula-Georgiou, E., Hendryx, M. S., Lemly, A. D., Likens, G. E., Loucks, O. L., Power, M. E., White, P. S., and P. R. Wilcock, 2010, "Mountaintop mining consequences," *Science*, 327: 148-149.

Phinn, S. R., 1998, "A framework for selecting appropriate remotely sensed data dimensions for environmental monitoring and management," *International Journal of Remote Sensing*, 19(17): 3457-3463.

Pontius, R. G., Jr., and M. Millones, 2011, "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment," *International Journal of Remote Sensing*, 32(15): 4407-4429.

RapidEye, 2009, *RapidEye Standard Image Product Specifications*, Berlin, Germany: RapidEye AG.

Rathore, C. S., and R. Wright, 1993, "Monitoring environmental impacts of surface coal-mining," *International Journal of Remote Sensing*, 14: 1021-1042.

R Development Core Team, 2012, *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.

Rodríguez-Galiano, V. F., Abarca-Hernández, F., Ghimire, B., Chica-Olmo, M., Atkinson, P. M., and C. Jeganathan, 2011, "Incorporating spatial variability measures in land-cover classification using random forest," *Procedia Environmental Sciences*, 3: 44-49.

Rodríguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and J. P. Rigol-Sanchez, 2012a, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, 67: 93-104.

Rodríguez-Galiano, V. F., Chica-Olma, M., Abarca-Hernández, F., Atkinson, P. M., and C. Jeganathan, 2012b, "Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture," *Remote Sensing of Environment*, 121: 93-107.

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., and C. Fischer, 2003, "Land-cover change monitoring with classification trees using Landsat TM and ancillary data," *Photogrammetric Engineering & Remote Sensing*, 69(7): 793-804.

Saylor, K. L., 2008, "Land Cover Trend Project: Central Appalachians," Washington, DC: U.S. Department of the Interior, U.S. Geological Survey [available online at http://landcovertrends.usgs.gov/east/eco69 Report.html].

Sen, S., Zipper, C. E., Wynne, R. H., and P. F., Donovan, 2012, "Identifying revegetated mines as disturbance/recovery trajectories using an interannual Landsat chronosequence," *Photogrammetric Engineering & Remote Sensing*, 78(3): 223-235.

Simmons, J. A., Currie, W. S., Eshleman, K. N., Kuers, K., Monteleone, S., Negley, T. L., Pohlad, B. R., and C. L. Thomas, 2008, "Forest to reclaimed mine land use change leads to altered ecosystem structure and function," *Ecological Applications*, 18(1): 104-118.

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

Stehman, S. V., and G. M. Foody, 2009, "Accuracy assessment," in *The SAGE Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D., and G. M. Foody (Eds.), London, UK: Sage Publications Ltd.

Su, L., and Y. Huang, 2009, "Support vector machines (SVM) classification: Comparison of linkage techniques using a cluster-based method for training data selection," *GIScience & Remote Sensing*, 46(4): 411-423.

Tarnavsky, E., Stow, D., Coulter, L., and A. Hope, 2004, "Spatial and radiometric fidelity of airborne multispectral imagery in the context of land-cover change analysis," *GIScience and Remote Sensing*, 41(1): 62-80.

Toutin, T., 2009, "Fine spatial resolution optical sensors," in *The SAGE Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D., and G. M. Foody (Eds.), London, UK: Sage Publications Ltd.

Townsend, P. A., Helmers, D. P., Kingdon, C. C., McNeil, B. E., de Beurs, K. M., and K. N. Eshleman, 2009, "Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976-2006 Landsat time series," *Remote Sensing of Environment*, 113: 62-72.

Tyc, G., Tulip, J., Schulten, D., Krischke, M., and M. Oxfort, 2005, "The RapidEye mission design," *Acta Astronautica*, 56: 213-219.

Vapnik, V. N., 1995, *The Nature of Statistical Learning Theory*, New York, New York: Springer-Verlag.

USEPA, 2005, *Final Programmatic Environmental Impact Statement (PEIS) on Mountaintop Mining/Valley Fills in Appalachia* (EPA 9-03-R-05002), Washington, DC: U.S. Environmental Protection Agency [available online at http://www.epa.gov/region3.mtntop/eis2005.htm].

Warner, T.A., 2011, "Kernel-based texture in remote sensing image classification," *Geography Compass*, 5(10): 781-798.

Warner, T. A., Nellis, M. D. and G. M. Foody, 2009, "Remote sensing scale and data selection issues," in *The SAGE Handbook of Remote Sensing*, Warner, T. A., Nellis, M. D., and G. M. Foody (Eds.), London, UK: Sage Publications Ltd.

Wickham, J. D., Riitters, K. H., Wade, T. G., Coan, M., and C. Homer, 2007, "The effect of Appalachian mountaintop mining on interior forest," *Landscape Ecology*, 22: 179-187.

Zégre, N. P., Maxwell, A. E., and S. Lamont, 2013, "Characterizing streamflow response of a mountaintop-mined watershed to changing land use," *Applied Geography*, 39: 5-15.

Zipper, C. E., Burger, J. A., Skousen, J. G., Angel, P. N., Barton, C. D., Davis, V., and J. A. Franklin, 2011, "Restoring forests and associated goods and services on Appalachian coal surface mines," *Environmental Management*, 47(5): 751-765.

Table 1: Characteristics of NAIP orthophotography and RapidEye satellite imagery used.

Sensor and Data Attributes	National Agriculture Imagery Program (NAIP) Orthophotography	RapidEye Satellite Imagery
Collection Date	14 July 2011	1 August 2011
Platform	Aircraft	Satellite
Sensor Type	Digital camera (Intergraph Z/I Imaging Digital Mapping Camera (DMC))	Multi-spectral push broom imager
Spatial Resolution	1 m GSD	6.5 m GSD (Resampled to 5 m)
Spectral Resolution	Blue (400 nm – 580 nm) Green (500 nm – 650nm) Red (590 nm – 675 nm) NIR (675 nm – 850 nm)	Blue (440 nm – 510 nm) Green (520 nm – 590 nm) Red (630 nm – 685 nm) Red Edge (690 nm – 730 nm) NIR (760 nm – 850 nm)
Radiometric Resolution	12-bit (Provided at 8-bit)	12-bit (Provided at 16-bit)
Temporal Resolution	Potentially every other growing season	Potentially daily (5.5 days at nadir)
Seasonality	Growing season (leaf-on)	Potentially multi-seasonal

Table 2: Land cover class definitions.

Class	Description
Forested	Land dominated by mature, woody vegetation that has not been directly disturbed by surface mining; mature forest that generally represents pre-mining conditions of the slopes
Reclaimed-herbaceous vegetation	Reclaimed areas dominated by herbaceous/non-woody vegetation
Reclaimed-woody vegetation	Reclaimed areas dominated by clumped or clustered woody plants that include shrubs and immature trees
Barren	Barren land lacking vegetation; manmade structures; haul roads; active quarries; lands disturbed by mining
Water	Water, including retention ponds, streams, and standing water

Table 3: Number of training polygons, areas (ha) of training polygons, and number of randomly selected training pixels manually interpreted for each land cover class based on manual interpretation of multiple sources.

Land Cover Class	**Number of Training Polygons**	**Area of Training Polygons (ha)**	**Number of Training Pixels Randomly Selected From Within The Training Polygons**
Forested	81	77.0	1000
Reclaimed-herbaceous vegetation	85	25.5	1000
Reclaimed-woody vegetation	55	8.7	1000
Barren	64	21.7	1000
Water	30	1.7	386

Table 4: Comparison of map accuracy for NAIP and RapidEye. A *z*-score for McNemar's test larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is more accurate than the other.

Method	**Parameter**	**NAIP**	**RapidEye**	**McNemar *z*-score**
Comparisons at Original Cell Size (1 m NAIP, 5 m RapidEye)				
SVM	OA (%)	81.2	90.8	8.062*
	K (%)	74.5	87.4	
	AD (%)	10.1	2.9	
	QD (%)	8.7	6.3	
RF	OA (%)	80.0	90.6	8.322*
	K (%)	73.0	87.1	
	AD (%)	10.2	3.4	
	QD (%)	9.8	6.0	
Comparisons at Common Cell Size (6.5 m NAIP and RapidEye)				
SVM	OA (%)	83.6	90.3	5.852*
	K (%)	77.8	86.6	
	AD (%)	7.0	3.5	
	QD (%)	9.4	6.2	
RF	OA (%)	83.4	90.9	6.041*
	K (%)	77.4	87.4	
	AD (%)	8.8	4.3	
	QD (%)	7.8	4.8	
Comparison at Same Radiometric Scale (8-bit) and Original Cell Size (1 m NAIP, 5 m RapidEye)				
Method	Parameter	NAIP	RapidEye	McNemar *z*-score
SVM	OA (%)	81.2	90.5	7.673*
	K (%)	74.5	87.1	
	AD (%)	10.1	3.3	
	QD (%)	8.7	6.2	
RF	OA (%)	80.0	91.1	8.806*
	K (%)	73.0	87.8	
	AD (%)	10.2	3.0	
	QD (%)	9.8	5.9	
Comparison at Common Cell Size (6.5 m) and Same Radiometric Scale (8-bit)				
SVM	OA (%)	83.6	90.2	5.826*
	K (%)	77.8	86.5	
	AD (%)	7.0	3.2	
	QD (%)	9.4	6.6	
RF	OA (%)	83.4	91.1	6.652*
	K (%)	77.4	87.7	
	AD (%)	8.8	3.3	
	QD (%)	7.8	5.6	

OA = Overall Accuracy K = Kappa Statistic
AD = Allocation Disagreement QD = Quantity Disagreement

Table 5: Error matrix for NAIP (6.5 m cell size, 8-bit) classification using RF. Overall accuracy is 83.4%.

		Reference Data					
		Forested	Reclaimed-herbaceous vegetation	Reclaimed-woody vegetation	Barren	Water	User's Accuracy
Classification	Forested	238	0	67	0	6	76.5%
	Reclaimed-herbaceous vegetation	10	232	11	8	12	84.0%
	Reclaimed-woody vegetation	12	21	182	0	15	79.1%
	Barren	0	7	0	241	22	89.3%
	Water	0	0	0	11	205	94.9%
	Producer's Accuracy	91.5%	89.2%	70.0%	92.7%	78.8%	

Table 6: Error matrix for leaf-on RapidEye (6.5 m cells size, 12-bit) classification using RF. Overall accuracy is 90.9%.

		Reference Data					
		Forested	Reclaimed-herbaceous vegetation	Reclaimed-woody vegetation	Barren	Water	User's Accuracy
Classification	Forested	245	0	29	0	0	89.4%
	Reclaimed-herbaceous vegetation	0	238	13	3	2	93.0%
	Reclaimed-woody vegetation	15	4	218	0	0	92.0%
	Barren	0	18	0	251	19	87.2%
	Water	0	0	0	6	239	97.6%
	Producer's Accuracy	94.2%	91.5%	83.8%	96.5%	91.9%	

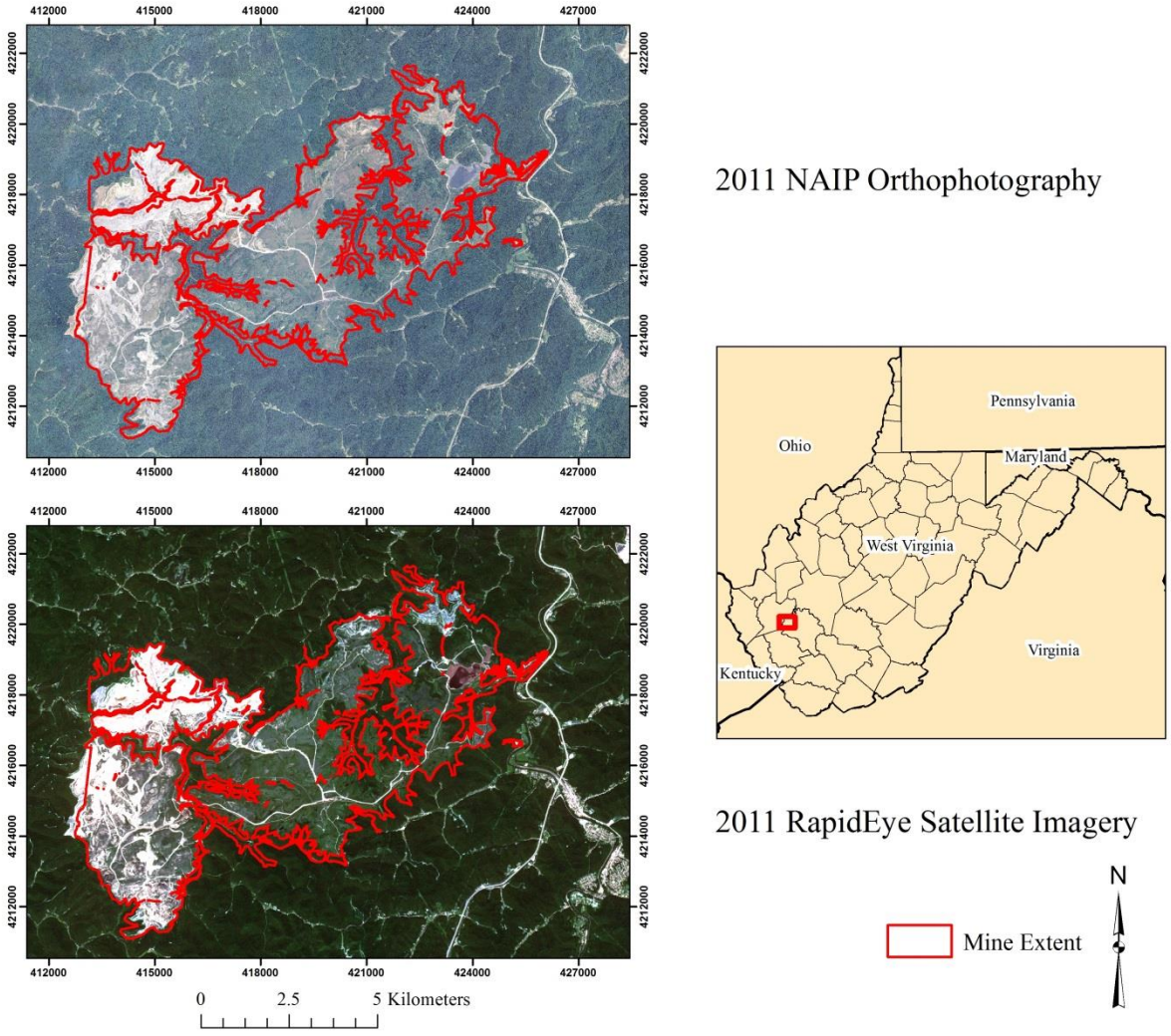

Figure 1: Hobet-21 surface mine complex. Base image for the top image is the NAIP orthophotography acquired on 14 July 2011. Base image for the bottom image is the RapidEye satellite imagery (© (2013) BlackBridge S.àr.l. All rights reserved.) acquired on 1 August 2011. The depicted mine extent is based on the surface mining permit obtained from WVDEP. The map is projected in NAD83 UTM Zone 17 N.

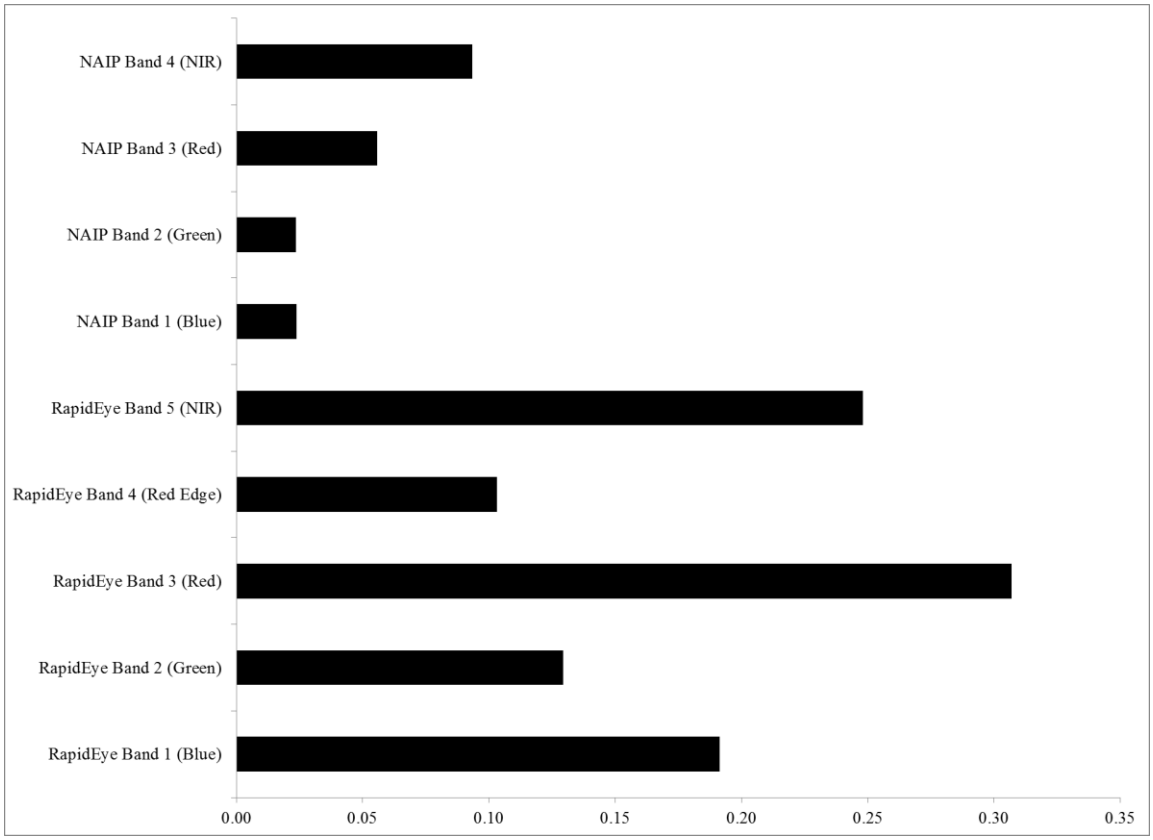

Figure 2: Relative importance of predictor variables for NAIP and RapidEye at provided cell size as estimated by the oob mean decrease in accuracy by RF.

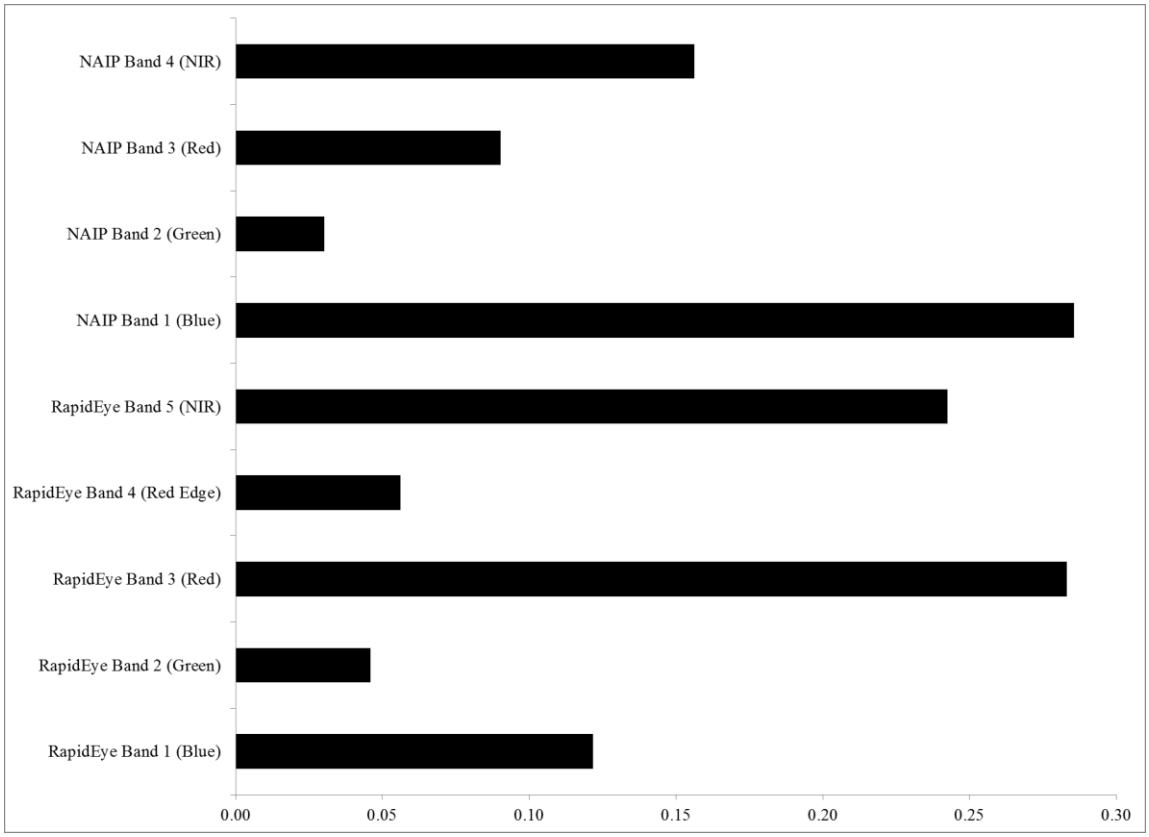

Figure 3: Relative importance of predictor variables for NAIP and RapidEye at common cell size (6.5 m) as estimated by the oob mean decrease in accuracy by RF.

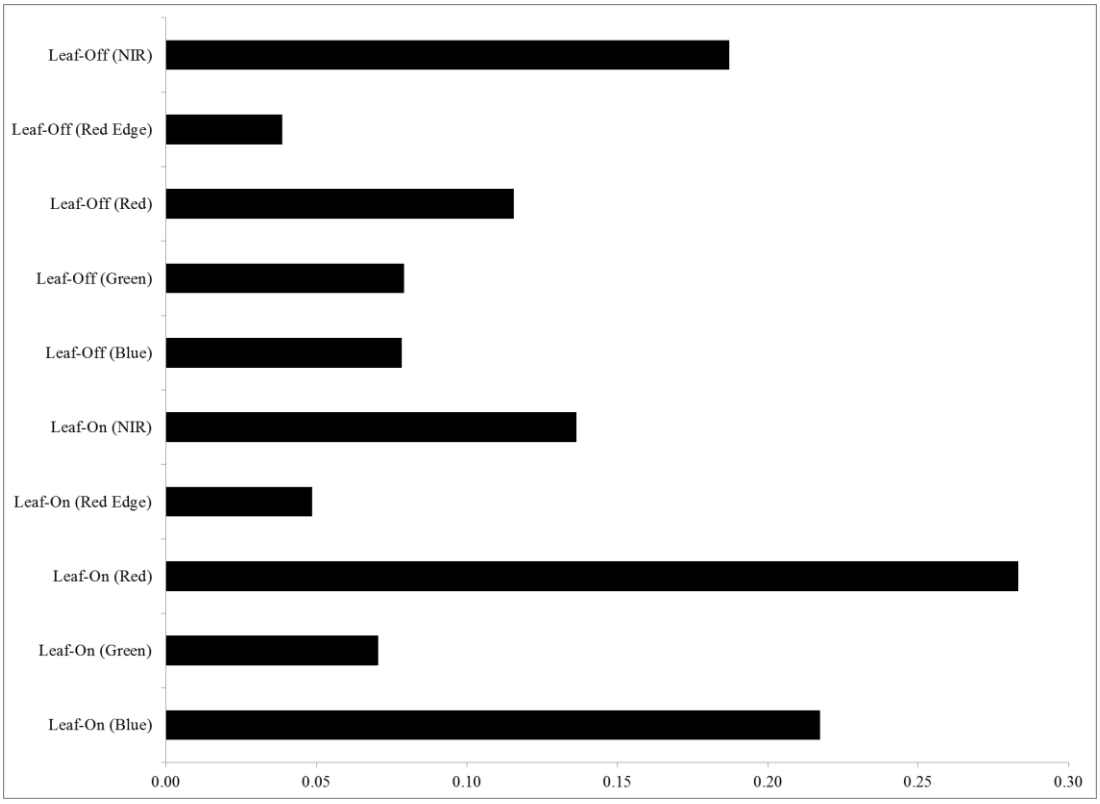

Figure 4: Relative importance of predictor variables for leaf-on and leaf-off RapidEye bands as estimated by the oob mean decrease in accuracy by RF.

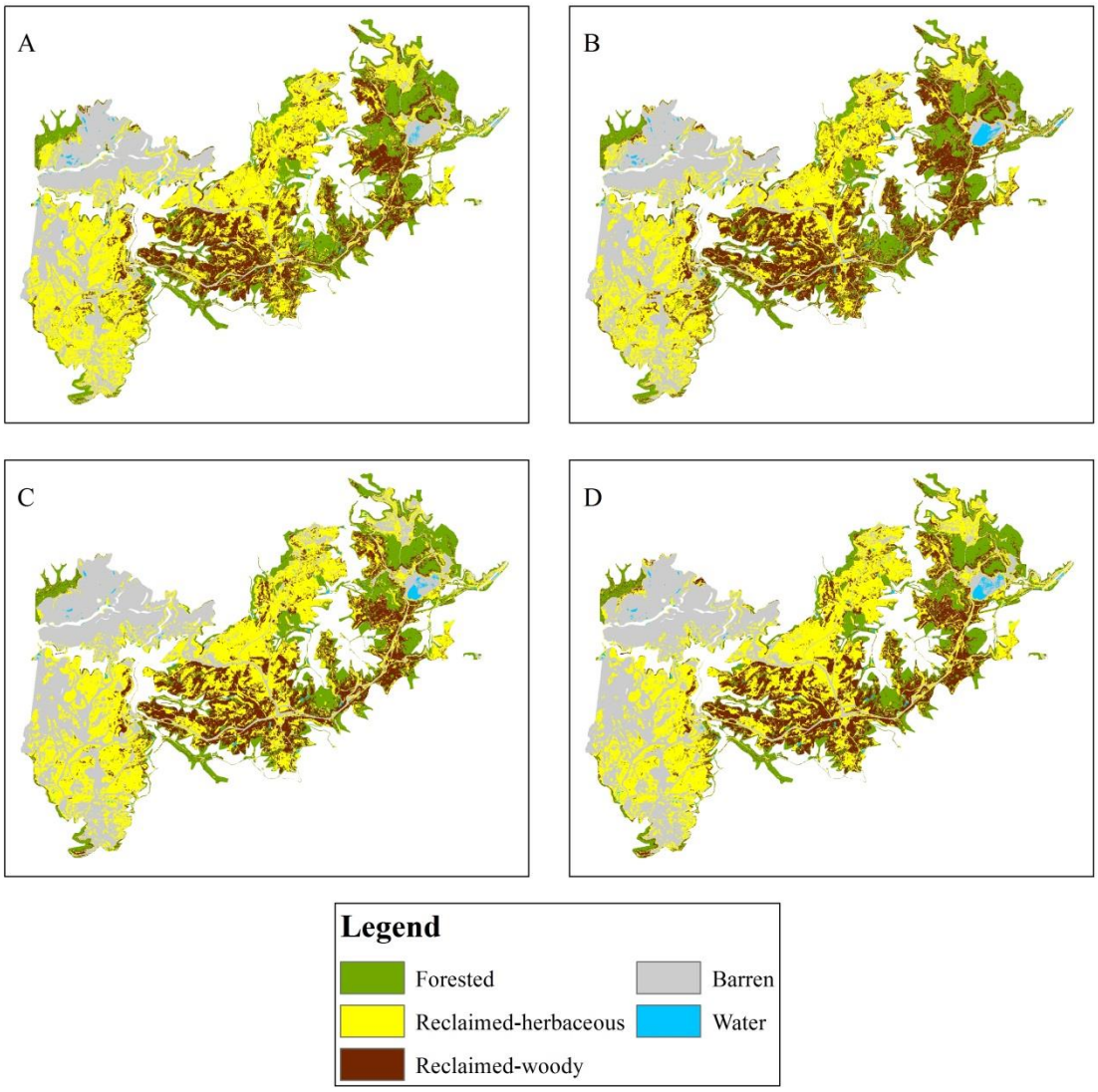

Figure 5: Land cover classification using (A) NAIP orthophotography (6.5 m) and SVM, (B) NAIP orthophotography (6.5 m) and RF, (C) RapidEye (6.5 m) and SVM, (D) RapidEye (6.5 m) and RF.

CHAPTER 4

Assessing machine learning algorithms and image- and LiDAR-derived variables for GEOBIA classification of mining and mine reclamation[1]

A.E. Maxwell, T.A. Warner, M.P. Strager, J.F. Conley, and A.L. Sharp

Abstract

This study investigates machine learning algorithms and measures derived from RapidEye satellite imagery and light detection and ranging (LiDAR) data for geographic object-based image analysis (GEOBIA) classification of mining and mine reclamation. Support vector machines (SVM), random forests (RF), and boosted classification and regression trees (CART) classification algorithms were assessed and compared to the *k*-nearest neighbor (*k*-NN) classifier. For GEOBIA classification of mine landscapes, the use of disparate data (i.e. LiDAR data) improved overall accuracy whereas the use of complex, object-oriented variables such as object geometry measures, first-order texture, and second-order texture from the grey level co-occurrence matrix (GLCM) decreased or did not improve the classification accuracy. SVM generally outperformed *k*-NN and the ensemble tree classifiers when only using the band means. With the incorporation of LiDAR-descriptive statistics, all four algorithms provided statistically comparable accuracies. *K*-NN suffered reduced classification accuracy with high dimensional feature spaces, suggesting that

[1] This is an Accepted Manuscript of an article published by Taylor & Francis in the *International Journal of Remote Sensing* on 17 February 2015, available online: http://www.tandfonline.com/doi/abs/10.1080/01431161.2014.1001086

Maxwell, A.E., T.A. Warner, M.P. Stager, J.F. Conley, and A.L. Sharp, 2015. Assessing machine-learning algorithms and image- and lidar-derived variables for GEOBIA classification of mining and mine reclamation, *International Journal of Remote Sensing*, 36(4): 954-978.

(Received 7 September 2014, Accepted 29 November 2014)

more complex machine learning algorithm may be more appropriate when a large number of predictor variables are used.

Keywords: GEOBIA; object-based classification; machine learning; LiDAR; RapidEye

1. Introduction

In recent years, pixel-based classification has been called into question, as image objects, defined as contiguous regions of pixels that are relatively spectrally homogenous, may better represent real objects on the ground than the original individual pixels. The process of segmenting an image and labeling the resulting image objects is commonly termed geographic object-based image analysis (GEOBIA). The GEOBIA approach has been described as a paradigm shift in remote sensing, and the number of papers referencing this technique has increased rapidly over the last decade (Blaschke and Strobl 2001; Walter 2004; Chubey, Franklin, and Wulder 2006; Drăgut and Blaschke 2006; Blaschke 2010; Baker et al. 2013; Meneguzzo, Liknes, and Nelson 2013; Blaschke et al. 2014). In applying supervised classification for labeling the image objects, GEOBIA differs from pixel-based classification in important ways that may have implications for designing an optimal classification strategy.

In GEOBIA, the original image digital number (DN) values are not normally used directly in the classification, instead summary attributes of the object are normally employed. These summary attributes can include features such as measures of central tendency of the individual bands or combinations of those bands, spectral variability (e.g. standard deviation (SD) or other texture measures), and object geometry (e.g. compactness) (Trimble 2011). Although many studies have demonstrated the benefit for classification of such object-derived variables, including object geometry (for example, Guo et al. 2007) and texture (Kim, Madden, and Warner 2009), there is not a clear consensus, especially with regards to their value in general land cover classification. For example, Yu et al. (2006) found that geometric features were of

little value for mapping detailed vegetation at the alliance level from Digital Airborne Imaging System (DAIS) imagery.

As an alternative to deriving additional variables from the objects, information from another sensor, such as light detection and ranging (LiDAR), can be integrated into the analysis to potentially improve the classification accuracy (Ke, Quackenbush, and Im 2010; Maxwell et al. 2014b), if such data are available. LiDAR data can provide information that is fundamentally different from that of spectral image data, and therefore may have more potential for increasing classification accuracy than the use of additional object-derived measures.

Because variables derived for image objects do not necessarily obey parametric statistical distributions, machine learning approaches are typically used for GEOBIA classification, with *k*-nearest neighbor (*k*-NN) classification, implemented in eCognition (Trimble 2011), the most commonly used method. However, *k*-NN is a relatively simple machine learning method, and may be less effective in the presence of high-dimensional data (Platt and Rapoza 2008). There are many other, potentially more powerful, machine learning algorithms that can be used for GEOBIA classification. For example, Duro, Franklin, and Dubé (2012a) found that support vector machines (SVM) and random forests (RF) were valuable for classifying image objects in agricultural landscapes. However, few studies have compared the effectiveness of *k*-NN classification to other machine learning approaches in the context of GEOBIA. One exception is Mallinis et al. (2008), who found that a decision tree (DT) classifier outperformed *k*-NN.

The goal of this research was to assess GEOBIA classification for mapping mining and mine reclamation land cover. It was anticipated that the incorporation of additional measures would improve the classification accuracy in comparison to just using spectral means and that more complex machine learning algorithms will outperform *k*-NN, resulting in improved

classifications for monitoring mining and mine reclamation. The following research questions were investigated:

1. Does incorporating object geometry or texture measures (first-order and second-order) improve GEOBIA classification accuracy?
2. Does incorporating LiDAR with multispectral data improve classification accuracy using GEOBIA?
3. How does the typically used *k*-NN compare to SVM, RF, and boosted classification and regression trees (CART) for GEOBIA classification?

2. Background

2.1. Mountaintop removal mining

Mountaintop removal coal mining is currently the leading cause of land cover change in the southern coalfields of the eastern United States including southern West Virginia, eastern Kentucky, and southwestern Virginia (Saylor 2008; Townsend et al. 2009; Drummond and Loveland 2010). It is estimated that 420,000 ha of land has been modified by surface mining in the Appalachian region between 1973 and 2000 (Wickham et al. 2007; Drummond and Loveland 2010). Surface mining reclamation typically results in the replacement of the original biodiverse forests (Master, Flack, and Stein 1998; Stein, Kutner, and Adams 2000; Wickham et al. 2007) with grasslands or shrublands (Simmons et al. 2008; Kazar and Warner 2013), removal of topsoil, and recontouring of the landscape (Palmer et al. 2010; Bernhardt and Palmer 2011), with associated impacts on water quality (for example, Merriam et al. 2003) and water quantity (Zégre et al. 2014).

Mapping this land cover alteration has historically focused on classification of moderate spatial resolution data such as Landsat Multispectral Scanner (MSS), Thematic Mapper (TM),

Enhanced Thematic Mapper Plus (ETM+), and Satellite Pour l'Observation de la Terre (SPOT) data (Anderson and Schubert 1976; Irons and Kennard 1986; Parks, Petersen, and Baumer 1987; Rathore and Wright 1993; Anderson et al. 1997; Prakash and Gupta 1998; Yuill 2003; Townsend et al. 2009; Sen et al. 2012). Less research has focused on the use of higher resolution data for mapping mining disturbance, although mine reclamation typically results in complex patterns of reclamation.

This paper is a part of a larger, on-going project that addresses the mapping of surface mines using high spatial resolution remotely sensed data. In our previous work, we addressed the relative benefits of different data and machine learning algorithms for pixel-based classification of mining and mine reclamation (Maxwell et al. 2014b). The incorporation of a LiDAR-derived normalized digital surface model (nDSM) statistically improved the classification accuracy for mapping land cover in comparison to only using spectral data from RapidEye satellite imagery. In addition, SVM outperformed the two ensemble decision tree classifiers, RF and boosted CART. A second study (Maxwell et al. 2014a) compared National Agriculture Imagery Program (NAIP) aerial orthophotography and satellite imagery for mapping surface mines, finding that satellite data resulted in a statistically more accurate classification, presumably due to the relatively more consistent viewing geometry and illumination provided by the satellite sensor. This paper expands upon our previous research by comparing *k*-NN to alternative machine learning algorithms for GEOBIA classification, investigating whether the incorporation of object attributes such as object geometry and object-specific first- and second-order texture increase classification accuracy, and evaluating any increase in comparison to adding disparate data, such as LiDAR.

2.2. Incorporating texture in GEOBIA

Image texture, a measure of local spatial variability in DN values, has been shown to improve land cover classification, especially when utilizing high spatial resolution data with low spectral resolution (Chica-Olmo and Abarca-Hernández 2000; Franklin et al. 2000; Asner et al. 2002; Chan, Laporte, and Defries 2003; Johansen et al. 2007; Agüera, Aguilar, and Aguilar 2008; Ghimire, Rogan, and Miller 2010; Rodríguez-Galiano et al. 2012b). However, previous studies provide contradictory evaluations regarding the benefit of the incorporation of texture (Warner 2011), possibly because of the difficulty in separating within-class texture, which is assumed to be of interest, from between-class texture (Ferro and Warner 2002). Texture derived over image objects may be more useful than the conventional texture derived from fixed-sized kernels generally used in per-pixel classification. This is because texture calculated over appropriately segmented objects should only capture within-class texture.

Texture can be categorized as either first- or second-order. First-order texture measures do not consider spatial location within the local region (i.e. kernel or object) over which texture is being calculated. The commonly used object variable SD is an example of a first-order texture. Second-order measures calculate texture only for pixels separated by a defined distance and direction. These second-order textural spatial associations are stored in a matrix, the grey level co-occurrence matrix (GLCM), and statistical measures of texture are then produced from this matrix (Haralick, Shanmugan, and Dinstein 1973; Haralick and Shanmugan 1974; Warner 2011).

Previous research has assessed the incorporation of texture in GEOBIA. For example, Kim, Madden, and Warner (2009) found an increase in classification accuracy when multiple measures of object-specific texture from the GLCM were incorporated for forest type mapping using IKONOS imagery. Yu et al. (2006) found that object-specific GLCM measures, including contrast, correlation, and dissimilarity, were of importance in detailed vegetation classification at

the alliance level from high spatial resolution DAIS imagery. However, Duro, Franklin, and Dubé (2012b) found that GLCM angular moment calculated from SPOT-5 normalized difference vegetation index (NDVI) was of little value in comparison to band means and SDs for GEOBIA classification of an agricultural landscape.

2.3. Combining multispectral data and LiDAR

Previous research has highlighted the improvement in classification accuracy when LiDAR is combined with optical data for mapping land cover (Cowen et al. 2000; Brennan and Webster 2006; Bork and Su 2007; Chust et al. 2008; Chen et al. 2009; Ke, Quackenbush, and Im 2010; Guo et al. 2011; Maxwell et al. 2014b). Also, LiDAR data have been integrated into different GEOBIA analysis approaches in various fields of application (for example, Blanchard, Jakubowski, and Kelly 2011; O'Neil-Dunne et al. 2013; Stal et al. 2013; Zhou 2013). The combination of imagery and LiDAR data has been investigated for mapping heterogeneous rangelands (Chen et al. 2009), urban landscapes (Brennan and Webster 2006; Guo et al. 2011), and coastal estuaries (Brennan and Webster 2006; Chust et al. 2008). Guo et al. (2011) specifically noted the usefulness of nDSM data for mapping urban landscapes. Ke, Quackenbush, and Im (2010) found that combining LiDAR and Quickbird multispectral imagery improved both the quality of segmentation and classification accuracy for forest species mapping using GEOBIA.

As an active remote sensing technology, LiDAR relies on recorded two-way travel time of transmitted laser pulses and precise geolocation derived from differential global positioning system (GPS) and inertial measurement unit (IMU) measurements (Lillesand, Kiefer, and Chipman 2008). In addition to elevation of the reflecting surface, most LiDAR systems record the return pulse intensity, which is in part a function of the reflectance of the surfaces returning

the laser pulse (Brantberg 2007). However, return intensity is also influenced by footprint size, scan angle, return number, and range distance (Lin and Mills 2010). Although some studies have demonstrated the benefit of using intensity for land cover classification (Brennan and Webster 2006), the use of LiDAR return intensity has not been as widely explored as elevation data in land cover mapping due to the difficulty of radiometric calibration of the returned laser intensity (Flood 2001; Kaasalainene et al. 2005).

2.4. Machine learning algorithms

In remote sensing, machine learning algorithms are of interest because they offer the potential to handle complex spectral measurement space, multidimensional data, and large volumes of data with reduced processing time compared to traditional classifiers (Hansen and Reed 2000). In this section, we discuss four machine learning algorithms: *k*-NN classification, SVM, RF, and boosted CART.

The *k*-NN classifier (Yu et al. 2006; Mallinis et al. 2008) classifies unknown samples by comparing their location in the feature space to those of the training samples. The unknown sample is assigned to the class most commonly found in the nearest *k* training object(s) within the feature space, where *k* is specified by the user (Steele, Winne, and Redmond 1998; Yu et al. 2006). *K*-NN is used widely in GEOBIA classification because it is nonparametric and because image segmentation commonly results in fewer samples to classify, so execution time is usually not an issue. *K*-NN is generally not used for pixel-based classification due to slow execution, when a large number of pixels must be classified (Hardin and Thomson 1992; Yu et al. 2006).

SVMs separate two classes with a multi-dimensional hyperplane that provides the maximum margin or best separation between the classes. A transformation to a higher dimensional space, where the training samples may be linearly separable, is accomplished using

a kernel function, such as a polynomial or radial basis function (RBF). A penalty parameter (*C*), penalizes training samples located on the "wrong" side of the decision boundary, allowing a degree of generalization (Vapnik 1995; Joachims 1998; Burges 1998; Tso and Mather 2003; Pal and Mather 2005; Pal 2005; Warner and Nerry 2009). Because SVM algorithms were originally designed for two class problems only, strategies have also been developed incorporating multiple SVM algorithms to produce a multi-class classification (Vapnik 1995; Tso and Mather 2003; Pal 2005; Pal and Mather 2005).

RF is a non-parametric learning algorithm that uses an ensemble of classification trees to improve upon the accuracy and consistency of single DT classifications. Each tree is generated from a subsample of the data obtained from random bootstrap sampling of the training data with replacement, a process known as bagging (Breiman 1996; Breiman 2001). The withheld, or out-of-bag (oob), samples can be used for accuracy assessment. A random subset of the predictor variables (the number of which is defined by user) is used for growing each tree of the ensemble. The random selection of training data and variables decreases the correlation between trees, and, in doing so, decreases the generalization error (Breiman 2001). RF has many attributes that make it attractive for image classification, including the capacity to model complex interactions between predictor variables, handling data with missing values, generating high classification accuracies, and providing measures of predictor variable importance (Steele 2000; Cutler et al. 2007).

Boosted CART, like RF, is also an ensemble classification of decision trees; however, the method used to generate the ensemble is fundamentally different. First, the entire training data set is used in each tree as opposed to a bootstrap sample (Freund and Schapire 1996; Ghimire et al. 2012). Second, misclassified samples in prior trees are given higher weights in subsequent

trees in order to address misclassification problems in the prior trees. This process, termed boosting, has been found to improve upon the classification of a single tree by as much as 50% (DeFries and Chan 2000; Muchoney et al. 2000; Friedl et al. 1999; Friedl et al. 2002; McIver and Friedl 2002; Lawrence et al. 2004; Ghimire et al. 2012), but may also result in overfitting of the training data (Bauer and Kohavi 1999).

3. Study area and data

The study area for this research is the 5,500 ha Hobet-21 mountaintop surface mine located in Boone and Lincoln counties, West Virginia, USA (Figure 1). In Figure 1, red hues generally indicate vegetation whereas blue and gray hues indicate barren areas within the mine in this leaf-off false colour composite. This mine is currently the largest permitted mine in the southern coalfields and the Appalachian region (Keene and Skousen 2010). The age of mine disturbance and reclamation varies; historical imagery shows that some of the mine disturbance predates 1987, while portions of the mine were still active at the time of the writing. This mine was selected because of its large spatial extent, wide variety in age of disturbance, vegetation, and land cover, and because imagery and LiDAR data were available that are nearly temporally coincident.

RapidEye data are the primary optical data in the study. The mine is covered by a single image tile, and this image was collected on 1 April 2010, prior to spring leaf out. The scene has an average center image view angle of -2.82˚, azimuth angle of 110.2˚, sun azimuth of 171.2˚, and sun elevation of 56.5˚. The RapidEye system consists of a constellation of five satellites that were launched in August 2008. The RapidEye sensors have five spectral bands: blue (440-510 nm), green (520-590 nm), red (630-730 nm), red edge (690-730 nm), and near infrared (NIR) (760-850 nm) (Tyc et al. 2005). The ground sampling distance of the sensors is 6.5 m. For this

project, the 3A product was used, which has radiometric, sensor, and geometric corrections applied, and is orthorectified to a 5 m grid.

LiDAR data were collected on 12 April 2010, within two weeks of the RapidEye data acquisition. Flight specifications were selected to support a nominal pulse spacing of 1 m. Using an aircraft flying at 1524 m above ground level and a flight speed of 125 knots (232 km hr^{-1}), the Optech ALTM 3100 C sensor was set to a pulse frequency of 70 kHz, a scan frequency of 35 Hz, and a scan angle of 36˚ (full swath). A 30% overlap was acquired between swaths. The LiDAR system recorded up to four returns per laser pulse, as well as both height and intensity information. The point data were classified by the vendor as ground, non-ground, or outliers, and delivered in LAS 1.2 format. The point classifications were utilized as they were provided, and no additional point classification or editing was performed.

4. Methods

4.1. Data segmentation

Image segmentation was performed in eCognition 8.0 (Trimble, Sunnyvale, California) using the multi-resolution image segmentation algorithm. This algorithm requires the user to define three parameters: scale, shape, and compactness (Trimble 2011). The scale parameter controls the size of the image objects (Liu and Xia 2010; Kim et al. 2011), and it has been suggested that this parameter has the largest impact on resulting classification accuracy (Blaschke 2003; Meinel and Neubert 2004, Kim, Madden, and Warner 2009; Liu and Xia 2010; Smith 2010; Myint et al. 2011). The shape parameter quantifies the relative weight assigned to the shape of the object versus the so-called color (or spectral properties), while compactness controls the balance between the form and edge length of the object (Baatz and Schäpe 2000). Methods to determine the optimal segmentation parameters have been proposed (for example,

Costa et al. 2008); however a test using the method suggested by Kim, Marguerite, and Warner (2009), which selects an optimal scale based on minimal spatial autocorrelation between adjacent objects, was not successful. Therefore, a trial-and-error approach using expert judgment to evaluate the segmentations was implemented, as is common in many object-based classifications (e.g. Laliberte, Fredrickson, and Rango 2007; Mathieu, Aryal, and Chong 2007, Dingle Robertson and King 2011; Myint et al. 2011; Pu, Landry, and Yu 2011; Duro, Franklin, and Dubé 2012a; Duro, Franklin, and Dubé 2012b). On this basis, the optimal parameters were judged to be 100 for scale, 0.1 for shape, and 0.5 for compactness. Only the spectral data were used to create the image objects, with each band equally weighted. This segmentation was used in all the classification experiments. The LiDAR data were not used in the segmentation process as one goal of this study was to assess the impact of including LiDAR data for classifying image objects. Including LiDAR in the segmentation process would have made all the classifications partially reliant on the LiDAR data, so this was avoided.

Table 1 lists the input predictor variables used in the GEOBIA classifications. The LiDAR variables were generated from point data gridded to match the 5 m RapidEye image. The nDSM was generated by subtracting the gridded ground surface from the gridded first return data. A first return intensity grid was also generated. The intensity data were not normalized for distance or other factors as this was not possible based on the limited LiDAR metadata, and our previous work indicated that despite not making these corrections, the LiDAR data were nevertheless very useful for land cover classification of mining and mine reclamation (Maxwell et al. 2014b). For each image object, a total of 74 predictor variables were calculated: five variables comprising the mean for all five spectral bands, five variables comprising the SDs for all five spectral bands, the eight object-specific GLCM textural measures calculated from each of

the five spectral bands (40 in total), 14 measures of object geometry, and five LiDAR-derived measures for both the nDSM and first return intensity (i.e. 10 in total).

The first- and second-order texture calculations resulted in 45 bands. It is well known that texture bands tend to have redundant information, which potentially could reduce classification accuracy. Therefore, to decrease the dimensionality of the texture data set, and reduce the correlation between the texture measures, principle component analysis was undertaken. In the resulting transformation, the first 10 principle components captured nearly 95% of the variability of the entire data set.

4.2. Data classification

Five land cover categories were mapped: forested, reclaimed-herbaceous vegetation, reclaimed-woody vegetation, barren, and water, as described in Table 2. These classes were chosen based on prior experience of typical land cover conditions within surface mines and reclaimed surface mines in the region. It should be noted that the term "forested land cover" is used to indicate forests in the permit area not yet removed for the mountaintop removal mining. The term does not necessarily imply that the area has never been disturbed at all; West Virginia's landscape records extensive historical disturbances ranging from clear-cutting to prior surface coal mining.

Table 3 describes the number of training objects used for each class. A total of 921 training objects were selected by manual interpretation of multiple data sources including the following: 1 m natural color imagery collected approximately one month prior to the LiDAR and RapidEye data by Pictometry International Corporation (Rochester, New York), the RapidEye imagery, LiDAR nDSM, and LiDAR bare earth contour data.

All classifications were performed using the statistical software tool R (R Core Development Team, 2012). *K*-NN classification was performed using the kknn package in R (Schliep and Hechenbichler 2014), SVM using the e1071 package (Meyer et al. 2012), RF using the randomForest package (Liaw and Wiener 2002), and boosted CART using the adabag package (Alfaro-Cortes, Gamez-Martinez, and Garcia-Rubio 2012). In order to obtain the most accurate classification for a specific classifier and input variable combination, parameter optimization was performed. For *k*-NN the *k* and the kernel parameters were optimized using leave-one-out cross validation as implemented in the kknn package in which one sample is withheld for validation and all other samples are used for training. The other classifiers were optimized using 10-fold cross validation. This optimization was conducted using a grid search of the specified parameters in R. Optimal parameters were estimated based on the minimum classification error obtained for the training data, with the data partitioned into 10 unique training sets, using a random assignment. The classifier was then trained 10 times using 90% of the data, and each time the remaining 10% of the data, not used for training in that instance, were used for validation. The goal of the optimization was to ensure that each classification algorithm used optimal parameters for the particular set of predictor variables tested. For RF and boosted CART a total of 500 trees were used in the ensemble, as this was found to be adequate to produce a stable classification result. For SVM a RBF kernel was used and the *C* and kernel-specific gamma (γ) parameters were optimized. For RF the number of predictor variables randomly sampled as candidates at each node (m) was optimized.

4.3. Error assessment

Classification accuracy assessment was performed at the object-level as suggested by Congalton and Green (2009) using 1,000 image objects that were randomly selected. The

validation objects were independent from the training objects, with no overlap. The dominant land cover class was assessed within the validation objects using Pictometry, RapidEye, and LiDAR data, and the interpretation was performed twice to ensure consistency. In constructing the error matrices, each randomly selected object was weighted by its area (Stehman and Czaplewski 1998; Congalton and Green 2009).

Quantity and allocation disagreement were calculated, as suggested by Pontius and Millones (2011). These two measures sum to overall error. Quantity disagreement provides a measure of error in the proportions of the categories, while allocation disagreement provides a measure of error in the spatial allocation of the categories (Pontius and Millones 2011).

In order to evaluate the statistical significance of any differences in the classifications the results were compared on a pairwise basis using McNemar's test (Dietterich 1998; Foody 2004). McNemar's test is a test of statistical difference that generates a *z*-score under the null hypothesis that the classifications are not different. A *z*-score larger than 1.645 indicates a 95% confidence of statistical significance for the one-directional test of whether one classification is more accurate than the other (Bradley 1968; Dietterich 1998; Foody 2004; Agresti 2007). Objects were weighted by area to calculate this statistical measure.

The relative importance of predictor variables was assessed from the oob mean decrease in accuracy from RF. The RF algorithm generates this measure during the training process by excluding each variable sequentially and recording the resulting increased oob error (Breiman 2001; Rodríguez-Galiano et al. 2012a; Rodríguez-Galiano et al. 2012b). This ancillary output of RF was used to assess the contribution of specific measures calculated for image objects (e.g. mean of image bands, SD of image bands, LiDAR descriptive statistics, object geometry measures, and measures of object-specific texture).

5. Results and discussion

5.1. Object geometry and texture measures for GEOBIA classification

Table 4 shows the overall accuracy, allocation disagreement, and quantity disagreement for the GEOBIA classifications using different predictor variable combinations. The overall accuracy varied from 70.7% (spectral means and geometry classified with *k*-NN) to 86.6% (spectral means and LiDAR classified with either SVM or boosted CART).

The incorporation of geometry decreased the classification accuracy for *k*-NN, SVM, and RF compared to only using spectral means, and this decrease was shown to be statistically significant in all cases (Table 5). The incorporation of geometry increased the classification accuracy for boosted CART; however, this increase was not statistically significant. Figure 2 shows the relative importance of predictor variables as band means and object geometry measures as estimated by the oob mean decrease in accuracy measure as determined by the RF classifier. These data suggest that object geometry measures were of little importance for classifying the land cover of interest and only complicated the feature space.

The incorporation of SD as a first-order texture calculated from the RapidEye spectral bands decreased the classification accuracy for all four classifiers compared to only using the spectral band means, and this decrease was statistically significant for *k*-NN and boosted CART (Tables 4 and 5). This is notable because SD is a standard object variable, typically used in GEOBIA classification. Further, the oob mean decrease in accuracy measure suggests that all band means are more important in the classification model than band SDs (Figure 3). Thus, object-specific first-order texture did not improve the classification accuracy over just using the band means when using RapidEye imagery. The incorporation of object-specific texture

measures from the GLCM also decreased the classification accuracy using *k*-NN, SVM and RF, and the classifications were shown to be statistically different (Tables 4 and 5).

Figure 4 showing the oob mean decrease in accuracy measure indicates that texture was of little importance in the classification, even after the principal component transformation was applied to the entire 45 texture variables to produce 10 uncorrelated variables. All spectral band means were of greater importance than the 10 principle component texture measures. The reduction of the number of texture variables from 45 to 10 did not improve the classification performance in comparison to just using the band means, confirming that texture is simply not of value for the classification of mining and mine reclamation using RapidEye satellite imagery. This finding may not be applicable if different image data were used or different classes were defined. To fully assess the use of texture for mapping mining and mine reclamation, it would be necessary to assess multiple data sets at varying spatial resolutions.

5.2. LiDAR for GEOBIA classification

As shown in Table 4, the inclusion of descriptive statistics derived from LiDAR improved the GEOBIA classification accuracy in comparison to just using the spectral band means alone and the classifications using the combined data set were statistically more accurate for all four algorithms (Table 5). Figure 5 further supports this conclusion; variable importance measures derived from RF as mean decrease in oob accuracy indicate that out of the top five predictor variables in the model, three were derived from the LiDAR nDSM data. Mean nDSM in the object was found to be the most important predictor variable in the model. Other important LiDAR-derived measures include SD for the nDSM, mean first return intensity, and maximum nDSM value within the object. Our previous research suggests that the incorporation of LiDAR-derived predictor variables can provide a substantial improvement in classification accuracy for

pixel-based classification in comparison to just using imagery data (Maxwell et al. 2014b). This research expands that conclusion to GEOBIA.

The most accurate classifications were achieved with the SVM and boosted CART algorithms using all band means plus 10 descriptive statistics derived from LiDAR as predictor variables. The SVM result is shown in Figure 6 with the error matrix provided as Table 6. An overall accuracy of 86.6% was obtained (Table 4). The 8.8% allocation disagreement and 4.6% quantity disagreement indicate that generally the error is a result of the mislabeling of objects, rather than the overall proportions of the areas of each class mapped. Producer's accuracy was 79.8% for the reclaimed-herbaceous vegetation class and 77.3% for the reclaimed-woody vegetation class. Confusion in the classification was noted between the reclaimed-herbaceous vegetation class and the barren class. This is attributed to mixed pixels as areas in an early stage of reclamation are often characterized by a sparse vegetative cover. Confusion was also noted between forest and reclaimed-woody vegetation; however, this confusion was reduced by the inclusion of LiDAR data.

In summary, the incorporation of LiDAR statistically improved the classification in comparison to just using the spectral band means whereas incorporation of geometry or texture did not improve or decreased the accuracy, especially for *k*-NN. This supports the assumption that one means to combat the low spectral resolution and high class heterogeneity inherent to high spatial resolution data, such as RapidEye satellite imagery, is to combine the spectral data with other data sources that provide additional predictor variables to augment the reduced spectral resolution. If classification accuracy is to be improved over that made available by just using image band means, it may be necessary to incorporate information from another sensor, such as LiDAR data.

Although the utility of geometry and texture may be sensor-specific, LiDAR data should be of value for this mapping task regardless of the spectral data source because it provides additional predictor variables that increase the separability of the land cover classes in the feature space. However, incorporating LiDAR will required further data collection whereas geometry and texture measures can be calculated directly from the spectral data or derived image objects. Issues of temporal alignment, practicality, and cost may limit the applicability of combining spectral and LiDAR data, especially over large areas.

5.3. Comparison of machine learning algorithms for GEOBIA

For *k*-NN the incorporation of GLCM measures decreased the accuracy by 8.8% in comparison to just using the spectral band means. Similarly, the incorporation of object geometry measures decreased *k*-NN accuracy by 12.4% in comparison to just using band means. Thus, in general, *k*-NN was not found to be robust within a complex measurement space, specifically a high-dimensional feature space with many redundant bands. This is notable because *k*-NN is commonly used in GEOBIA classification because it is provided in the eCognition program (Trimble 2011). If a large number of predictor variables are to be used, a more complex learning algorithm may provide a more accurate classification in comparison to *k*-NN.

For boosted CART, the accuracy was increased by the incorporation of GLCM measures by 0.3% in comparison to just using the band means, however this was not a statistically significant difference. We attribute this difference in performance between boosted CART and the other three classifiers to the means by which boosted CART selects variables for splitting at each node. Boosted CART does not implement random variable selection for each tree as is the case for RF. This means that the best predictor variables are potentially available for splitting in

each tree and less important variables can be ignored in the model. In contrast, in a RF model in which many variables are not useful, many individual trees may be dominated by not useful variables, potentially resulting in a reduction in accuracy. The accuracy of SVM may decrease as a hyperplane must be selected using many variables that complicate the feature space and do not greatly increase the separability of the classes. Thus, boosted CART may be more appropriate for classifying data in which a large number of predictor variables of low importance are included in the model.

Tables 7 shows the *z*-scores comparing the *k*-NN GEOBIA classification to classifications using the other three algorithms when using just the spectral band means and the spectral band means plus LiDAR descriptive statistics as predictor variables. When only the spectral band means were used, SVM statistically outperformed *k*-NN. The ensemble tree algorithms provided lower overall accuracies for this classification in comparison to *k*-NN (Table 4), however the difference was not statistically significant. For a classification using the spectral band means plus the LiDAR descriptive statistics, none of the algorithms provided statistically more accurate classification accuracies than *k*-NN. This suggests that the choice of classification algorithm may be of greater importance when classes of interest are less separable in the feature space. With the addition of LiDAR data, overall accuracy increased, and the choice of classification algorithm was less important.

Table 8 shows the *z*-scores comparing the SVM GEOBIA classification to classifications using the other three algorithms when using just the band means and the band means plus LiDAR. SVM significantly outperformed the other three algorithms when using just the spectral band means. Although it provided the highest classification accuracy when using the band means plus the LiDAR data, SVM was not statistically more accurate than the other three algorithms

given this combination of predictor variables. SVM generally provided higher classification accuracy than the ensemble tree classifiers for this classification task, similar to the results observed from our previous pixel-based analysis (Maxwell et al., 2014b), suggesting that SVM is a robust classifier for both GEOBIA and pixel-based classification. However, for GEOBIA classification, this increase may not be statistically significant given a feature space where the classes are more separable, such as when LiDAR is combined with optical data.

For most of the classifications allocation disagreement was greater than quantity disagreement (Table 4). A notable exception was RF, which for all data sets except spectral means and GLCM measures had a higher quantity disagreement than allocation disagreement, suggesting RF is less successful than the other classifiers at balancing errors of omission and commission and thus is less successful at estimating the proportions of the various cover types accurately.

6. Conclusions

For the mapping of mining and mine reclamation using RapidEye satellite data and LiDAR, SVM generally outperformed *k*-NN and the ensemble tree classifiers for GEOBIA classification. *K*-NN was not robust when used with high dimensional, redundant data, in that it produced reduced classification accuracy with the incorporation of object geometry and GLCM texture measures. This suggests that a more complex machine learning algorithm than *k*-NN, which is commonly used in GEOBIA analysis, may be more appropriate with a high dimensional feature space.

For GEOBIA, LiDAR-derived summary statistics improved the classification accuracy whereas object-specific geometry, first-order texture (including SD, a commonly used object variable), and second-order texture generally decreased or did not improve the accuracy. If

classification accuracy is to be improved from the relatively low values obtained from using image spectral band means, it may be necessary to incorporate information form another sensor, such as LiDAR data.

Although the methods outlined above show promise for mapping and monitoring change within individual mine permit areas at a fine scale, these methods are not without challenges. Combining disparate data, such as satellite imagery and LiDAR, requires multiple acquisitions, raising issues of practicality, temporal alignment, and cost. Such problems are exacerbated in a landscape experiencing rapid change, and where planning and coordinating multiple collections within a short period of time may be of importance.

Nevertheless, this study demonstrates that the combination of high spatial resolution multispectral satellite data and LiDAR can offer a means to map mining and mine reclamation with more spatial detail in comparison to moderate resolution sensors. Using machine learning algorithms to classify data from multiple sensors potentially allows monitoring environmental change and implementation of these technologies at an operational level.

Acknowledgments

LiDAR data were provided by the West Virginia Department of Environmental Protection (WVDEP) and the Natural Resource Analysis Center (NRAC) at West Virginia University. We would specifically like to acknowledge Adam Riley and Paul Kinder for their assistance in obtaining and processing the LiDAR data. We would like to thank two anonymous reviewers for their helpful comments that greatly improved the manuscript.

Funding

Funding support for this study was provided by West Virginia View, Alderson Broaddus University, and the Appalachian Research Initiative for Environmental Science (ARIES). The project described in this publication was also supported in part by Grant/Cooperative Agreement Number

[08HQGR0157] from the United States Geological Survey via a subaward from AmericaView. This work was supported by the Appalachian Research Initiative for Environment Science (ARIES) and United States Geologic Survey under Grant [number 08HQGR0157]. The views, opinions and recommendations expressed herein are solely those of the authors and do not imply any endorsement by ARIES employees, other ARIES-affiliated researchers or industrial members. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS.

References

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. 2nd ed., 400 pp. Hoboken, New Jersey: Wiley-Interscience.

Agüera, F., F .J. Aguilar, and M. A. Aguilar. 2008. "Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses." *ISPRS Journal of Photogrammetry and Remote Sensing* 63(6): 635-646.

Alfaro-Cortes, E., M. Gamez-Martinez, and N. Garcia-Rubio. 2012. "Adabag: Applies multiclass AdaBoost.M1, AdaBoost-SAMME and bagging, R Package Version 3.1." (accessed June 1, 2013). http://CRAN.R-project.org/package=adabag.

Anderson, A. T., D. Schultz, N. Buchman, and H. M. Nock. 1997. "Landsat imagery for surface-mine inventory." *Photogrammetric Engineering & Remote Sensing* 43(8): 1027-1036.

Anderson, A. T., and J. Schubert, 1976. "ERTS-1 data applied to strip mining." *Photogrammetric Engineering & Remote Sensing* 42: 211-219.

Asner, G. P., M. Keller, R. Pereira, Jr., and J. C. Zweede. 2002. "Remote sensing of selective logging in Amazonia: Assessing limitation based on detailed field observations, Landsat ETM+, and textural analysis." *Remote Sensing of Environment* 80(3): 483-496.

Baatz, M., and A. Schäpe. 2010. "Multiresolution segmentation – An optmimization approach for high quality multi-scale image segmentation." In *Angewandte Geographische Informationsverarveitung XII*, edited by J. Strobl, T. Blaschke, and G. Griesebener, 12-23. Berlin, Germany: Herbert Wichmann Verlag.

Baker, B. A., T. A. Warner, J. F. Conley, and B. E. McNeil. 2013. "Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations." *International Journal of Remote Sensing* 34(5): 1633-1651.

Bauer, E., and R. Kohavi. 1999. "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants." *Machine Learning* 36(1): 105-139.

Bernhardt, E. S., and M. A. Palmer. 2011. "The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians." *Annals of the New York Academy of Sciences* 1223(1): 39-57.

Blanchard, S. D., M. K. Jakubowski, and M. Kelly. 2011. "Object-based image analysis of downed logs in disturbed forested landscapes using LiDAR." *Remote Sensing* 3: 2420-2439.

Blaschke, T. 2003. "Object-based contextual image classification built on image segmentation." *2003 IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data* Greenbelt, Maryland, Ocotber 27-28, 13-119.

Blaschke, T. 2010. "Object based image analysis for remote sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2-16.

Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede. 2014. "Geographic object-based image analysis – Towards a new paradigm" *ISPRS Journal of Photogrammetry and Remote Sensing* 87(100): 180-191.

Blaschke, T, and J. Strobl. 2001. "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS." *GIS – Zeitschrift für Geoinformationssysteme* 14(6): 12-17.

Bork, E. W. and J. G. Su. 2007. "Integrating LiDAR data and multispectral imagery for enhanced classification of rangeland vegetation: A meta analysis." *Remote Sensing of Environment* 111(1): 11-24.

Bradley, J. V. 1968. *Distribution-Free Statistical Tests*. 388 pp. Englewood Cliffs, New Jersey: Prentice Hall.

Brantberg, T. 2007. "Classifying individual tree species under leaf-off and leaf-on conditions using airborne lidar." *ISPRS Journal of Photogrammetry and Remote Sensing* 61(5): 325-340.

Breiman, L. 1996. "Bagging predictors." *Machine Learning* 24(2): 123-140.

Breiman, L. 2001. "Random forests." *Machine Learning* 54(1): 5-32.

Brennan, R., and T. L. Webster. 2006. "Object-oriented land cover classification of lidar-derived surfaces." *Canadian Journal of Remote Sensing* 32(2): 162-172.

Burges, C. J .C., 1998. "A tutorial on support vector machines for pattern recognition." *Data Mining and Knowledge Discovery* 2(2): 121-167.

Chan, J. C. W., N. Laporte, and R. S. Defries. 2003. "Textural classification of logged forests in tropical Africa using machine-learning algorithms." *International Journal of Remote Sensing* 24(6): 1401-1407.

Chen, Y., W. Su, J. Li, and Z. Sun. 2009. "Hierarchical object oriented classification using very high resolution imagery and LiDAR data over urban areas." *Advances in Space Research* 43(7): 1101-1110.

Chica-Olmo, M., and F. Abarca-Hernández. 2000. "Computing geostatistical image texture for remotely sensed data classification." *Computers & Geoscience* 26(4): 373-383.

Chubey, M. S., S. E. Franklin, and M. A. Wulder. 2006. "Object-based analysis of IKONOS-2 imagery for extraction of forest inventory parameters." *Photogrammetric Engineering & Remote Sensing* 72(4): 383-394.

Chust, G., I. Galparsoro, Á Borja, J. Franco, and A. Uriarte. 2008. "Coastal and estuarine habitat mapping, using LiDAR height and intensity and multi-spectral imagery." *Estuarine, Coastal and Shelf Science* 78(4): 633-643.

Congalton, R. G., and K. Green. 2009. *Assessing the accuracy of remotely sensed data: principles and practices*. 2nd ed., 183 pp. London, United Kingdom: Taylor and Francis.

Costa, G. A. O. P., R .Q. Feitosa, T. B. Cazes, and B. Feijó. 2008. "Genetic adaption of segmentation parameters." In *Object-based Image Analysis*, edited by T. Blaschke, S. Lang, and G.J. Hay, 679-695. Berlin, Germany: Springer-Verlag.

Cowen, D .C., J. R. Jensen, C. J. Hendrix, M. E. Hodgson, and S. R. Schill. 2000. "A GIS-assisted rail construction econometric model that incorporates LiDAR data." *Photogrammetric Engineering and Remote Sensing* 66(11): 1323-1328.

Cutler, D. R., T. C. Edwards, Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. "Random Forests for classification in ecology." *Ecology* 88(11): 2783-2792.

DeFries, R. S., and J. C. W. Chan. 2000. "Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data." *Remote Sensing of Environment* 74(3): 503-515.

Dietterich, T. G. 1998. "Approximate statistical test for comparing supervised classification learning algorithms." *Neural Computation* 10(7): 1895-1923.

Dingle Robertson, L., and D. J. King. 2011. "Comparison of pixel- and object-based classification in land cover mapping." *International Journal of Remote Sensing* 32: 1505-1529.

Drăgut, L., and T. Blaschke. 2006. "Automated classification of landform elements using object-based image analysis." *Geomorphology* 81: 330-344.

Drummond, M. A. and T. R. Loveland. 2010. "Land-use pressure and a transition to forest-cover loss in the eastern United States." *Bioscience* 60(4): 286-298.

Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012a. "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." *Remote Sensing of Environment* 118: 259-272.

Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012b. "Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests." *International Journal of Remote Sensing* 33(14): 4502-4526.

Ferro, C. H. S., and T. A. Warner. 2002. "Scale and texture in digital image classification." *Photogrammetric Engineering & Remote Sensing* 68(1): 51-63.

Flood, M. 2001. "Laser altimetry: From science to commercial lidar mapping." *Photogrammetric Engineering & Remote Sensing* 67(11): 1209-1217.

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

Franklin, S. E., R .J. Hall, L. M. Moskal, A. J. Maudie, and M. B. Lavigne. 2000. "Incorporating texture into classification of forest species composition from airborne multispectral images." *International Journal of Remote Sensing* 21(1): 61-79.

Freund, Y. and R. E. Schapire. 1996. "Experiments with the new boosting algorithm." *Machine Learning: Proceeding of the Thirteenth International Conference,* San Francisco, California, 148-156.

Friedl, M. A., C. E. Brodley, and A. H. Strahler. 1999. "Maximizing land cover classification accuracies produced by decision trees at continental to global scales." *IEEE Transactions on Geoscience and Remote Sensing* 37(2): 969-977.

Friedl, M. A., D. K. McIver, J .C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf. 2002. "Global land cover mapping from MODIS: Algorithms and early results." *Remote Sensing of Environment* 83(1-2): 287-302.

Ghimire, B., J. Rogan, and J. Miller. 2010. "Contextual land-cover classification: Incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic." *Remote Sensing Letters* 1(1): 45-54.

Ghimire, B., J. Rogan, V. Rodríguez-Galiano, P. Panday, and N. Neeti. 2012. "An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA." *GIScience & Remote Sensing* 49(5): 623-643.

Guo, Q., M. Kelly, P. Gong, and D. Liu. 2007. "An object-based classification approach in mapping tree mortality using high spatial resolution imagery." *GIScience & Remote Sensing* 44(1): 24-47.

Guo, L., N. Chehata, C. Mallet, and S. Boukir. 2011. "Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests." *ISPRS Journal of Photogrammetry and Remote Sensing* 66: 56-66.

Hansen, M. C., and B. Reed. 2000. "A comparison of the IGBP DISCover and University of Maryland 1 km global land-cover products." *International Journal of Remote Sensing* 21(6-7): 1365-1373.

Haralick, R. M., K. Shanmugam, and I.H. Dinstein. 1973. "Textural features for image classification." *IEEE Transactions on Systems, Man and Cybernetics* 3(6): 610-621.

Haralick, R. M., and K. S. Shanmugam. 1974. "Combined spectral and spatial processing of ERTS imagery data." *Remote Sensing of Environment* 3(1): 3-13.

Hardin, P. J., and C. N. Thomson. 1992. "Fast nearest neighbor classification methods for multispectral imagery." *Professional Geographer* 44(2): 191-202.

Irons, J. R., and R. L. Kennard. 1986. "The utility of Thematic Mapper sensor characteristics for surface mine monitoring." *Photogrammetric Engineering & Remote Sensing* 52: 389-396.

Joachims, T. 1998. "Text categorization with support vector machines: Learning with many relevant features." *Proceedings of European Conference on Machine Learning* Chemnitz, Germany, Aril 21-23, 137-142.

Johansen, K., N. C. Coops, S. E. Gergel, and Y. Stange. 2007. "Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification." *Remote Sensing of Environment* 110: 29-44.

Kaasalainen, S., E. Ahoka, E. Hyyppa, and J. Suomalainen,. 2005. "Study of surface brightness from backscattering laser intensity: Calibration of laser data." *IEEE Geoscience and Remote Sensing Letters* 2(3): 255-259.

Kazar, S. A., and T. A. Warner. 2013. "Assessment of carbon storage and biomass on minelands reclaimed to grassland environments using Landsat spectral indices." *Journal of Applied Remote Sensing* 7(1): 073583. doi: 10.1117/1.JRS.7.073583.

Keene, T. and J. Skousen. 2010. "Mine spoil reclamation with switchgrass for biofuel production." *2010 National Meeting of the American Society for Mining and Reclamation*, Pittsburgh, Pennsylvania, June 5-11, 1-16.

Ke, Y., L. J. Quackenbush, and J. Im. 2010. "Synergistic use of Quickbird multispectral imagery and LiDAR data for object-based forest species classification." *Remote Sensing of Environment* 114: 1141-1154.

Kim, M., M. Madden, and T. A. Warner. 2009. "Forest type mapping using object-specific texture measures from multispectral Ikonos imagery: Segmentation quality and image classification issues." *Photogrammetric Engineering & Remote Sensing* 75(7): 819-829.

Kim, M., T. A. Warner, M. Madden, and D. Atkinson. 2011. "Multi-scale texture segmentation and classification of salt marsh using digital aerial imagery with very high spatial resolution." *International Journal of Remote Sensing* 32: 2825-2850.

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

Lawrence, R., A. Bunn, S. Powell, and M. Zambon. 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, *Remote Sensing of Environment*, 90(3): 331-336.

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

Lillesand, T. M., R. W. Kiefer, and J. W. Chipman. 2008. *Remote Sensing and Image Interpretation*. 6th ed., 756 pp. Hoboken, New Jersey: Wiley & Sons.

Lin, Y., and J. P. Mills. 2010. "Factors influencing pulse width of small footprint, full waveform airborne laser scanning data." *Photogrammetric Engineering & Remote Sensing* 76(1): 49-59.

Liu, D., and F. Xia. 2010. "Assessing object-based classification: Advantages and limitations." *Remote Sensing Letters* 1: 187-194.

Mallinis, G., N. Koutsias, M. Tsakiri-Strati, and M. Karteris. 2008. "Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site." *ISPRS Journal of Photogrammetry and Remote Sensing* 63: 237-250.

Master, L. L., S. R. Flack, and B. A. Stein. 1998. *Rivers of Life: Critical Watersheds for Protecting Freshwater Biodiversity*. 77 pp. Arlington, Virginia: The Nature Conservancy.

Mathieu, R., J. Aryal, and A. K. Chong. 2007. "Object-based classification of Ikonos imagery for mapping large-scale vegetation communities in urban areas." *Sensors* 7: 2860-2880.

Maxwell, A. E., Strager, M. P., Warner, T. A., Zegre, N. P., and C. B. Yuill. 2014a. "Comparison of NAIP orthophotography and RapidEye satellite imagery for mapping of mining and mine reclamation." *GIScience & Remote Sensing* 51(3): 301-320.

Maxwell, A. E., Warner, T. A., Strager, M. P., and M. Pal. 2014b. "Combining RapidEye satellite imagery and LiDAR for mapping of mining and mine reclamation." *Photogrammetric Engineering & Remote Sensing* 80(2): 179-189.

McIver, D. K. and M. A. Friedl. 2002. "Using prior probabilities in decision-tree classification of remotely sensed data." *Remote Sensing of Environment* 81(2-3): 235-261.

Meinel, G., and M. Neubert. 2004. "A comparison of segmentation programs for high resolution remote sensing data." *International Archives of the ISPRS* 35: 1097-1105.

Meneguzzo, D. M., G. C. Liknes, and M. D. Nelson. 2013. "Mapping trees outside of forests using high-resolution aerial imagery: a comparison of pixel- and object-based classification approaches." *Environmental Monitoring and Assessment* 185: 6261-6275.

Merriam, E. R., J. T. Petty, M. P. Strager, A. E. Maxwell, and P. F. Ziemkiewicz. 2013. "Scenario analysis predicts context-dependent stream response to landuse change in a heavily mined Central Appalachian watershed." *Freshwater Science* 32(4): 1246-1259.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2012. "e1071: Misc functions of the department of statistics (e1071)." *R Package Version 1.6-1* (accessed June 1, 2012). http://CRAN.R-project.org/package=e1071.

Muchoney, D., J. Borak, H. Chi, M. Friedl, S. Gopal, J. Hodges, N. Morrow, and A. Strahler. 2000. "Applications of the MODIS global supervised classification model to vegetation and land cover mapping of Central America." *International Journal of Remote Sensing* 21(6-7): 1115-1138.

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

O'Neil-Dunne, J. P. M., S. W. MacFaden, A. R. Royar, and K. C. Pelletier. 2013. "An object-based system for LiDAR data fusion and feature extraction." *Geocarto International* 28(3): 227-242.

Pal, M. 2005. "Random forest classifiers for remote sensing classification." *International Journal of Remote Sensing* 26(1): 217-222.

Pal, M, and P. M. Mather. 2005. "Support vector machines for classification in remote sensing." *International Journal of Remote Sensing* 5(10): 1007-1011.

Palmer, M. A., E. S. Bernhardt, W. H. Schlesinger, K. N. Eshleman, E. Fourfoula-Georgiou, M. S. Hendryx, A. D. Lemly, G. E. Likens, O. L. Loucks, M. E. Power, P. S. White, and P. R. Wilcock. 2010. "Mountaintop mining consequences." *Science* 327: 148-149.

Parks, N. F., G. W. Petersen, and G. M. Baumer. 1987. "High-resolution remote sensing of spatially and spectrally complex coal surface mines of Central Pennsylvania: A comparison between SPOT, MSS, and Landsat-5 Thematic Mapper." *Photogrammetric Engineering & Remote Sensing* 53(4): 415-420.

Platt, R. V., and L. Rapoza. 2008. "An evaluation of an object-oriented paradigm for land use/land cover classification." *The Professional Geographer* 60(1): 87-100. doi:10.1080/00330120701724152.

Pontius, R.G., Jr., and M. Millones. 2011. "Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment." *International Journal of Remote Sensing* 32(15): 4407-4429.

Prakash, A., and R. P. Gupta. 1998. "Land-use mapping and change in a coal mining area: A case study in the Jharia coalfields, India." *International Journal of Remote Sensing* 19(3): 391-410.

Pu, R., S. Landry, and Q. Yu. 2011. "Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery." *International Journal of Remote Sensing* 32: 3285-3308.

Rathore, C. S., and R. Wright. 1993. "Monitoring environmental impacts of surface coal-mining." *International Journal of Remote Sensing* 14(6): 1021-1042.

R Core Development Team. 2012. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (accessed June 1, 2013). http://www.R-project.org.

Rodríguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012a. "An assessment of the effectiveness of a random forest classifier for land-cover classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93-104.

Rodríguez-Galiano, V. F., M. Chica-Olma, F. Abarca-Hernández, P. M. Atkinson, and C. Jeganathan. 2012b. "Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture." *Remote Sensing of Environment* 121: 93-107.

Saylor, K. L. 2008. "Land Cover Trend Project: Central Appalachians." U.S. Department of the Interior, U.S. Geological Survey. Washington, DC (accessed June 1, 2013). http://landcovertrends.usgs.gov/east/eco69 Report.html

Schliep, K., and K. Hechenbichler. 2014. "Package 'kknn'." (Accessed July 7, 2014). http://140.247.115.176/mirrors/cran.r-project.org/web/packages/kknn/kknn.pdf

Sen, S., C. E. Zipper, R. H. Wynne, and P. F., Donovan. 2012. "Identifying revegetated mines as disturbance/recovery trajectories using an interannual Landsat chronosequence." *Photogrammetric Engineering & Remote Sensing* 78(3): 223-235.

Simmons, J. A., W. S. Currie, K. N. Eshleman, K. Kuers, S. Monteleone, T. L. Negley, B. R. Pohlad, and C. L. Thomas. 2008. "Forest to reclaimed mine land use change leads to altered ecosystem structure and function." *Ecological Applications* 18(1): 104-118.

Smith, A. 2010. "Image segmentation scale parameter optimization and land cover classification using Random Forest algorithm." *Journal of Spatial Science* 55: 69-79.

Stal, C., F. Tack, P. De Maeyer, A. De Wulf, and R. Goossens. 2013. "Airborne photogrammetry and LiDAR for DSM extraction and 3D change detection over an urban area- a comparative study." *International Journal of Remote Sensing* 34(4): 1087-1110.

Steele, B. M., J. C. Winne, and R. L. Redmond. 1998. "Estimation and mapping of misclassification probabilities for thematic land-cover maps." *Remote Sensing of Environment* 66(2): 192-202.

Steele, B. M. 2000. "Combining multiple classifiers: An application using spatial and remotely sensed information for land cover mapping." *Remote Sensing of Environment* 74(3): 545-556.

Stehman, S. V., and R. L. Czaplewski. 1998. "Design and analysis for thematic map accuracy assessment: fundamental principles." *Remote Sensing of Environment* 64: 331-344.

Stein, B. A., L. S. Kutner, and J. S. Adams. 2000. *Precious Heritage: The Status of Biodiversity in the United States*. 399 pp. New York, New York: Oxford University Press.

Townsend, P. A., D. P. Helmers, C. C. Kingdon, B. E. McNeil, K. M. de Beurs, and K. N. Eshleman. 2009. "Changes in the extent of surface mining and reclamation in the Central Appalachians detected using a 1976-2006 Landsat time series." *Remote Sensing of Environment* 113: 62-72.

Trimble, 2011. *eCognition Developer 8.64.1 User Guide*. Trimble, Munich, Germany.

Tso, B., and P. Mather. 2003. *Classification Methods for Remotely Sensed Data*. 352 pp. New York, New York: CRC Press.

Tyc, G., J. Tulip, D. Schulten, M. Krischke, and M. Oxfort. 2005. "The RapidEye mission design." *Acta Astronautica* 56(1): 213-219.

Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. 188 pp. New York, New York: Springer-Verlag.

Walter, V. 2004. "Object-based classification of remote sensing data for change detection." *ISPRS Journal of Photogrammetry and Remote Sensing* 58(3-4): 225-238.

Warner, T. 2011. "Kernel-based texture in remote sensing image classification." *Geography Compass* 5(10): 781-798.

Warner, T. A., and F. Nerry. 2009. "Does a single broadband or multispectral thermal data add information for classification of visible, near- and shortwave infrared imagery of urban areas?." *International Journal of Remote Sensing* 30(9): 2155-2171.

Wickham, J. D., K. H. Riitters, T. G. Wade, M. Coan, and C. Homer. 2007. "The effect of Appalachian mountaintop mining on interior forest." *Landscape Ecology* 22: 179-187.

Yuill, C., 2003. "Landscape use assessment: Mountaintop mining and the mountaintop mining region of West Virginia." *Draft Programmatic Environmental Impact Statement on Mountaintop Mining/Valley Fills in Appalachia* P. III, F-12.

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

Zégre, N. P., A. J. Miller, A. Maxwell, and S. J. Lamont. 2014. "Multiscale analysis of hydrology in a mountaintop min-impacted watershed." *Journal of the American Water Resources Association* doi: 10.1111/jawr.12184.

Zhou, W. 2013. “An object-based approach for urban land cover classification: Integrating LiDAR height and intensity data.” *Geoscience and Remote Sensing Letters, IEEE* 10(4): 928-931.

Table 1: Input predictor variables for GEOBIA classification.

Classification method	Total number of variables	Variables		
		Spectral and texture bands	Object geometry	LiDAR
GEOBIA	74	For each of the 5 spectral bands: Mean SD GLCM 2nd angular moment GLCM contrast GLCM correlation GLCM dissimilarity GLCM entropy GLCM mean GLCM SD GLCM homogeneity	Area Asymmetry Border index Border length Compactness Density Elliptic fit Length Length/width Rectangular fit Roundness Shape index Volume Width	nDSM mean nDSM SD nDSM minimum nDSM maximum nDSM range Intensity mean Intensity SD Intensity minimum Intensity maximum Intensity range

Table 2: Land cover class definitions.

Class	Description
Forested	Land dominated by mature, woody vegetation that has not been recently disturbed by surface mining; mature forest that generally represents pre-mining conditions of the slopes
Reclaimed-herbaceous vegetation	Reclaimed areas dominated by herbaceous/non-woody vegetation
Reclaimed-woody vegetation	Reclaimed areas dominated by clumped or clustered woody plants that include shrubs and immature trees
Barren	Barren land lacking vegetation; manmade structures; haul roads; active quarries; lands disturbed by mining
Water	Water, including retention ponds, streams, and standing water

Table 3: Number of training objects.

Land cover class	**Number of training objects for GEOBIA classification**
Forested	230
Reclaimed-herbaceous vegetation	241
Reclaimed-woody vegetation	82
Barren	342
Water	26
Total	**921**

Table 4: Comparison of classification accuracy using different input variable combinations for GEOBIA.

Classification method	Accuracy measure	Predictor variables					
		Spectral means	**Spectral means + geometry**	**Spectral means + SDs**	**Spectral means + GLCM measures**	**Spectral means + texture PCA**	**Spectral means + LiDAR**
***k*-NN**	OA (%)	83.1	70.7	79.0	74.3	80.0	85.9
	AD (%)	9.6	15.3	13.1	14.0	12.7	8.7
	QD (%)	7.3	14.0	7.9	11.8	7.3	5.3
SVM	OA (%)	84.5	80.9	84.0	80.6	80.8	86.6
	AD (%)	8.2	11.4	6.5	9.6	6.5	8.8
	QD (%)	7.3	7.7	9.5	9.8	12.8	4.6
RF	OA (%)	81.8	79.1	81.3	80.2	80.4	85.6
	AD (%)	8.1	7.9	8.0	10.1	8.0	6.9
	QD (%)	10.1	13.1	10.6	9.7	11.6	7.5
Boosted CART	OA (%)	82.6	83.1	80.7	82.9	81.4	86.6
	AD (%)	9.4	8.6	9.8	8.6	8.9	6.5
	QD (%)	8.0	8.3	9.5	8.6	9.7	6.9

Note: OA = Overall accuracy (%), AD = Allocation disagreement (%), and QD = Quantity disagreement (%)

Table 5: *Z*-scores comparing the classification to a classification using just the spectral means.

Classification method	***z*-score compared to classification using just the spectral means**				
	Spectral means + geometry	**Spectral means + SDs**	**Spectral means + GLCM measures**	**Spectral means + texture PCA**	**Spectral means + LiDAR**
***k*-NN**	8.828*	3.593*	6.455*	2.582*	2.497*
SVM	3.862*	0.560	3.847*	3.685*	1.983*
RF	4.573*	0.790	1.941*	2.424*	3.729*
Boosted CART	0.863	2.801*	0.442	1.784*	3.544*

Note: A *z*-score larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is more accurate than the other.

Table 6: Error matrix for GEOBIA classification using SVM with imagery bands and LiDAR-derivatives based on relative area of the sampled objects. Overall accuracy is 86.6%

		Reference data						
		Barren	Forested	Reclaimed-herbaceous vegetation	Reclaimed-woody vegetation	Water	Totals	User's accuracy
Classified data	Barren	0.136	0.000	0.028	0.000	0.002	0.167	81.6%
	Forested	0.004	0.339	0.015	0.023	0.001	0.383	88.6%
	Reclaimed-herbaceous vegetation	0.022	0.000	0.258	0.014	0.000	0.293	87.8%
	Reclaimed-woody vegetation	0.001	0.001	0.021	0.126	0.000	0.149	84.7%
	Water	0.001	0.000	0.000	0.000	0.007	0.008	83.5%
	Totals	0.164	0.340	0.323	0.163	0.011		
	Producer's accuracy	83.3%	99.8%	79.8%	77.3%	63.7%		

Table 7: *Z*-score comparing k-NN to other classifiers using band means and band means + LiDAR.

		***z*-score**		
		SVM compared to *k*-NN	**RF compared to *k*-NN**	**Booted CART compared to *k*-NN**
Input variables	Sepctral means	2.034*	1.639	0.773
	Spectral means + LiDAR	1.002	0.537	0.756

Note: A *z*-score larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is more accurate than the other.

Table 8: *Z*-score comparing SVM to other classifiers using band means and band means + LiDAR.

		***z*-score**		
		***k*-NN compared to SVM**	**RF compared to SVM**	**Booted CART compared to SVM**
Input Variables	Spectral means	2.034*	3.506*	2.814*
	Spectral means + LiDAR	1.326	0.011	1.002

Note: A *z*-score larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is more accurate than the other.

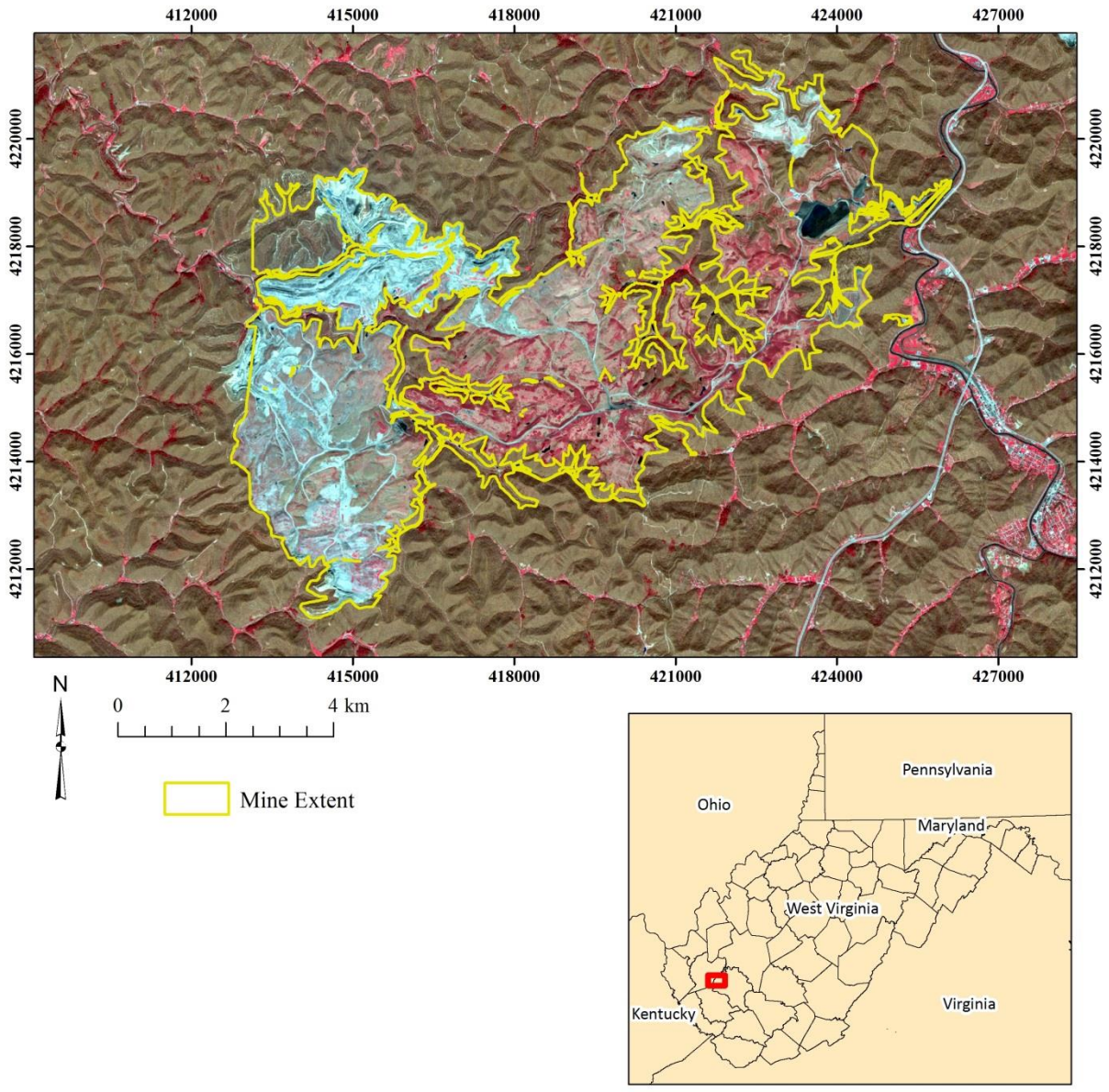

Figure 1: Hobet-21 surface mine complex. Base image is RapidEye scene acquired on 1 April 2010 and displayed in simulated color infrared (Bands 5, 3, 2 as red, green, and blue) (© (2015) BlackBridge S.àr.l. All rights reserved). The depicted mine extent is based on a surface mining permit obtained from the West Virginia Department of Environmental Protection (WVDEP). The map is projected in NAD83 UTM Zone 17 N.

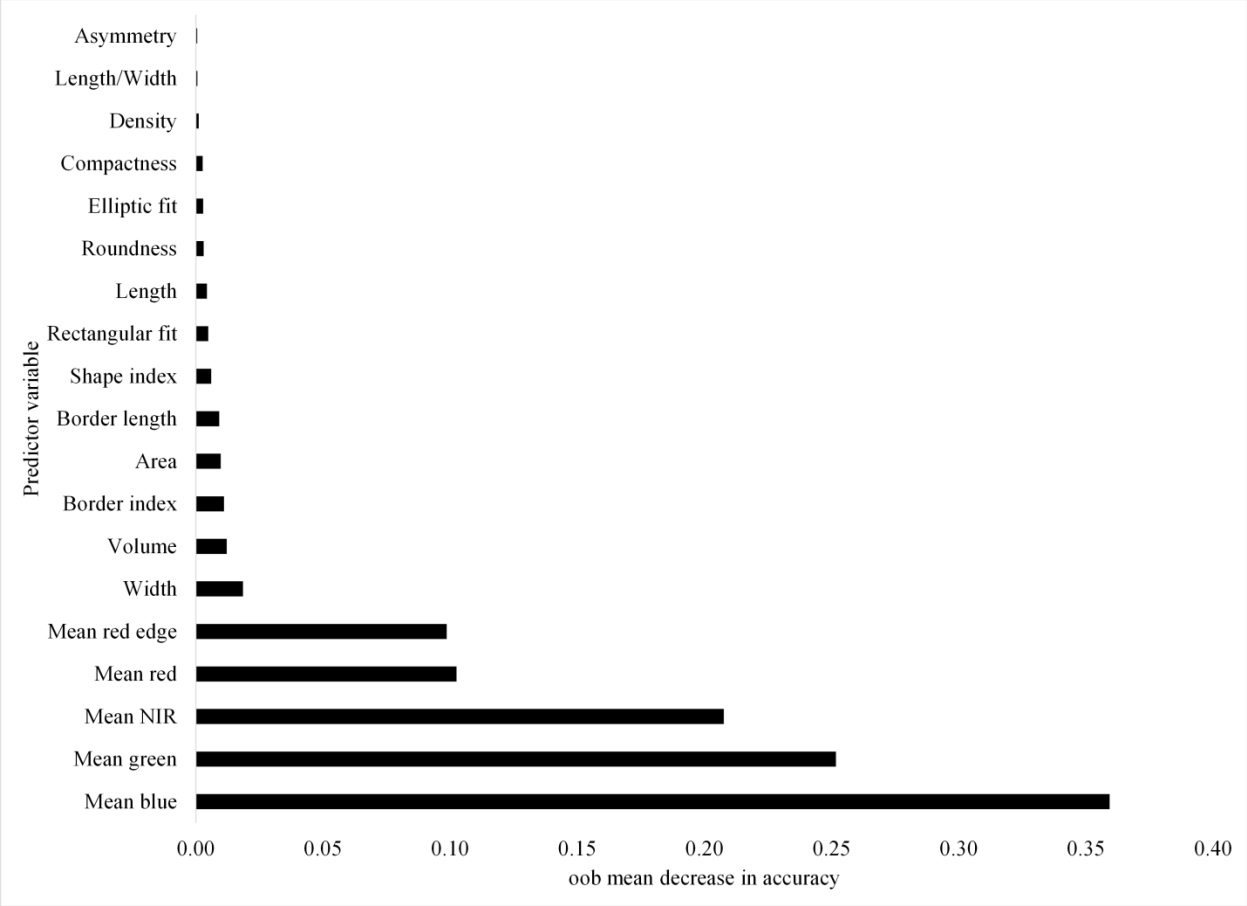

Figure 2: Relative importance of predictor variables as estimated by oob mean decrease in accuracy by RF for model using object band means and object geometry variables.

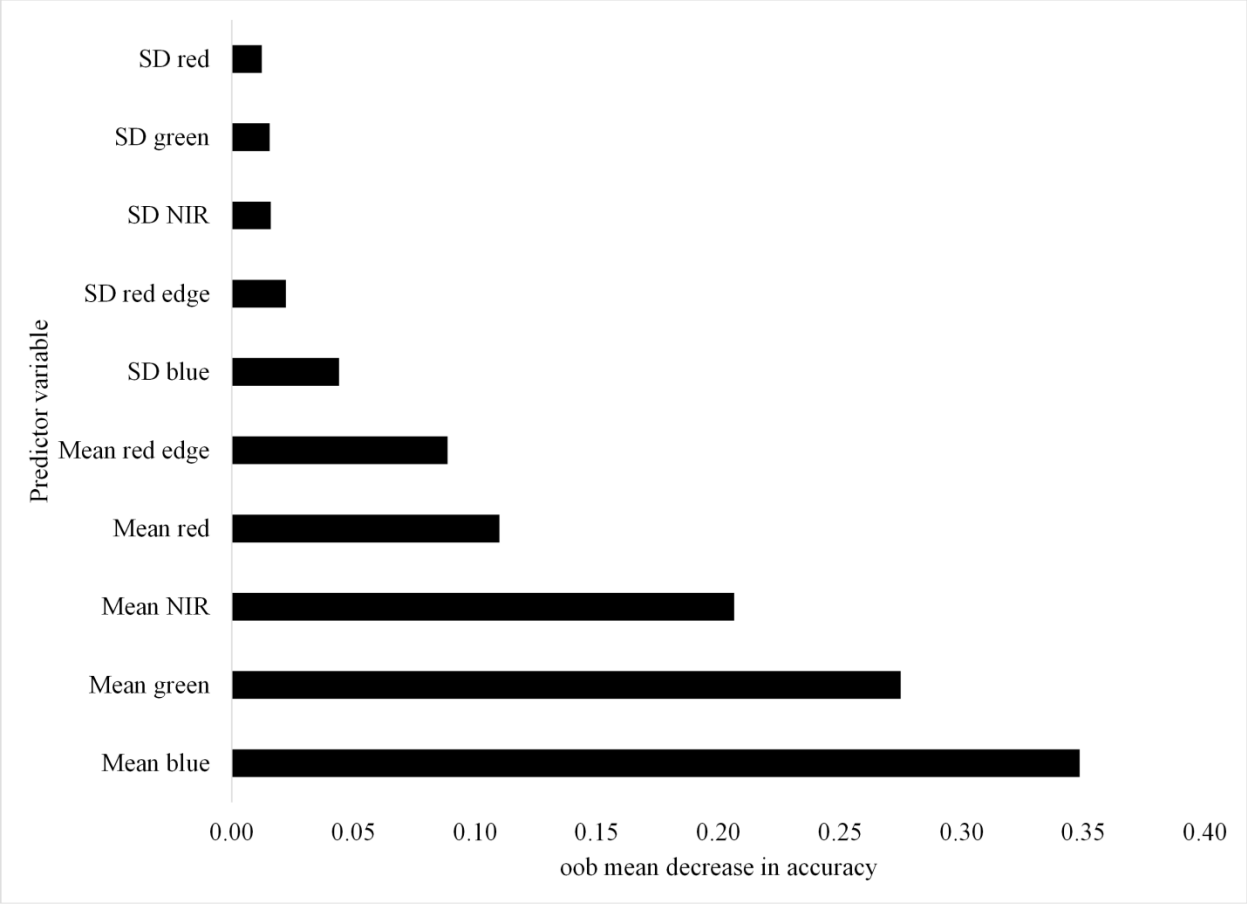

Figure 3: Relative importance of predictor variables as estimated by oob mean decrease in accuracy by RF for model using object spectral means and SDs.

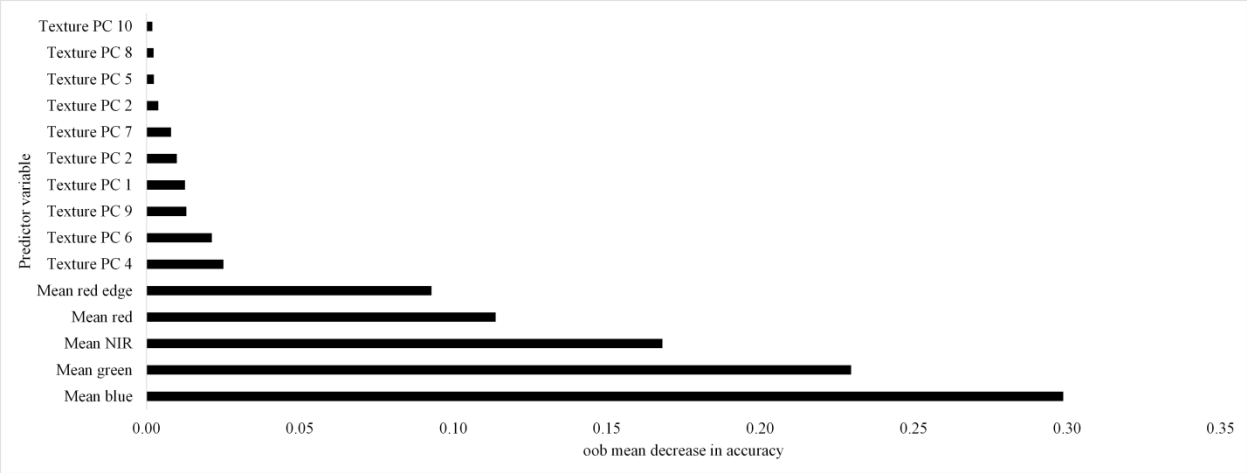

Figure 4: Relative importance of predictor variables as estimated by oob mean decrease in accuracy by RF for model using spectral data as spectral means and texture principle components (PCs).

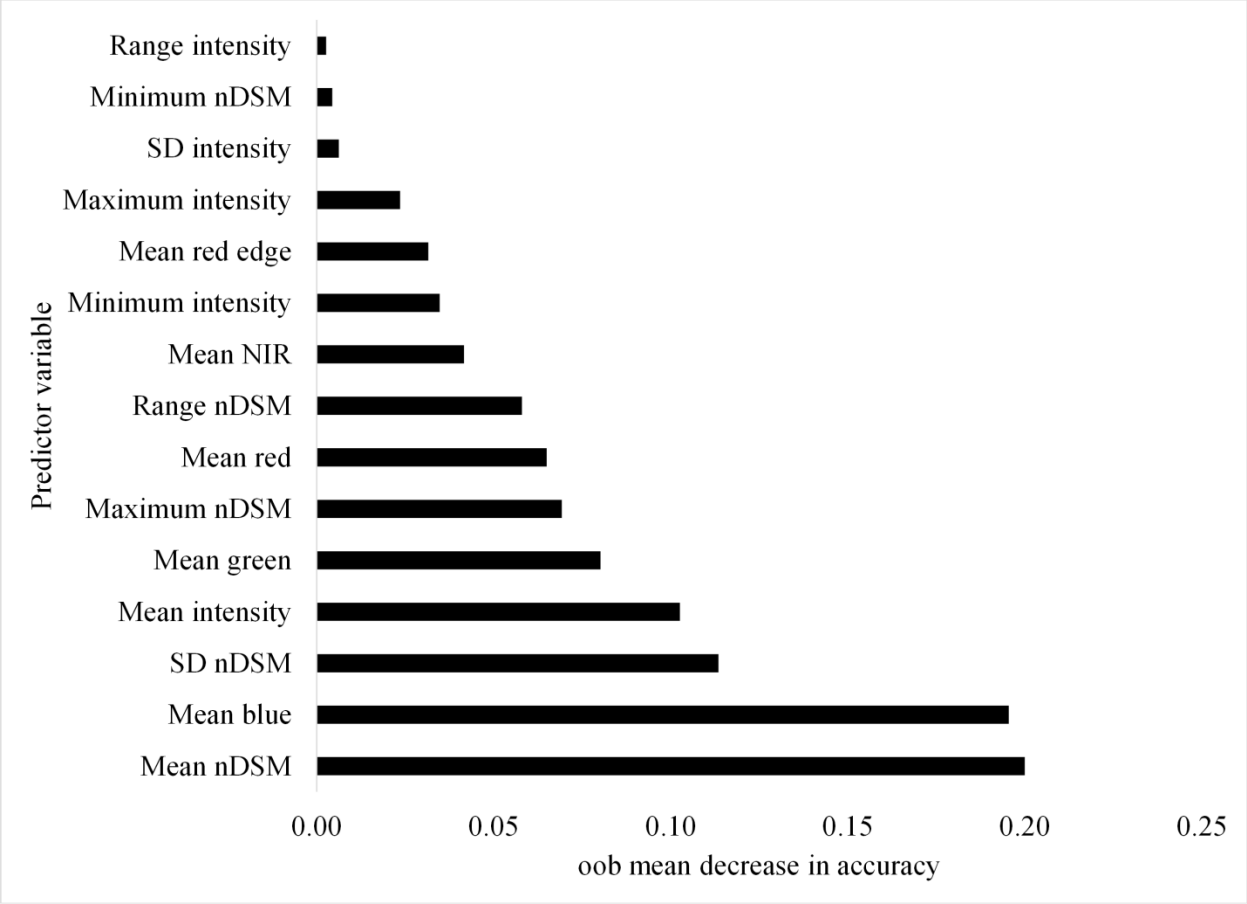

Figure 5: Relative importance of predictor variables as estimated by oob mean decrease in accuracy by RF for model using object spectral means and descriptive statistics derived from LiDAR

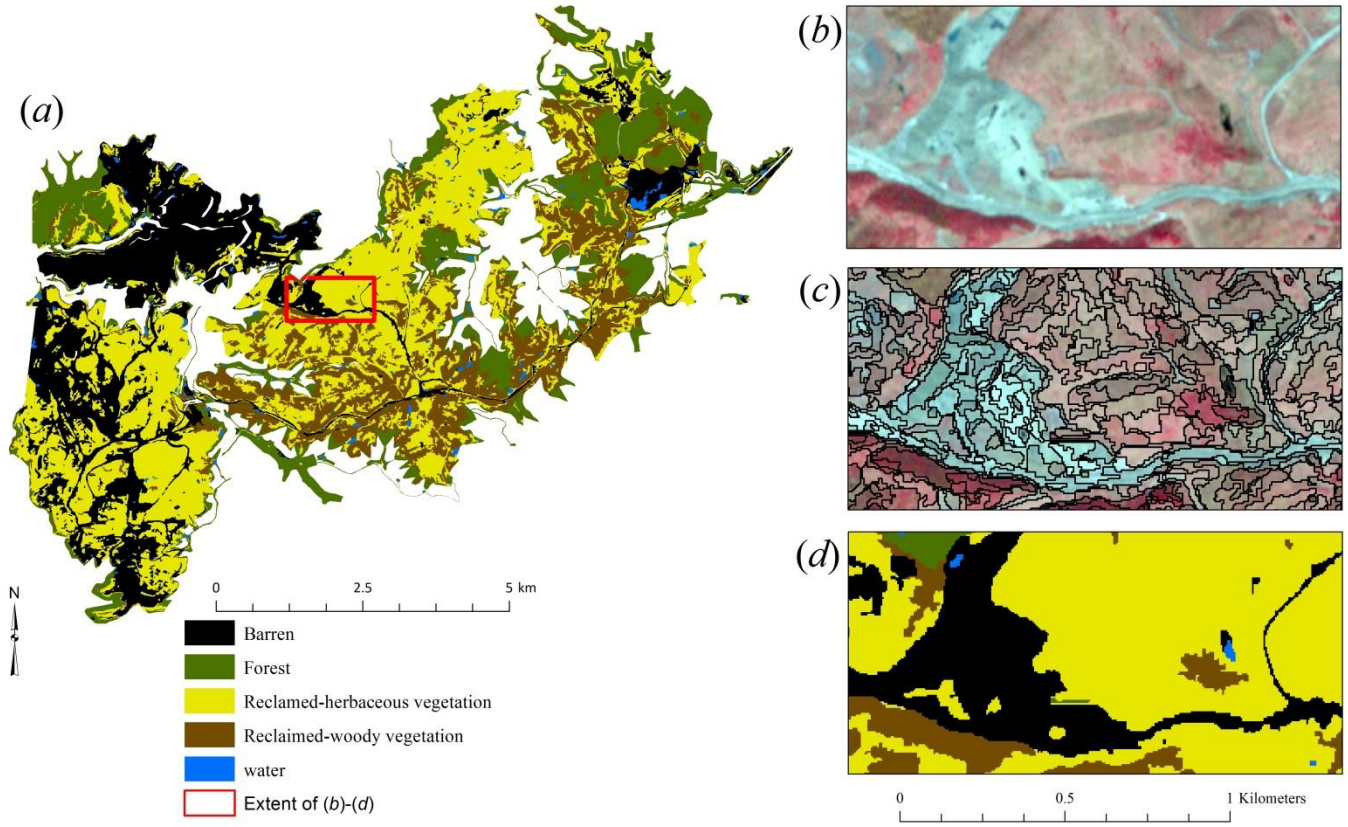

Figure 6: (*a*) GEOBIA land cover classification of the Hobet-21 mine complex using all RapidEye imagery band means and 10 LiDAR-derived descriptive statistics. SVM was used with a radial basis function (RBF) kernel, a gamma value (γ) of 0.0001, and a cost value (C) of 1300. Refer to Figure 1 for location information. (*b*) Example of RapidEye data (© (2015) BlackBridge S.àr.l. All rights reserved). (*c*) Example of image segmentation. (*d*) Example of classification results.

CHAPTER 5

Differentiating mine-reclaimed grasslands from spectrally similar land cover using terrain variables and object-based machine learning classification[1]

Aaron E. Maxwell and Timothy A. Warner

Abstract

Incorporating ancillary, non-spectral data may improve the separability of land use/land cover (LULC) classes. This study investigates the use of multi-temporal digital terrain data combined with aerial National Agriculture Imagery Program (NAIP) imagery for differentiating mine-reclaimed grasslands from non-mining grasslands across a broad region (6085 km^2). The terrain data were derived from historical digital hypsography and a recent light detection and ranging (LiDAR) data set. A geographic object-based image analysis (GEOBIA) approach, combined with machine learning algorithms, random forests (RF) and support vector machines (SVM), was used because these methods facilitate the use of ancillary data in classification. The results suggest that mine-reclaimed grasslands can be mapped accurately, with user's and producer's accuracies above 80%, due to a distinctive topographic signature in comparison to other spectrally similar grasslands within this landscape. The use of multi-temporal digital elevation model (DEM) data and pre-mining terrain data only generally provided statistically significant increased classification accuracy in comparison to post-mining terrain data. Elevation change data were of value, and terrain

[1] This is an Accepted Manuscript of an article scheduled to be published by Taylor & Francis in the *International Journal of Remote Sensing*, available online: http://www.tandfonline.com/doi/abs/10.1080/01431161.2015.1083632

Maxwell, A.E., and T.A. Warner, 2015. Differentiating mine-reclaimed grasslands from spectrally similar land cover using terrain variables and object-based machine learning classification, *International Journal of Remote Sensing* (In Press). (Received 6 April 2015, Accepted 11 August 2015)

shape variables generally improved the classification. GEOBIA and machine learning algorithms were useful in exploiting this non-spectral data, as data gridded at variable cell sizes can be summarized at the scale of image objects, allowing complex interactions between predictor variables to be characterised.

Keywords: GEOBIA, object-based classification, machine learning, land cover mapping, LULC mapping, LiDAR, hypsography

1. Introduction

Mapping land use change is important for studies of anthropogenic global change (Anderson et al. 1976; Folke et al. 1998; Walker 1998; Cihlar and Jansen 2001). Nevertheless, some land use/land cover (LULC) classes may not be spectrally distinctive, resulting in low classification accuracy when multispectral data are used to produce a thematic map. This is especially true when attempting to map land use classes, since the use of land does not necessarily result in spectrally distinctive properties. Ancillary, non-spectral data may help differentiate these spectrally similar LULC classes (Gislason, Benediktsson, and Sveinsoon 2006; Treitz and Howarth 2000; Knight et al. 2013).

Mine reclamation is an example of a land use that may not be spectrally distinguishable in aerial and satellite imagery. Mountaintop removal with valley fills (MTR/VF) mining is a resource extraction approach practiced in the Appalachian region of the United States of America (USA), especially in southern West Virginia, eastern Kentucky, and southwestern Virginia. In this region, heavy machinery and explosives are used to expose coal seams of the Pennsylvanian geologic subperiod. MTR/VF mining results in faster and more pervasive terrain alteration than more traditional mining techniques (Fritz et al. 2010). Excavation and the subsequent reclamation associated with MTR/VF results in considerable physical terrain alteration, as 50 to 200 m of rock material is commonly removed from mountaintops and the unconsolidated rock waste is disposed of in the adjacent valleys as so-called valley fills, filling

headwater streams and generally raising valley elevations (Hooke 1994; Hooke 1999; Fritz et al. 2010; Palmer et al. 2010; Bernhardt and Palmer 2011; Bernhardt et al. 2012; Maxwell and Strager 2013). This practice alters the pre-mining landforms of the affected mountaintops and valleys, flattening the upper slopes of a landscape that was originally characterised by moderate to strong relief and steep slopes dissected by a dendritic stream network (Ehlke et al. 1982; Maxwell and Strager 2013). The resulting topographically altered terrain, which was generally forested before mining, is commonly reclaimed to grasslands or shrublands (Simmons et al. 2008; Kazar and Warner 2013). Because this mining practice results in such characteristic topographic alteration, multi-temporal terrain data may facilitate the differentiation of MTR/VF reclaimed grasslands from other spectrally similar grasslands within this landscape.

In this paper, we explore the use of multi-temporal, digital elevation model (DEM)-derived terrain data combined with high resolution aerial orthophotography for differentiating mine-reclaimed grasslands from other grasslands. The study site comprises three watersheds covering 6085 km^2 in West Virginia, USA, a region where extensive landscape alteration has taken place due to surface coal mining, epecially MTR/VF. A geographic object-based image analysis (GEOBIA) approach and machine learning algorithms are used to integrate the disparate data at differing scales. The following research questions are addressed:

1. Can mine reclaimed grasslands be separated from other grasslands across broad regions using DEM-derived terrain characteristics?
2. Is it necessary to use both pre- and post-mining terrain characteristics to obtain an accurate separation of these classes? Or, can an accurate separation be obtained using only topography from a single time period (e.g. the current landscape)?

3. Do derived terrain attributes help differentiate these grassland classes, and, if so, what terrain attributes are most important?

2. Background

2.1. Importance of mapping mine reclaimed grasslands

Grasslands resulting from surface mine reclamation have been shown to be fundamentally different from other grasslands in terms of their impact on hydrology (Negley and Eshelman 2006; Ferrari et al. 2009; McCormick et al. 2009; Zégre, Maxwell, and Lamont 2013; Miller and Zégre 2014; Zégre et al. 2014), terrestrial habitat (Weakland and Wood 2005; Wood, Bosworth, and Dettmers 2006; Simmons et al. 2008; Wickham et al. 2007; Wickham et al. 2013), and aquatic ecosystems (Hartman et al. 2005; Pond et al. 2008; Fritz et al. 2010; Pond, 2010; Merriam et al. 2011; Bernhardt et al. 2012; Merriam et al. 2013). Negley and Eshleman (2006) found that watersheds affected by mining and mine reclamation produce increased storm runoff and higher peak hourly runoff rates for storm events in comparison to watersheds not affected by mining. These observations were attributed to the loss of tree canopy and reduced evapotranspiration as well as decreased infiltration due to soil compaction. However, in a review of the hydrologic impacts of MTR/VF mining and mine reclamation, Miller and Zégre (2014) suggest that hydrology of such systems are not well understood. Although traditional mining practices generally increase peak and total runoff, the hydrologic impacts of MTR/VF reclamation are confounded by the increased storage of water in valley fill spoil and the reduced infiltration resulting from the compaction of soils above the fill.

Concerning the effect on terrestrial habitats, Wood et al. (2006) suggest that mine reclamation and loss of forest negatively affect Cerulean Warbler (*Dendroica cerulean*)

populations, a species of conservation concern. Simmons et al. (2008) document nutrient limitations within terrestrial ecosystems impacted by mine reclamation that may persist for decades or centuries. Within aquatic ecosystems, Pond (2010) found that the number and richness of assemblages of mayflies (Ephemeroptera), especially sensitive aquatic insect taxa, were reduced in streams impaired by mining in comparison to reference sites. Merriam et al. (2013) found a direct correlation between selenium (Se) concentrations in streams and the extent of surface mining and reclamation upstream. Thus, because mine reclamation has unique and profound impacts on hydrological processes, terrestrial habitat, and aquatic ecosystems, it is important to be able to map and differentiate such land use from spectrally similar classes. In particular, information on the extent and location within the modified topographic landscape of reclaimed grasslands is a foundational data layer for environmental studies of MTR/VF landscapes.

2.2. Terrain data for mapping and modeling

Terrain data have been integrated into LULC mapping in many previous studies to improve classification accuracy. For example, Gislason, Benediktsson, and Sveinsson (2006) combined elevation, topographic slope, and topographic aspect derived from DEM data with Landsat Multispectral Scanner (MSS) data for mapping forest types in Colorado, USA, and noted the value of elevation in the classification. Treitz and Howarth (2000) found DEM data improve classification accuracy for forest ecosystems in northern Ontario, Canada. Although terrain information alone provided a weak separation of the forest ecosystem classes, combining these data with spectral data improved the classification. Knight et al. (2013) also noted an improvement in classification accuracy when topographic derivatives such as compound topographic moisture index (CTMI), topographic slope, and slope curvature were combined with

spectral data for mapping palustrine wetlands. Terrain data have also been used as predictor variables for spatial modeling (for example, Prasad, Iverson, and Liaw 2006; Wright and Gallant 2007; Pino-Mejías et al. 2010; Evans and Kiesecker 2014). However, a review of the current literature suggests that multi-temporal terrain data have not been explored for mapping LULC classes within landscapes characterized by extensive anthropogenic topographic alteration.

2.3. Mapping surface mining and reclamation

Because mine reclamation generally persists as a legacy landscape alteration, mapping reclamation is of particular interest. However, as Rathore and Wright (1993) note, mine reclamation has proven more difficult to map than active mining. Research investigating the mapping of LULC resulting from mining and mine reclamation has traditionally focused on moderate spatial resolution multispectral data, such as MSS, Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Satellite Pour l'Observation de la Terre (SPOT) data (Anderson and Schubert 1976; Irons and Kennard 1986; Parks, Petersen, and Baumer 1987; Rathore and Wright 1993; Anderson et al. 1997; Prakash and Gupta 1998; Yuill 2003; Townsend et al. 2009; Sen et al. 2012). In contrast, our previous research has explored the use of high spatial resolution satellite and aerial data, the combination of spectral and light detection and ranging (LiDAR) data, and the implementation of GEOBIA and machine learning algorithms for mapping mining and mine reclamation at the scale of a single mine (Maxwell et al. 2014a; Maxwell et al. 2014b; Maxwell et al. 2015). This research expands upon our previous work by focusing particularly on the potential benefit of multi-temporal terrain data, integrated with high resolution aerial imagery, for differentiating mine-reclaimed grasslands for regional mapping of reclamation. Indeed, to our knowledge, this is the first paper on mapping grasslands associated with mine reclamation at a fine resolution (5 m) across a broad region.

A notable example of previous mapping of mining and mining reclamation with moderate scale data is the work of Townsend et al. (2009), who developed a classification approach for mapping mining and mine reclamation across a region encompassing eight river basins in the Central Appalachian Mountain region of the Eastern United States. Four Landsat MSS, TM, and ETM+ scenes from 1976, 1987, 1999, and 2006 were classified using ISODATA clustering to map spectrally separable land cover classes. They then used ancillary data to produce a mine mask and a decision tree process utilising characteristic transitions of land cover to separate mining and mine reclamation from other classes within the mine mask. Accuracies for mapping mining and mine reclamation were generally above 85% using this method. Sen et al. (2012) expanded upon this work using a time series of Landsat TM and ETM+ data to differentiate re-vegetated mines from other forest-displacing disturbance such as urbanization, using disturbance and subsequent recovery trajectories and a GEOBIA approach across four counties impacted by MTR/VF mining in southwestern Virginia. An accuracy of 89% was obtained.

These previous studies suggest that reclaimed mine lands can be separated from spectrally similar land cover using multi-temporal data. However, a time series is not commonly available when working with high resolution satellite or aerial data, suggesting that other methods must be explored if mapping is to be undertaken at a high spatial resolution.

2.4. GEOBIA and machine learning algorithms

GEOBIA, the process of segmenting an image into objects, or contiguous groups of pixels that are relatively spectrally homogeneous, and labeling each resulting object as a single unit, has been described as a paradigm shift in remote sensing (Blaschkie et al. 2014). GEOBIA has been shown to be particularly applicable for the classification of high spatial resolution data (Blaschke and Strobl 2001; Walter 2004; Chubey, Franklin, and Wulder 2006; Drăgut and

Blaschke 2006; Blaschke 2010; Baker et al. 2013; Meneguzzo, Liknes, and Nelson 2013). A feature of GEOBIA of great significance for incorporating information from ancillary layers is that the spatial support (i.e. data structure and resolution) of the ancillary data does not need to be the same as the pixel grid used to develop the objects.

Once image objects are produced, a wide variety of summary spectral (e.g. mean, standard deviation, range, minimum, maximum of image bands) and spatial (e.g. shape, size, and association with neighboring objects) properties can be summarized for each object (Trimble 2011). However, many of these variables may not meet the assumptions of multivariate normality required for statistical classifiers, and, as a result, nonparametric, machine learning approaches are commonly used in GEOBIA (Ke, Quakenbush, and Im 2010; Trimble 2011; Duro, Franklin, and Dubé 2012a; Duro, Franklin, and Dubé 2012b; Guan et al. 2013). Machine learning algorithms offer the potential to handle high dimensional complex spectral measurement spaces and large volumes of data, with the added benefit of reduced processing time compared to traditional classifiers (Hansen and Reed 2000). In this study, two machine learning algorithms were used, support vector machines (SVM) (Vapnik 1995; Joachims 1998; Burges 1998; Tso and Mather 2003; Pal and Mather 2005; Pal 2005; Warner and Nerry 2009) and random forests (RF) (Breiman 2001).

SVMs separate two classes by constructing a multi-dimensional hyperplane that is optimized as the maximum margin that provides the best separation between the classes. To create this decision boundary, it is usually necessary to transform the data to a higher dimensional space in order for the data to be linearly separable. This is accomplished using a kernel function, such as a polynomial or radial basis function (RBF). To facilitate generalization of the decision boundary, a penalty parameter (C) penalizes training samples located on the

"wrong" side of the decision boundary (Vapnik 1995; Joachims 1998; Burges 1998; Tso and Mather 2003; Pal and Mather 2005; Pal 2005; Warner and Nerry 2009). SVM algorithms were originally designed for two class problems, and, as a result, strategies are required to allow for the separation of more than two classes. For example, the "one-against-one" approach uses binary classifiers and a voting scheme to separate multiple classes (Vapnik 1995; Tso and Mather 2003; Pal 2005; Pal and Mather 2005; Meyer et al. 2012).

RF uses an ensemble of classification trees to improve upon the accuracy and consistency of single decision tree (DT) classifications. RF differs from other ensemble DT methods because each tree is generated from a subsample of the data obtained from random bootstrap sampling of the training data with replacement, a process known as bagging (Breiman 1996; Breiman 2001). The withheld, or out-of-bag (oob), samples can be used for map accuracy assessment, assuming the training data were collected in a random and unbiased manner. Also, a random subset of the predictor variables (the number of which is defined by the user) is used for growing each tree in the ensemble. This is done to decrease the correlation between trees, and thereby decrease the generalization error (Breiman 2001). RF has many attributes that make it attractive for classification, including the capacity to model complex interactions between predictor variables, handling data with missing values, generating high classification accuracies, and providing measures of predictor variable importance (Steele 2000; Cutler et al. 2007).

3. Study area

The study area was defined relative to Hydrologic Unit Code (HUC 8) watershed extents within the MTR/VF region of West Virginia, USA (Figure 1). Three adjacent watersheds were mapped: the Upper Kanawha, Upper Guyandotte, and Coal River, totaling 6085 km^2. These watersheds were selected due to the availability of pre- and post-mining terrain data, digital mine

permit extents, and aerial orthophotography. Also, a prior land cover analysis found that within these watersheds surface mining and mine reclamation comprise a major component of the landscape, as much as 6% of the surface area (Maxwell et al. 2011). Watersheds were used to define the study area boundary because watersheds tend to be used as the spatial unit for both management and environmental research.

4. Methods

4.1. Overview of mapping process

Prior to a detailed description of the methods, we first give a brief overview of the mapping process (Figure 2). The classification is hierarchical, with two stages. First, after pre-processing, the aerial orthophotography was classified using a GEOBIA approach, in which the imagery was segmented and then classified using SVM to produce four classes: woody vegetation, herbaceous vegetation, barren areas, and water. The resulting land cover classification was then generalized using a sieving operation to remove land cover patches less than 1 ha, the minimum mapping unit (MMU) for the study. Contiguous areas of herbaceous cover (i.e. grasses) were then merged as single objects for the second stage of the classification, in which the RF algorithm along with pre- and post-mining terrain variables were used to differentiate mine reclaimed grasslands from other grasslands. The results were then assessed using randomized validation data. The following sections elaborate on these methods.

4.2. Input data and pre-processing

National Agriculture Imagery Program (NAIP) orthophotography was the primary image data used in this study. The images were collected during the growing season of 2011 between 10 July 2011 and 6 October 2011 with an Integraph Z/I Imaging Digital Mapping Camera (DMC). The data were provided at a 1 m ground sampling distance (GSD) with four spectral

bands (blue, green, red, and near-infrared (NIR)) by the United States Department of Agriculture (USDA) Farm Service Agency. NAIP orthophotography has been used for LULC classification in previous studies (for example, Baker et al. 2013; Davies et al. 2010; Meneguzzo Liknes, and Nelson 2013; Maxwell et al. 2014a). In a previous study (Maxwell et al. 2014a) using NAIP imagery, and focusing only on land cover classes within a single mine in this region, we found that classification accuracies were above 90% when the number of classes was limited and the spatial resolution was decreased from 1 m to 5 m. In order to prepare the imagery for segmentation and classification, each uncompressed quarter quadrangle was resampled to a 5 m cell size using pixel aggregation (i.e. average of the input cells) within Erdas Imagine 2014 (ERDAS 2013). The resampled quarter quadrangles were then mosaicked to produce a single image for the entire study area.

A pre-mining, historic DEM was produced from United States Geologic Survey (USGS) digital line graph (DLG) contour data derived from 1:24,000 topographic maps. These data do not represent a single date; source data range from 1951 to 1989, with the majority of the data representing topographic conditions of the 1960s and 1970s. A review of available terrain data for this region suggested that this is the most appropriate historic elevation data set for this analysis, as pre-mining DEM data are limited. Further, visual inspection of a hillshade image produced from these data and an elevation change image produced by subtracting the historic and recent DEMs both suggest that the DLG data predate almost all large scale MTR/VF activity in the study area, which began as early as the 1960s but was not widespread until the 1990s (Milici 2000; Wickham et al. 2013). The contour data were gridded on a 9 m raster using the Topo to Raster tool in ArcMap 10.2 (ESRI 2012). A 9 m cell size was chosen, as opposed to a 5 m cell size to match the image data, to reflect the inherent resolution of the DLG data.

A recent, post-mining DEM was made available by the West Virginia Department of Environmental Protection (WVDEP). This DEM was produced using aerial LiDAR, which is an active remote sensing technique that uses the two-way travel time of emitted laser pulses and precise geolocation derived from differential global positioning system (GPS) and inertial measurement unit (IMU) data to calculate the elevation of the ground surface and the height of objects above the ground surface (Hyyppä et al. 2009). The LiDAR data were collected between 9 April 2010 and 29 March 2011 during leaf-off conditions to maximize the number of ground returns. Flight specifications were selected to support a nominal average pulse spacing of 1 m. The Optech ALTM 3100 C sensor was set to a pulse frequency of 70 kHz, a scan frequency of 35 Hz, and a scan angle of 36˚ (full swath). A 30% overlap was acquired between swaths. The aircraft flew at an average of 1524 m above ground level and at a speed of 125 knots (232 km hr^{-1}). The LiDAR system recorded up to four returns per laser pulse, and each return was classified by the vendor as ground, non-ground, or as an outlier, and delivered in LAS 1.2 format. The DEM provided by the WVDEP was resampled using pixel aggregation to a 9 m cell size, to match that of the pre-mining DEM.

4.3. Image segmentation and classification

The NAIP orthophotography was segmented using the multi-resolution image segmentation algorithm within eCognition 8.0 (Trimble, Sunnydale, California). This algorithm requires the user to define three parameters: scale, shape, and compactness. The scale parameter controls the size of the image objects (Liu and Xia 2010; Kim et al. 2011), and a number of studies have suggested that this parameter has the largest impact on subsequent classification accuracy (Blaschke 2003; Meinel and Neubert 2004; Kim, Madden, and Warner 2009; Liu and Xia 2010; Smith 2010; Myint et al. 2011). The shape parameter controls the relative importance

assigned to the shape of the object versus the "color," which relates to spectral properties. Compactness controls the balance between the edge length and form of the object (Baatz and Schäpe 2000). As is common in GEOBIA (e.g. Laliberte, Fredrickson, and Rango 2007; Mathieu, Aryal, and Chong 2007, Dingle Robertson and King 2011; Myint et al. 2011; Pu, Landry, and Yu 2011; Duro, Franklin and Dubé 2012a; Duro, Franklin and Dubé 2012b) trial-and-error and expert judgment were used to select the optimal settings of 30 for scale, 0.1 for shape, and 0.5 for compactness. All four image bands were equally weighted in the segmentation. Over 500,000 objects were generated.

The resulting image objects were classified using the implementation of the SVM algorithm available in the e1071 package (Meyer et al. 2012) within the statistical software tool R (R Core Development Team 2012). SVM was chosen because our prior research within this landscape suggested that SVM typically provides more accurate spectral classifications in comparison to RF and boosted classification and regression trees (CART) (Maxwell et al. 2014a; Maxwell et al. 2014b; Maxwell et al. 2015). A total of 9,409 objects were used to train the algorithm. These objects were selected based on manual interpretation of the 2011 NAIP orthophotography, prior 2007 NAIP orthophotography, mine permit data made available by the WVDEP, and the digital terrain data. For the classification, a RBF kernel was used and the user-defined parameters C and kernel-specific gamma (γ) were optimized using 10-fold cross validation in which the training data were partitioned into 10 unique training sets, using a random assignment, and the classifier was trained 10 times using 90% of the data and withholding the other 10% for validation.

The primary aim of the first stage of the classification was to map grasslands. However, as an intermediate step, the objects were classified as woody vegetation, herbaceous vegetation,

barren areas, and water. These classes were chosen based on our prior experience of common land cover conditions in this landscape (Maxwell et al. 2014a; Maxwell et al. 2014b; Maxwell et al. 2015), which suggests these classes are separable given only spectral data. After classification, contiguous areas of land cover smaller than the 1 ha MMU were removed using a sieving operation and replaced with the dominant surrounding class. A 1 ha MMU was selected because reclamation practices in this area commonly result in large patches, typically over 1 ha, of similar land cover. Groups of adjacent objects that were classified as grassland cover were then combined into single objects. The resulting objects were further classified in the next classification stage.

4.4. Differentiation of grasslands

In the second stage of the classification, image objects that were previously labeled as herbaceous vegetation in the initial classification were divided into mine-reclaimed grasslands and non-mining grasslands (Table 1) using pre- and post-mining terrain characteristics of each object. Although SVM was used for the first stage of the classification discussed above (i.e. the land cover classification using spectral data), RF was used for the second stage of the analysis, separating the grassland classes, as RF was assumed to be more suitable for modeling the complex interactions between the highly correlated terrain variables (Burkholder et al. 2011). In addition, RF was chosen because it offers measures of variable importance, as well as an estimate of error from the oob samples.

Predictor variables for the RF classification were derived from the pre- and post-mining DEM data. These variables are summarized in Table 2. Topographic slope (in degrees) was calculated using the Spatial Analyst Extension of ArcMap 10.2 (Burrough and McDonell 1998; ESRI 2012), whereas the other terrain attributes were calculated using the ArcGIS

Geomorphometry & Gradient Metrics Toolbox (Evans et al. 2014). The metrics calculated included CTMI (Moore et al. 1993; Gessler et al. 1995), slope position (Berry 2002), roughness (Blaszczynski 1997; Riley, DeGloria, and Elliot 1999), and dissection (Evans 1972). Slope position, roughness, and dissection rely on focal statistics calculated using a moving window; thus, the result is dependent on the window size used. For this study, we used window sizes of 11 pixels × 11 pixels, 21 pixels × 21 pixels, and 31 pixels × 31 pixels (i.e. 99 m × 99 m, 189 m × 189 m, and 279 m × 279 m) and averaged the results. These window sizes were assumed to approximate the hillslope scale, the scale of interest, and were selected by estimating the range of typical valley to ridge distances in this landscape. An elevation change grid was also produced by subtracting the pre-mining DEM from the post-mining DEM. Positive values indicate increases in elevation (e.g. fills) whereas negative values indicate decreased elevation (e.g. excavation).

Within each image object, summary mean, minimum, maximum, and standard deviation were calculated for pre- and post-mining elevation and slope, and also elevation change. For all other variables, only the mean was calculated (Table 2). Classifications were produced using the DEM-derived input variable combinations described in Table 3. As a baseline for the comparisons, a classification was also performed using only the spectral data derived from the NAIP imagery as band means and standard deviations (8 predictor variables) for each grassland object.

A total of 200 randomly chosen objects were used to train the model, 100 from each of the grassland classes. The RF algorithm from the randomForest package (Liaw and Wiener 2002) within the statistical software tool R (R Core Development Team 2012) was used. A total of 500 trees were used in the ensemble, as this was found to be adequate to produce a stable

classification result. The number of variables randomly sampled as candidates at each node (m) was optimized for each input variable combination using 10-fold cross validation.

4.5. Classification assessment

A variety of methods were used to assess the classifications including the following: traditional accuracy assessment using randomized validation data and error matrices, oob estimates of generalization error and variable importance provided by RF, and an assessment of confusion between mine reclaimed grasslands and non-mining grasslands using a randomized sample within areas classified as herbaceous vegetation. These methods will be discussed in more detail below.

In order to estimate overall map accuracy, an accuracy assessment was performed using 3,000 randomly selected point locations across the entire study area of the three watersheds. The classification of three classes was assessed: mine-reclaimed grasslands, non-mining grasslands, and non-grass cover. The random points were validated using visual interpretation of a variety of layers including 2011 NAIP orthophotography, 2007 NAIP orthophotography, pre-mining slope data, post-mining slope data, a pre-mining hillshade image, a post-mining hillshade image, the elevation change raster grid, and WVDEP surface mine permit data. Five of the 3,000 points were removed from the analysis as the correct class was uncertain due to change in land cover between the dates of the NAIP orthophotography and post-mining terrain data collection (i.e. the LiDAR data).

One strength of the RF algorithm is its ability to estimate classification error using the withheld, or oob, data (Breiman 2001; Rodríguez-Galiano et al. 2012a; Rodríguez-Galiano et al. 2012b). Lawrence, Wood, and Sheley (2006) and Rodríguez-Galiano et al. (2012b) suggest that this accuracy assessment is reliable and unbiased when randomized validation data are used, as

in this study. As a result, this estimate of classification accuracy was used to assess the accuracy of separating mine reclaimed grasslands from other grasslands, the second stage of the classification. The algorithm also generates a measure of variable importance during the training process by excluding each variable sequentially and recording the resulting oob error (Breiman 2001; Rodríguez-Galiano et al. 2012a; Rodríguez-Galiano et al. 2012b). This ancillary output of RF was used to assess the contribution of specific terrain measures calculated for the grassland objects (e.g. pre-mining mean elevation, post-mining mean elevation, mean elevation change, pre-mining mean slope position, post-mining mean terrain roughness, etc.).

In order to evaluate the statistical significance of any differences in the classifications the results were compared on a pairwise basis using McNemar's test (Dietterich 1998; Foody 2004). McNemar's test is a test of statistical difference that generates a *z*-score under the null hypothesis that the classifications are not different. A *z*-score larger than 1.645 indicates a 95% confidence of statistical significance for the one-directional test of whether one classification is more accurate than the other (Bradley 1968; Dietterich 1998; Foody 2004; Agresti 2007). This statistical test was used to assess the statistical difference between the classifications. It was also used to assess the separability or differentiation of mine reclaimed grasslands and other grasslands (i.e. the second stage of the classification) using a second set of 1,000 random validation points from within areas mapped as grasslands. Of the 1,000 original sample points, 33 were removed from the analysis as they could not be interpreted due to landscape change between the dates of the 2011 NAIP orthophotography and post-mining terrain data collection or because they were interpreted as not being grasslands (i.e. were incorrectly mapped as grasslands).

5. Results and discussion

5.1. Classification accuracy

Using different terrain input variable combinations, the overall accuracy of the classifications ranged from 97.4% to 97.9% (Table 4). Overall, the most accurate classifications generally used a combination of pre- and post-mining terrain variables or variables derived from the pre-mining surface only. Figure 3 shows the grassland classification for the entire study area produced using all pre-and post-mining terrain variables plus elevation change variables in the second stage of the classification. Figure 4 shows an example of the same result in greater detail, but for a smaller area and overlaid on the NAIP imagery. Table 5 summarizes the confusion matrix for this classification.

For the various combinations of terrain data, user's accuracy for mine-reclaimed grasslands ranged from 77% to 89% and producer's accuracy ranged from 77% to 83% (Table 4). For non-mining grasslands user's accuracy ranged from 76% to 85% and producer's accuracy ranged from 57% to 72%. The lower producer's accuracy for non-mining grasslands is due to confusion with both mine-reclaimed grasslands and non-grassland cover. Non-grassland cover was generally differentiated from grassland cover with user's and producer's accuracies above 98%.

Overall, these data suggest that grasslands can be accurately differentiated from other land cover types using GEOBIA, SVM, and NAIP orthophotography. In addition, the results indicate that terrain variables are useful for differentiating non-mining and mine-reclaimed grasslands, as using only spectral data in the second stage of the classification yielded the lowest overall accuracy (97.2%), and, most importantly, the lowest user's and producer's accuracies for both grassland classes (57% to 78%). This result is not particularly surprising, since we expected mine-reclaimed grasslands to be very similar spectrally to non-mining grasslands.

Because grasslands are only a small part of the overall landscape, the difference in overall accuracy between the classifications with different topographical variables varied by only 0.4%. However, the user's and producer's accuracies for the two grassland categories varied widely, and consequently the classifications were statistically different for many of the combinations, as shown by McNemar's test (Table 6). This confirms that the choice of input terrain predictor variables affects the accuracy of the classification. Further, all classifications that used terrain variables, with the exception of the combination utilising only post-mining terrain summary statistics for elevation and slope, were statistically more accurate than the classification using spectral data.

5.2. Importance of pre- and post-mining terrain data for differentiating mine-reclaimed and non-mining grasslands

The discussion so far has focused on statistics generated from the entire classification map. In order to explore the second stage of the classification more closely, we now focus exclusively on the differentiation of the mine-reclaimed grasslands and the non-mining grasslands.

Figure 5 shows oob error rates for the separation of non-mining and mine-reclaimed grasslands using different input variable combinations. The differentiation error rates using various combinations of terrain variables, as estimated using the oob data, range from 4.5% (using all pre- and post-mining terrain variables but not elevation change variables) to 16.0% (using only post-mining descriptive statistics for elevation and slope). The error rate using only spectral data in the second stage of the classification was 19.0%, the highest error rate obtained.

Using the random samples within the grassland classes, Table 7 shows the McNemar's test results for assessing the differentiation of mine-reclaimed and non-mining grasslands.

Statistical significance was observed between the majority of the input variable combinations, with the spectral data and the post-mining elevation and slope classifications being significantly different (i.e. having a lower accuracy) than all other classifications. The McNemar's test also confirms that a classification using pre- and post-mining terrain data (i.e. not including statistics derived from elevation change) provided a statistically more accurate differentiation of the two classes than a classification using only post-mining data (z-score = 2.635) but not in comparison to a classification using pre-mining data (z-score = 0.333). Pre-mining data only also produced a statistically more accurate differentiation than post-mining data only (z-score = 2.457).

In summary, these data suggest that mine-reclaimed grasslands have a unique topographic signature compared to other grasslands in this terrain, and can thus be separated from other grasslands using terrain characteristics extracted from DEM data. This is especially true when both pre- and post-mining characteristics are used or when just pre-mining characteristics are used. We attribute the usefulness of pre-mining data to the nature of the terrain alteration resulting from MTR/VF mining. A pre-mining topography characterised by steep slopes and an upper slope position may be more predictive than post-mining terrain characteristics, in which the landscape has been flattened and therefore became more similar topographically to non-mining grasslands.

5.3. Importance of terrain attributes for differentiating mine-reclaimed and non-mining grasslands

The McNemar's test (Table 7) comparing the grassland differentiation using all pre- and post-mining predictor variables with and without including the elevation change data yielded a z-score of 1.604. This suggests that the incorporation of descriptive statistics derived from the elevation change surface did not statistically improve the classification accuracy. However, a

classification using just the elevation change data was not statistically different from a classification using all of the pre- and post-mining predictor variables excluding elevation change variables (*z*-score = 0.277). A combination of summary statistics for elevation change and pre- and post-mining elevation and slope was statistically more accurate than a classification using just pre- and post-mining elevation and slope variables (*z*-score = 2.333). As shown in Figure 6, measures derived from the elevation change surface were of particular importance in the model as estimated by the oob mean decrease in accuracy measure. These data suggest that there is merit in including elevation change variables, especially when the number of terrain variables used to characterize the pre- and post-mining terrain are limited.

The incorporation of pre- and post-mining CTMI, slope position, roughness, and dissection in the classification was also assessed. These topographic variables statistically improved the differentiation of the two grassland classes in comparison to only using measures derived from elevation and slope when only post-mining terrain data were used (*z*-score = 6.359) and when only pre-mining data were used (*z*-score = 4.619); however, no statistical difference was observed when using a combination of pre- and post-mining data (*z*-score = 1.270). Figure 7 shows variable importance for the post-mining model as estimated from the oob mean decrease in accuracy. These data suggest that the additional terrain variables, beyond elevation and slope, contribute to the model, especially dissection. Figure 8 shows variable importance for the pre-mining model. These data suggest that the added variables, beyond elevation and slope, contribute to the model, especially dissection and roughness. However, the most important variable appears to be the pre-mining mean slope. The reason for this is likely because pre-mining slopes of MTR/VF sites are often steep, thus, pre-mining slopes differentiate mine-reclaimed grasslands from non-mining grasslands, which are often found on flatter surfaces (e.g.

valley bottoms) in this landscape. These data tend to suggest that derived topographic variables are of value, especially when only pre- or post-mining terrain data are available.

5.4. Practical considerations

Pre- and post-mining terrain data were found to be of value for differentiating spectrally similar non-mining and mine-reclaimed grasslands in this study area in West Virginia. However, there are some practical limitations. First, the pre-mining terrain data were of greater importance for differentiating the cover types than post-mining data. The availability of older DEM data for characterizing the pre-mining terrain is generally limited, and in this study it was necessary to produce a DEM from the available DLG data as a historical DEM was not readily available. Further, USGS DLG contour data were collected over a wide range of dates and were derived from photogrammetric methods, making comparison to the recent LiDAR-derived data complex (DeWitt, Warner, and Conley 2015). In addition, the DLG data may not pre-date all mining. Second, post-mining terrain data may not be temporally coincident with the available imagery. For example, in this study, the NAIP orthophotography was collected over a period of nearly three months, and the LiDAR data were collected over a period of nearly a year. Planning temporally coincident collections of high resolution imagery and LiDAR may be difficult, especially over large spatial extents. For studies with limited budgets that must exploit data originally collected for other purposes, as in this study, the challenges of finding data of similar dates are even greater.

Many of the terrain attributes calculated rely on focal statistics calculated using a moving window. Selecting the appropriate window size can be difficult as the optimal window size may be case-specific and guidance from the literature on the appropriate scale is limited. This presents a challenge when working with DEM-derived terrain attributes for LULC classification.

Despite these limitations, our results suggest that multi-temporal terrain data summarized for image objects offers a means to differentiate spectrally similar LULC classes that have a characteristic topographic signature. This is especially true when machine learning algorithms are used to classify such data. Such methods may be appropriate for augmenting available data sets to support a specific modeling or analysis task, such as the National Land Cover Dataset (NLCD).

6. Conclusions

This research investigated the use of multi-temporal terrain data for differentiating these topographically distinctive features. Surface mining produces extensive landscape alterations that persists as a legacy LULC alteration. Mine reclaimed land cover has been shown to have important impacts on hydrology, terrestrial habitats, and aquatic ecosystems. Thus, it is of importance to differentiate such grasslands from other grasslands on the landscape.

The classification approach employed, which makes use of GEOBIA, machine learning algorithms, high resolution aerial imagery, and multi-temporal terrain characteristics derived from DEMs, provided an accurate means to differentiate grassland cover from other land cover.

Mine-reclaimed grasslands were mapped with user's and producer's accuracies between 77% and 89% using multi-temporal terrain data. Classifications using either a combination of pre- and post-mining terrain variables or pre-mining terrain variables only generally outperformed classifications using only post-mining terrain data. Elevation change data were of value, and terrain characteristics as CTMI, slope position, roughness, and dissection generally improved the classification.

GEOBIA was a valuable tool for combining data collected using different sensors and gridded at variable cell sizes (i.e. the image and digital terrain data). In addition, GEOBIA

provided a mechanism to characterise the terrain data using summary variables (e.g. mean, maximum, minimum, standard deviation, etc.) at the object scale. Differentiating landscape position from attributes derived from DEMs is not straightforward because the scale of the landscape is complex, with multiple potential topographic scales present. Also, a single site might include more than one topographic class. With GEOBIA, by integrating over an object, these problems can potentially be overcome.

The machine learning algorithms were particularly useful in incorporating the ancillary data derived from the DEMs, since these most likely would not have met the basic assumptions of multivariate normality required for parametric classifiers. In addition, the RF classifier was particularly useful due to its ability to provide estimates of accuracy and also variable importance.

This study highlights the importance of maintaining legacy elevation products (e.g. DLG) with descriptive metadata regarding year of acquisition or creation, since the pre-mining terrain data were shown to be of great value in this study. We know of no formal effort to archive historical elevation data sets analogous to the extensive image archives that are maintained by the USGS, the National Aeronautic and Space Administration (NASA) and other government agencies. We recommend that developing such archives should be a priority.

Acknowledgments

LiDAR data were provided by the West Virginia Department of Environmental Protection (WVDEP) and the Natural Resource Analysis Center (NRAC) at West Virginia University. We would specifically like to acknowledge Adam Riley and Paul Kinder for their assistance in obtaining and processing the LiDAR data.

Funding

Funding support for this study was provided by West Virginia View and the Appalachian College Association (ACA). The project described in this publication was also supported in part by Grant Number G14AP00002 from the Department of the Interior, United States Geological Survey to AmericaView. Its contents are solely the responsibility of the authors; the views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Government. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Government.

References

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*, 2nd ed., 400. Hoboken, NJ: Wiley-Interscience.

Anderson, A. T., and J. Schubert. 1976. "ERTS-1 Data Applied to Strip Mining." *Photogrammetric Engineering & Remote Sensing* 42: 211-219.

Anderson, A. T., D. Schultz, N. Buchman, and H. M. Nock. 1997. "Landsat Imagery for Surface-Mine Inventory." *Photogrammetric Engineering & Remote Sensing* 43 (8): 1027-1036.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. "A Land-Use and Land Cover Classification System for Use with Remote Sensor Data." *Geologic Survey Professional Paper No. 964*. Washington, DC: US Government Printing Office.

Baatz, M., and A. Schäpe. 2000. "Multiresolution Segmentation – An Optmimization Approach for High Quality Multi-Scale Image Segmentation." In *Angewandte Geographische Informationsverarveitung XII*, edited by J. Strobl, T. Blaschke, and G. Griesebener, 12-23. Berlin: Herbert Wichmann Verlag.

Baker, B. A., T. A. Warner, J. F. Conley, and B. E. McNeil. 2013. "Does Spatial Resolution Matter? A Multi-Scale Comparison of Object-Based and Pixel-Based Methods for Detecting Change Associated with Gas Well Drilling Operations." *International Journal of Remote Sensing* 34 (5): 1633-1651.

Bernhardt, E. S., B. D. Lutz, R. S. King, J. P. Fay, C. E. Carter, A. M. Helton, D. Campagna, and J. Amos. 2012. "How Many Mountains Can We Mine? Assessing the Regional Degradation of

Ventral Appalachian Rivers by Surface Coal Mining." *Environmental Science & Technology* 46 (15): 8115-8122.

Bernhardt, E. S., and M. A. Palmer. 2011. "The Environmental Costs of Mountaintop Mining Valley Fill Operations for Aquatic Ecosystems of the Central Appalachians." *Annals of the New York Academy of Sciences* 1223 (1): 39-57.

Berry, J. K. 2002. "Beyond Mapping Use Surface Area for Realistic Calculations." *Geo World* 15 (9): 20-21.

Blaschke, T. 2003. "Object-Based Contextual Image Classification Built on Image Segmentation." 2003 IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, Greenbelt, MD, Ocotber 27-28, 13-119.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1): 2-16.

Blaschke, T., G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Q. Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede. 2014. "Geographic object-based image analysis – Towards a new paradigm." *ISPRS Journal of Photogrammetry and Remote Sensing* 87 (100): 180-191.

Blaschke, T, and J. Strobl. 2001. "What's Wrong with Pixels? Some Recent Developments Interfacing Remote Sensing and GIS." *GIS – Zeitschrift Für Geoinformationssysteme* 14 (6): 12-17.

Blaszczynski, J. S. 1997. "Landform Characterization with Geographic Information Systems." *Photogrammetric Engineering & Remote Sensing* 63 (2): 183-191.

Bradley, J. V. 1968. *Distribution-Free Statistical Tests*, 388. Englewood Cliffs, NJ: Prentice Hall.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24 (2): 123-140.

Breiman, L. 2001. "Random Forests." *Machine Learning* 54 (1): 5-32.

Burges, C. J. C. 1998. "A Tutorial on Support Vector Machines for Pattern Recognition." *Data Mining and Knowledge Discovery* 2 (2): 121-167.

Burkholder, A., T. A. Warner, M. Culp and R. E. Landenberger. 2011. "Seasonal trends in separability of leaf reflectance spectra for *Ailanthus altissima* and four other tree species." *Photogrammetric Engineering & Remote Sensing* 77 (8): 793-804.

Burrough, P. A., and R. A. McDonell. 1998. *Principles of Geographical Information Systems*, 2nd ed., 356. New York: Oxford University Press.

Chubey, M. S., S. E. Franklin, and M. A. Wulder. 2006. "Object-Based Analysis of IKONOS-2 Imagery for Extraction of Forest Inventory Parameters, *Photogrammetric Engineering & Remote Sensing* 72 (4): 383-394.

Cihlar, J., and L. J. M. Jensen. 2001. "From Land Cover to Land Use: A Methodology for Efficient Land Use Mapping Over Large Areas." *Professional Geographer* 53 (2): 275-289.

Cutler, D. R., T. C. Edwards, Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random Forests for Classification in Ecology, *Ecology* 88 (11): 2783-2792.

Davies, K. W., S. L. Petersen, D. D. Johnson, D. B. Davis, M. D. Madsen, D. L. Zvirzdin, and J. D. Bates. 2010. "Estimating Juniper Cover from National Agriculture Imagery Program (NAIP)

Imagery and Evaluating Relationships between Potential Cover and Environmental Variables." *Rangeland Ecology & Management* 63 (6): 630-637.

DeWitt, J. D., T. A. Warner, and J. F. Conley. 2015. "Comparison of DEMs Derived from USGS DLG, SRTM, a Statewide Photogrammetric Program, ASTER GDEM and Lidar." *GIScience & Remote Sensing* 52 (2): 179-197.

Dietterich, T. G. 1998. "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms." *Neural Computation* 10 (7): 1895-1923.

Dingle Robertson, L., and D. J. King. 2011. "Comparison of Pixel- and Object-Based Classification in Land Cover Change Mapping." *International Journal of Remote Sensing* 32 (6): 1505-1529.

Drăgut, L., and T. Blaschke. 2006. "Automated Classification of Landform Elements Using Object-Based Image Analysis." *Geomorphology* 81 (3): 330-344.

Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012a. "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." *Remote Sensing of Environment* 118: 259-272.

Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012b. "Multi-Scale Object-Based Image Analysis and Feature Selection of Multi-Sensor Earth Observation Imagery using Random Forests." *International Journal of Remote Sensing* 33 (14): 4502-4526.

Ehlke, T. A., G. S. Runner, and S. C. Downs. 1982. "Hydrology of Area 9, Eastern Coal Province, West Virginia." *United States Geological Survey Water-Resources Investigation*, 63. Open File Report 81-803

ESRI. 2012. *ArcGIS Desktop: Release 10.1*. Redlands, CA: Environmental Systems Research Institute.

ERDAS Imagine. 2013. *ERDAS Field Guide*, 792. Huntsville, Alabama: Intergraph Corporation.

Evans, I. S. 1972. "General geomorphometry, derivatives of altitude, and descriptive statistics." In *Spatial Analysis in Geomorphology*, edited by R. J. Chorley, 17-90. New York: Harper & Row.

Evans, J. S., and J. M. Kiesecker. 2014. "Shale Gas, Wind and Water: Assessing the Potential Cumulative Impact of Energy Development on Ecosystem Services within the Marcellus Play, *PLOS ONE* 9 (2): e89210.

Evans, J. S., J. Oakleaf, S. A. Cushman, and D. Theobald. 2014. "An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling." Version 2.0-0. Accessed March, 12, 2015. http://evansmurphy.wix.com/evansspatial

Ferrari, J. R., T. R. Lookingbill, B. McCormick, P. A. Townsend, and K. N. Eshleman. 2009. "Surface Mining and Reclamation Effects on Flood Response of Watersheds in the Central Appalachian Plateau Region." *Water Resources Research*, 45 (W04407): 1-11.

Folke, C. L. Pritchard, Jr., F. Berkes, J. Colding, U. Svedin. 1998. "The Problem of Fit between Ecosystems and Institutions." Paper presented at the International Human Dimensions Programme on Global Environmental Change, Bonn, Germany.

Foody, G. M. 2004. "Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy." *Photogrammetric Engineering & Remote Sensing* 70 (5): 627-634.

Fritz, K. M., S. Fulton, B. R. Johnson, C. D. Barton, J. D. Jack, D. A. Word, and R. A. Burke. 2010. "Structural and Functional Characteristics of Natural and Constructed Channels Draining a Reclaimed Mountaintop Removal and Valley Fill Coal Mine." *Journal of the North American Benthological Society* 29 (2): 637-689.

Gessler, P. E., I. D. Moore, N. J. McKenzie, and P. J. Ryan. 1995. "Soil-Landscape Modelling and Spatial Prediction of Soil Attributes." *International Journal of GIS* 9 (4): 421-432.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. "Random Forests for Land Cover Classification." *Pattern Recognition Letters* 27 (4): 294-300.

Guan, H., J. Li, M. Chapman, F. Deng, Z. Ji., and X. Yang. 2013. "Integration of Orthoimagery and Lidar for Object-Based Urban Thematic Mapping using Random Forests." *International Journal of Remote Sensing* 34 (14): 5166-5186.

Hansen, M. C., and B. Reed. 2000. "A Comparison of the IGBP DISCover and University of Maryland 1 km Global Land Cover Products." *International Journal of Remote Sensing* 21 (6-7): 1365-1373.

Hartman, K. J., M. D. Kaller, J. W. Howell, and J. A. Sweka. 2005. "How Much Do Valley Fills Influence Headwater Streams?" *Hyrdobiologia,* 532: 91-102.

Hooke, R. L. 1994. "On the Efficacy of Humans as Geomorphic Agents." *GSA Today* 4 (9): 217, 224-225.

Hooke, R. L. 1999. "Spatial Distribution of Human Geomorphic Activity in the United States: Comparison to Rivers." *Earth Surface Processes and Landforms* 24 (8): 687-692.

Hyyppä, J., W. Wagner, M. Hollaus, and H. Hyyppä. 2009. "Airborne Laser Scanning." In *The SAGE Handbook of Remote Sensing*, edited by T. A. Warner, M. D. Nellis, and G. M. Foody, 199-212. London: Sage Publications Ltd.

Irons, J. R., and R. L. Kennard. 1986. "The Utility of Thematic Mapper Sensor Characteristics for Surface Mine Monitoring." *Photogrammetric Engineering & Remote Sensing* 52: 389-396.

Joachims, T. 1998. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features." *Proceedings of European Conference on Machine Learning*, Chemnitz, April 21-23, 137-142.

Kazar, S. A., and T. A. Warner. 2013. "Assessment of Carbon Storage and Biomass on Minelands Reclaimed to Grassland Environments using Landsat Spectral Indices." *Journal of Applied Remote Sensing* 7 (1): 073583. doi: 10.1117/1.JRS.7.073583.

Ke, Y., L. J. Quackenbush, and J. Im. 2010. "Synergistic Use of Quickbird Multispectral Imagery and LiDAR Data for Object-Based Forest Species Classification." *Remote Sensing of Environment* 114: 1141-1154.

Kim, M., M. Madden, and T. A. Warner. 2009. "Forest Type Mapping using Object-Specific Texture Measures from Multispectral Ikonos Imagery: Segmentation Quality and Image Classification Issues." *Photogrammetric Engineering & Remote Sensing* 75 (7): 819-829.

Kim, M., T. A. Warner, M. Madden, and D. S. Atkinson. 2011. "Multi-Scale GEOBIA with Very High Spatial Resolution Digital Aerial Imagery: Scale, Texture, and Image Objects." *International Journal of Remote Sensing* 32 (10): 2825-2850.

Knight, J. F., B. P. Tolcser, J. M. Corcoran, and L. P. Rampi. 2013. "The Effects of Data Selection and Thematic Detail on the Accuracy of High Spatial Resolution Wetland Classification" *Photogrammetric Engineering & Remote Sensing* 79 (7): 613-623.

Laliberte, A. S., E. L. Fredrickson, and A. Rango. 2007. "Combining Decision Trees with Hierarchical Object-Oriented Image Analysis for Mapping Arid Rangelands." *Photogrammetric Engineering & Remote Sensing* 73 (2): 197-207.

Lawrence, R. L., S. D. Wood, and R. L. Sheley. 2006. "Mapping Invasive Plants using Hyperspectral Imagery and Breiman Cutler Classifications (RandomForest)." *Remote Sensing of Environment* 100: 356-362.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2 (3): 18-22.

Liu, D., and F. Xia. 2010. "Assessing Object-Based Classification: Advantages and Limitations." *Remote Sensing Letters* 1 (4): 187-194.

Mathieu, R., J. Aryal, and A. K. Chong. 2007. "Object-Based Classification of Ikonos Imagery for Mapping Large-Scale Vegetation Communities in Urban Areas." *Sensors* 7 (11): 2860-2880.

Maxwell, A. E., M. P. Strager, C. Yuill, J. T. Petty, E. Merriam, and C. Mazzarella. 2011. "Disturbance Mapping and Landscape Modeling of Mountaintop Mining using ArcGIS." Paper

presented at the International ESRI User Conference, 11-15 July, San Diego, California, July 11-15.

Maxwell, A. E., and M. P. Strager 2013. "Assessing Landform Alteration Induced by Mountaintop Mining." *Natural Science* 5 (2A): 229-237.

Maxwell, A. E., M. P. Strager, T. A. Warner, N. P. Zégre, and C. B. Yuill. 2014a. "Comparison of NAIP Orthophotography and RapidEye Satellite Imagery for Mapping of Mining and Mine Reclamation." *GIScience & Remote Sensing* 51 (3): 301-320.

Maxwell, A. E., T. A. Warner, M. P. Strager, and M. Pal. 2014b. "Combining RapidEye Satellite Imagery and LiDAR for Mapping of Mining and Mine Reclamation." *Photogrammetric Engineering and Remote Sensing* 80 (2): 179-189.

Maxwell, A. E., T. A. Warner, M. P. Strager, J. F. Conley, and A. L. Sharp. 2015. "Assessing Machine Learning Algorithms and Image- and LiDAR-Derived Variables for GEOBIA Classification of Mining and Mine Reclamation." *International Journal of Remote Sensing* 36 (4): 954-978.

McCormick, B. C., K. N. Eshleman, J. L. Griffith, and P. A. Townsend. 2009. "Detection of Flooding Responses at the River Basin Scale Enhanced by Land Use Change." *Water Resources Research* 45 (W08401): 1-15.

Meinel, G., and M. Neubert. 2004. "A Comparison of Segmentation Programs for High Resolution Remote Sensing Data." *International Archives of the ISPRS* 35: 1097-1105.

Meneguzzo, D. M., G. C. Liknes, and M. D. Nelson. 2013. "Mapping Trees Outside Forests Using High-Resolution Aerial Imagery: A Comparison of Pixel- and Object-Based Classification Approaches." *Environmental Monitoring and Assessment* 185: 6261-6275.

Merriam, E. R., J. T. Petty, G .T. Merovich, J. B. Fulton, and M. P. Strager, 2011. "Additive Effects of Mining and Residential Development on Stream Conditions in a Central Appalachian Watershed." *Journal of the North American Benthological Society* 30 (2): 399-418.

Merriam, E. R., J. T. Petty, M. P. Strager, A. E. Maxwell, and P. F. Ziemkiewicz. 2013. "Scenario Analysis Predicts Context-Dependent Stream Response to Landuse Change in a Heavily Mined Central Appalachian Watershed." *Freshwater Science* 32 (4): 1246-1259.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2012. "e1071: Misc Functions of the Department of Statistics (e1071)." R Package Version 1.6-1. Accessed March 12, 2015. http://CRAN.R-project.org/package=e1071

Milici, R. C. 2000. "Depletion of Appalachian Coal Reserves – How Soon?" *International Journal of Coal Geology* 44 (3-4): 251-266.

Miller, A. J., and N. P. Zégre. 2014. "Mountaintop Removal Mining and Catchment Hydrology." *Water* 6 (3): 472-499.

Moore, I. D., P. E. Gessler, G. Z. Nielsen, and G. A. Petersen. 1993. "Terrain Attributes: Estimation Methods and Scale Effects. In *Modeling Change in Environmental Systems*, edited by A .J. Jakeman, M. B. Beck, and M. McAleer, 189-214. London: Wiley.

Myint, S. W., P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng. 2011. "Per-Pixel vs. Object-Based Classification of Urban Land Cover Extraction using High Spatial Resolution Imagery." *Remote Sensing of Environment* 115: 1145-1161.

Negley, T. L., and K. N. Eshleman. 2006. "Comparison of Streamflow Responses of Surface-Mined and Forest Watersheds in the Appalachian Mountain, USA." *Hydrological Processes* 20 (16): 3467-3483.

Pal, M. 2005. "Random Forest Classifiers for Remote Sensing Classification." *International Journal of Remote Sensing* 26 (1): 217-222.

Pal, M, and P. M. Mather. 2005. "Support Vector Machines for Classification in Remote Sensing." *International Journal of Remote Sensing* 5 (10): 1007-1011.

Palmer, M. A., E. S. Bernhardt, W. H. Schlesinger, K. N. Eshleman, E. Fourfoula-Georgiou, M. S. Hendryx, A. D. Lemly, G. E. Likens, O. L. Loucks, M. E. Power, P. S. White, and P. R. Wilcock. 2010. "Mountaintop Mining Consequences." *Science* 327: 148-149.

Parks, N. F., G. W. Petersen, and G. M. Baumer. 1987. "High Resolution Remote Sensing of Spatially and Spectrally Complex Coal Surface Mines of Central Pennsylvania: A Comparison Between SPOT, MSS, and Landsat-5 Thematic Mapper." *Photogrammetric Engineering & Remote Sensing* 53(4): 415-420.

Pino-Mejías, R., M. D. Cubiles-de-la-Vega, M. Anaya-Romero, A. Pascual-Acosta, A. Jordán López, and N. Bellinfante-Crocci. 2010. "Predicting the Potential Habitat of Oaks with Data Mining Models and the R System." *Environmental Modelling & Software* 25 (7): 826-836.

Pond, G. J. 2010. "Patterns of Ephemeroptera Taxa Loss in Appalachian Headwater Streams (Kentucky, WV)." *Hydrobiologia* 641: 185-201.

Pond, G. J., M. E. Passmore, F. A. Boruk, L. Reynolds, and C. J. Rose. 2008. "Downstream Effects of Mountaintop Coal Mining: Comparing Biological Conditions Using Family- and Genus-Level Macroinvertebrate Bioassessment Tools." *Journal of the North American Benthological Society* 27 (3): 717-737.

Prakash, A., and R. P. Gupta, 1998. "Land-Use Mapping and Change Detection in a Coal Mining Area: A Case Study in the Jharia Coalfields, India." *International Journal of Remote Sensing* 19 (3): 391-410.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. "Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction." *Ecosystems* 9: 181-199.

Pu, R., S. Landry, and Q. Yu. 2011. "Object-Based Urban Detailed Land Cover Classification With High Spatial Resolution IKONOS Imagery." *International Journal of Remote Sensing* 32 (12): 3285-3308.

Rathore, C. S., and R. Wright. 1993. "Monitoring Environmental Impacts of Surface Coal-Mining." *International Journal of Remote Sensing* 14 (6): 1021-1042.

R Core Development Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Accessed March 12, 2015. URL: http://www.R-project.org.

Riley, S. J., S. D. DeGloria, and R. Elliot. 1999. "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity." *Intermountain Journal of Science* 5: 1-4.

Rodríguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez. 2012a. "An Assessment of the Effectiveness of a Random Forest Classifier for Land-Cover Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93-104.

Rodríguez-Galiano, V. F., M. Chica-Olma, F. Abarca-Hernández, P. M. Atkinson, and C. Jeganathan. 2012b. "Random Forest Classification of Mediterranean Land Cover using Multi-Seasonal Imagery and Multi-Seasonal Texture." *Remote Sensing of Environment* 121: 93-107.

Sen, S., C. E. Zipper, R. H. Wynne, and P. F., Donovan. 2012. "Identifying Revegetated Mines as Disturbance/Recovery Trajectories Using an Interannual Landsat Chronosequence." *Photogrammetric Engineering & Remote Sensing* 78 (3): 223-235.

Simmons, J. A., W. S. Currie, K. N. Eshleman, K. Kuers, S. Monteleone, T. L. Negley, B. R. Pohlad, and C. L. Thomas. 2008. "Forest to Reclaimed Mine Land Use Change Leads to Altered Ecosystem Structure and Function." *Ecological Applications* 18 (1): 104-118.

Smith, A. 2010. "Image Segmentation Scale Parameter Optimization and Land Cover Classification Using the Random Forest Algorithm." *Journal of Spatial Science*, 55 (1): 69-79.

Steele, B. M. 2000. "Combining Multiple Classifiers: An Application Using Spatial and Remotely Sensed Information for Land Cover Mapping." *Remote Sensing of Environment* 74 (3): 545-556.

Townsend, P. A., D. P. Helmers, C. C. Kingdon, B. E. McNeil, K. M. de Beurs, and K. N. Eshleman. 2009. "Changes in the Extent of Surface Mining and Reclamation in the Central Appalachians Detected Using a 1976-2006 Landsat Time Series." *Remote Sensing of Environment* 113: 62-72.

Treitz, P., and P. Howarth. 2000. "Integrating Spectral, Spatial, and Terrain Variables for Forest Ecosystem Classification." *Photogrammetric Engineering & Remote Sensing* 66 (3): 305-318.

Trimble. 2011. *eCognition Developer 8.64.1 User Guide*, Munich: Trimble.

Tso, B., and P. Mather. 2003. *Classification Methods for Remotely Sensed Data*, 352. New York: CRC Press.

Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*, 188. New York: Springer-Verlag.

Walker, B. 1998. "GCTE and LUCC – A Natural and Timely Partnership." *LUCC Newsletter* 3: 3-4.

Walter, V. 2004. "Object-Based Classification of Remote Sensing Data for Change Detection." *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-4): 225-238.

Warner, T. A., and F. Nerry. 2009. "Does a Single Broadband or Multispectral Thermal Data Add Information for Classification of Visible, Near- and Shortwave Infrared Imagery of Urban Areas?" *International Journal of Remote Sensing* 30 (9): 2155-2171.

Weakland, C. A., and P. B. Wood. 2005. "Cerulean Warbler (*Dendroica cerulean*) Microhabitat and Landscape-Level Habitat Characteristics in Southern West Virginia." *The Auk* 122 (2): 497-508.

Wickham, J. D., K. H. Riitters, T. G. Wade, M. Coan, and C. Homer. 2007. "The Effect of Appalachian Mountaintop Mining on Interior Forest." *Landscape Ecology* 22: 179-187.

Wickham, J., P. B. Wood, M. C. Nicholson, W. Jenkins, D. Druckenbrod, G. W. Suter, M. P. Strager, C. Mazzarella, W. Galloway, and J. Amos. 2013. "The Overlooked Terrestrial Impacts of Mountaintop Mining." *BioScience* 63 (5): 335-348.

Wood, P. B, S. B. Bosworth, and R. Dettmers. 2006. "Cerulean Warbler Abundance and Occurrence Relative to Large-Scale Edge and Habitat Characteristics." *The Condor* 108 (1): 154-165.

Wright, C., and A. Gallant. 2007. "Improved Wetland Remote Sensing in Yellowstone National Park using Classification Trees to Combine TM Imagery and Ancillary Environmental Data." *Remote Sensing of Environment* 107 (4): 582-605.

Yuill, C. 2003. "Landscape Use Assessment: Mountaintop Mining and the Mountaintop Mining Region of West Virginia." *Draft Programmatic Environmental Impact Statement on Mountaintop Mining/Valley Fills in Appalachia* III, F-12.

Zégre, N. P, A. J. Miller, A. Maxwell, and S. J. Lamont. 2014. "Multiscale Analysis of Hydrology in a Mountaintop Mine-Impacted Watershed." *Journal of the American Water Resources Association* 50 (5): 1257-1272.

Zégre, N. P., A. Maxwell, and S. Lamont. 2013. "Characterizing Streamflow Response of a Mountaintop-Mined Watershed to Changing Land Use." *Applied Geography* 39: 5-15.

Table 1: Grassland class definitions.

Land Cover Class	Description
Non-mining grasslands	Grasslands not resulting from mine reclamation, including pastureland, herbaceous dominated residential development, and other areas on the landscape dominated by herbaceous vegetation.
Mine-reclaimed grasslands	Grasslands resulting from mine reclamation, including reclaimed lands within mine sites and valley fills dominated by herbaceous vegetation.

Table 2: Descriptions of terrain characteristics.

Measure	Description	Reference	Object summary statistics
Elevation (Ele)	Elevation Z	NA	Mean, minimum, maximum, standard deviation
Slope (°) (Slp)	Slope (gradient or rate of maximum change in Z) $\text{atan}\sqrt{\frac{(\text{Rise})^2}{(\text{Run})^2}} \times 57.29578$	Burrough and McDonell 1998	Mean, minimum, maximum, standard deviation
Compound topographic moisture index (CTMI)	Measure of steady state wetness as estimated from terrain characteristics $\ln\left(\frac{(\text{Upstream contibuting area})}{\tan(\text{Slope})}\right)$	Gessler et al. 1995 Moore et al. 1993	Mean
Slope position	Scalable slope position $Z - Z_{\text{mean}}$	Berry 2002	Mean
Roughness	Roughness or terrain complexity index $\sqrt{Z_{\text{standard\_deviation}}}$	Riley et al. 1999 Blaszczynski 1997	Mean
Dissection	Dissection of landscape index $\frac{(Z - Z_{\text{minimum}})}{(Z_{\text{maximum}} - Z_{\text{minimum}})}$	Evans 1972	Mean
Elevation change	Pre-Mining Elevation – Post-Mining Elevation $Z_{\text{post-mining}} - Z_{\text{pre-mining}}$	NA	Mean, minimum, maximum, standard deviation

Table 3: Input variable combinations compared.

Classification	Variables used in the classification	Number of variables
1	All pre- and post-mining DEM-derived variables plus elevation change summary variables	28
2	All pre- and post-mining DEM-derived variables (but excluding elevation change)	24
3	Pre- and post-mining elevation and slope plus elevation change summary variables	20
4	Pre- and post-mining elevation and slope summary variables	16
5	All pre-mining DEM-derived summary variables	12
6	All post-mining DEM-derived summary variables	12
7	Pre-mining elevation and slope summary variables	8
8	Post-mining elevation and slope summary variables	8
9	Elevation change summary variables	4
10	Spectral data (band means and standard deviations)	8

Table 4: Summary statistics for classification of three classes (not grassland, non-mining grasslands, mine-reclaimed grasslands). The best value in each column is shaded gray.

			Mine-reclaimed grasslands		Non-mining grasslands		Not grassland	
Variables used	**Data set**	**OA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**	**UA (%)**	**PA (%)**
All Pre-/Post-mining + Ele Change	**1**	97.9	89	81	82	72	99	99
All Pre-/Post-mining	**2**	97.9	88	80	81	72	99	99
Pre-/Post-mining Ele and Slp + Ele Change	**3**	97.9	86	83	85	69	99	99
Pre-/Post-mining Ele and Slp	**4**	97.7	83	78	80	67	99	99
All Pre-mining	**5**	97.9	88	79	80	72	99	99
All Post-mining	**6**	97.6	82	77	78	65	99	99
Pre-mining Ele and Slp	**7**	97.8	87	79	79	70	99	99
Post-mining Ele and Slp	**8**	97.4	77	77	76	57	99	99
Ele Change	**9**	97.9	86	83	84	69	99	99
Spectral Data (no terrain data)	**10**	97.2	78	72	66	57	99	99

Note: OA = overall accuracy, UA = user's accuracy, PA = producer's accuracy.

Table 5: Error matrix for classification using all predictor variables (All pre- and post-mining variables plus elevation change in the second stage of the classification). Overall accuracy is 97.9% for separating the three classes.

		Reference data				
		Not grassland	Non-mining grasslands	Mine-reclaimed grasslands	**Total**	**User's accuracy (%)**
Classified data	Not grasslands	2780	18	19	2817	99
	Non-Mining grasslands	10	60	3	73	82
	Mine-reclaimed grasslands	7	5	93	105	89
	Total	2797	83	115		
	Producer's accuracy (%)	99	72	81		

Table 6: McNemar's test results for classification of three classes (not grassland, non-mining grasslands, mine-reclaimed grasslands).

	Data set	1	2	3	4	5	6	7	8	9
All Pre-/Post-mining + Ele Change	**1**									
All Pre-/Post-mining	**2**	0.447								
Pre-/Post-mining Ele and Slp + Ele Change	**3**	0.000	0.333							
Pre-/Post-mining Ele and Slp	**4**	2.111*	1.897*	2.111*						
All Pre-mining	**5**	0.817	1.000	0.632	1.508					
All Post-mining	**6**	2.673*	2.324*	2.673*	0.655	2.000*				
Pre-mining Ele and Slp	**7**	1.414	1.134	1.265	1.000	0.707	1.414			
Post-mining Ele and Slp	**8**	3.400*	3.138*	3.710*	2.041*	2.887*	1.698*	2.837*		
Ele Change	**9**	0.333	0.000	0.447	1.500	0.258	2.065*	0.832	3.138*	
Spectral Data	**10**	3.888*	3.667*	4.131*	2.744*	3.452*	1.982*	3.212*	0.949	4.017*

Note: A *z*-score larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is more accurate than the other.

Table 7: McNemar's test results for differentiation of non-mining and mine-reclaimed grasslands.

	Data set	1	2	3	4	5	6	7	8	9
All Pre-/Post-mining + Ele Change	**1**									
All Pre-/Post-mining	**2**	1.604								
Pre-/Post-mining Ele and Slp + Ele Change	**3**	0.200	1.238							
Pre-/Post-mining Ele and Slp	**4**	2.491*	1.270	2.333*						
All Pre-mining	**5**	1.692*	0.333	1.309	1.116					
All Post-mining	**6**	4.281*	2.635*	3.878*	1.497	2.457*				
Pre-mining Ele and Slp	**7**	5.128*	4.621*	4.587*	2.994*	4.619*	0.539			
Post-mining Ele and Slp	**8**	8.433*	7.209*	8.275*	6.683*	7.030*	6.359*	5.031*		
Ele Change	**9**	1.134	0.277	1.180	1.192	0.368	2.652*	3.133*	6.982*	
Spectral Data	**10**	7.898*	6.548*	7.929*	5.933*	6.438*	4.820*	4.353*	0.254	7.233*

Note: A *z*-score larger than 1.645 (*) indicates a 95% confidence interval of statistical significance for the one-directional test of whether one classification is more accurate than the other.

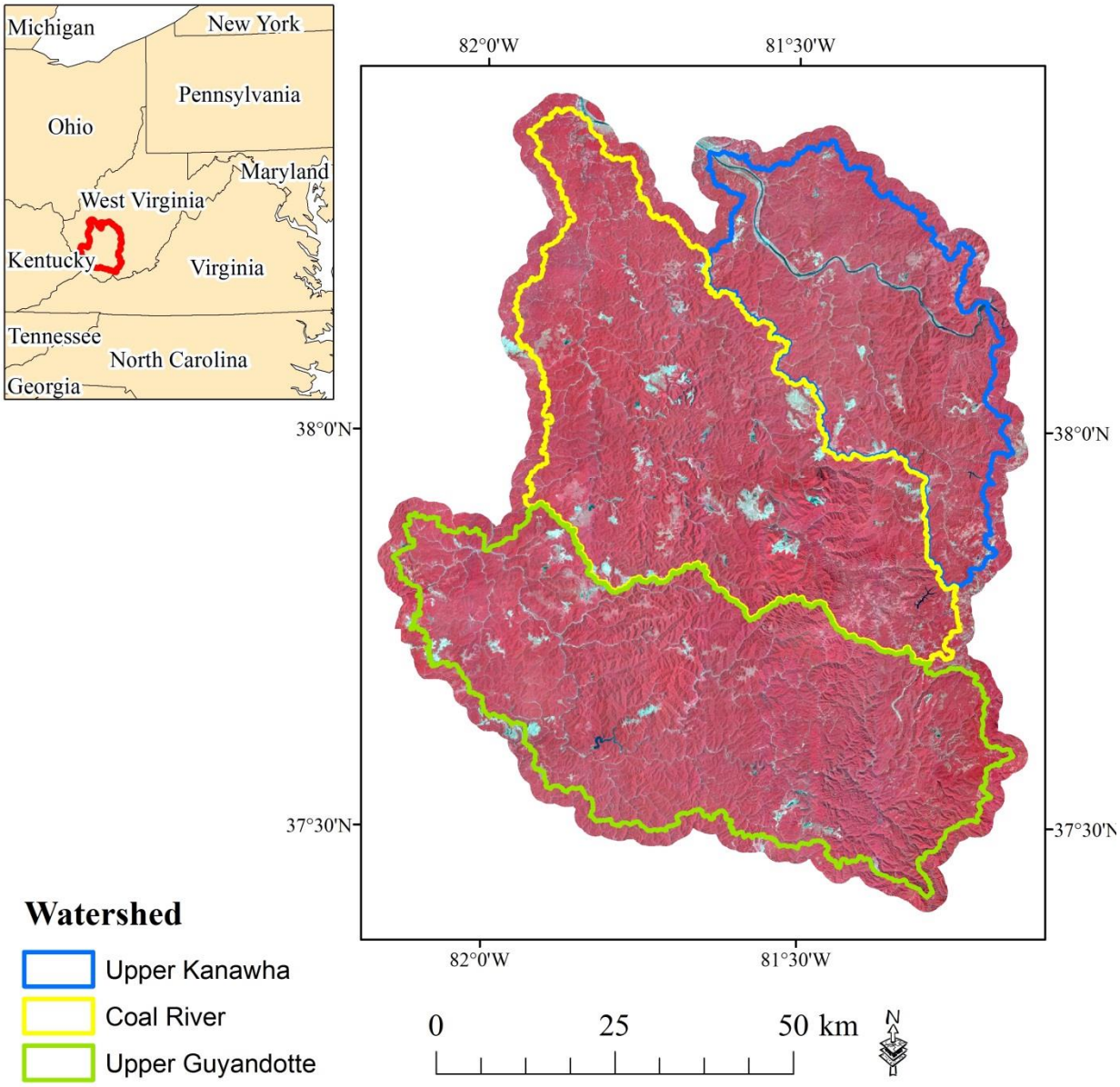

Figure 1: Location map showing the state of West Virginia and the study area extent. Base imagery is 2011 NAIP orthophotography displayed in false color (Bands 4, 3, 2 as red, green and blue). Large cyan patches generally correspond to active mining areas. Surface mining is extensive throughout these watersheds.

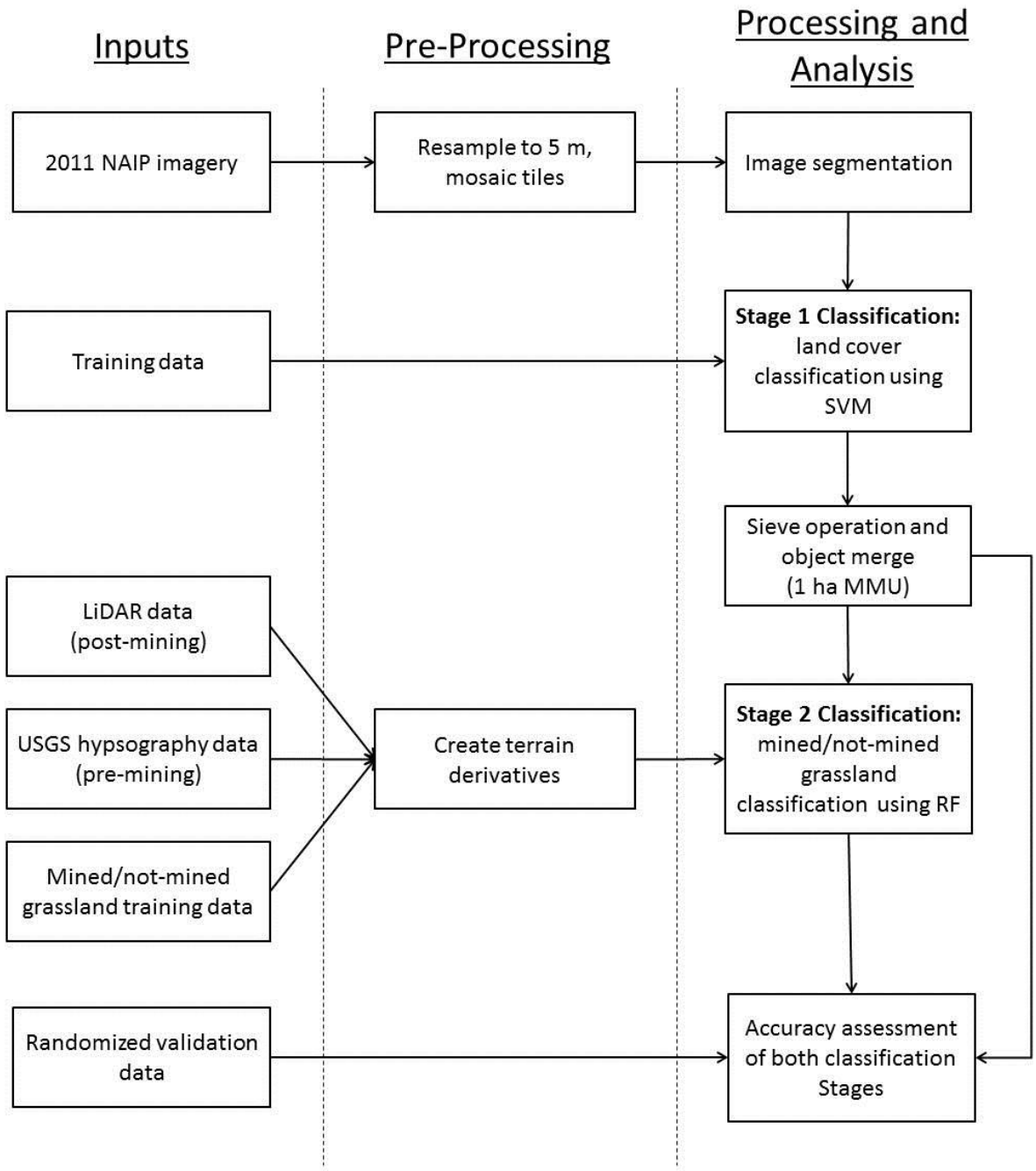

Figure 2: Overview of mapping and assessment process.

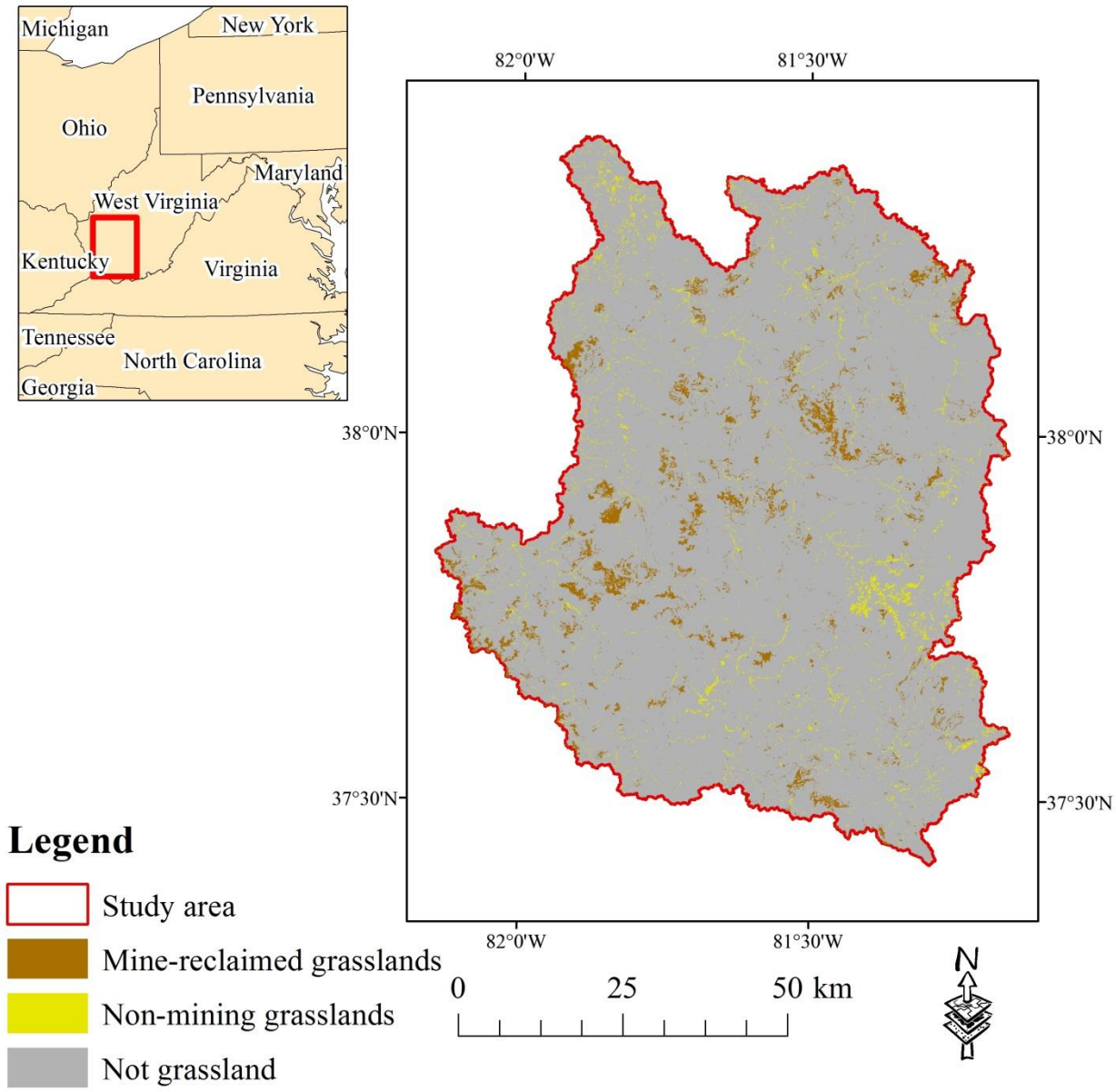

Figure 3: Grassland classification for entire study area using all predictor variables (All pre- and post-mining variables plus elevation change in the second stage of the classification).

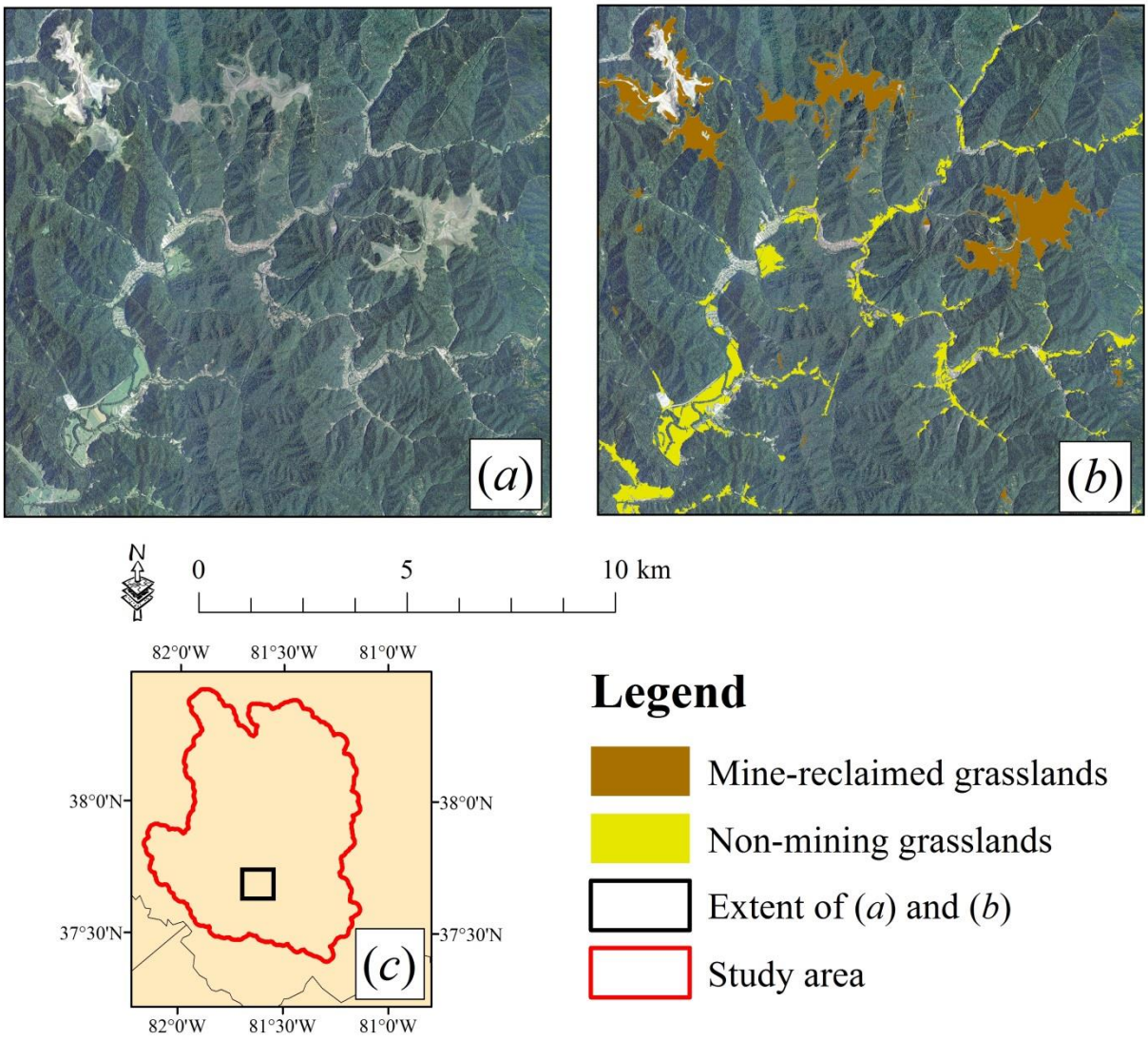

Figure 4 (*a*) NAIP simulated natural color image (bands 3, 2, 1 as red, green and blue). (*b*) NAIP image with example grassland classification using all predictor variables (All pre- and post-mining variables plus elevation change in the second stage of the classification). (*c*) Location map.

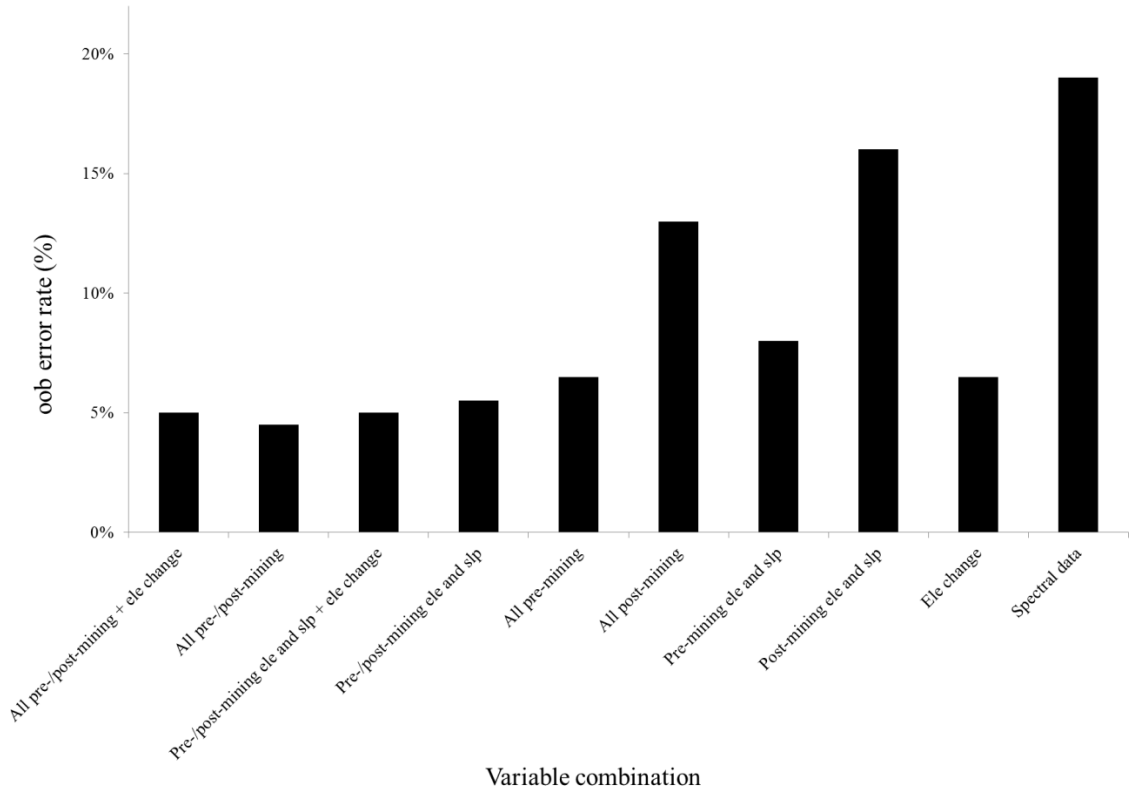

Figure 5: oob error rate estimated by RF algorithm for different input variable combinations.

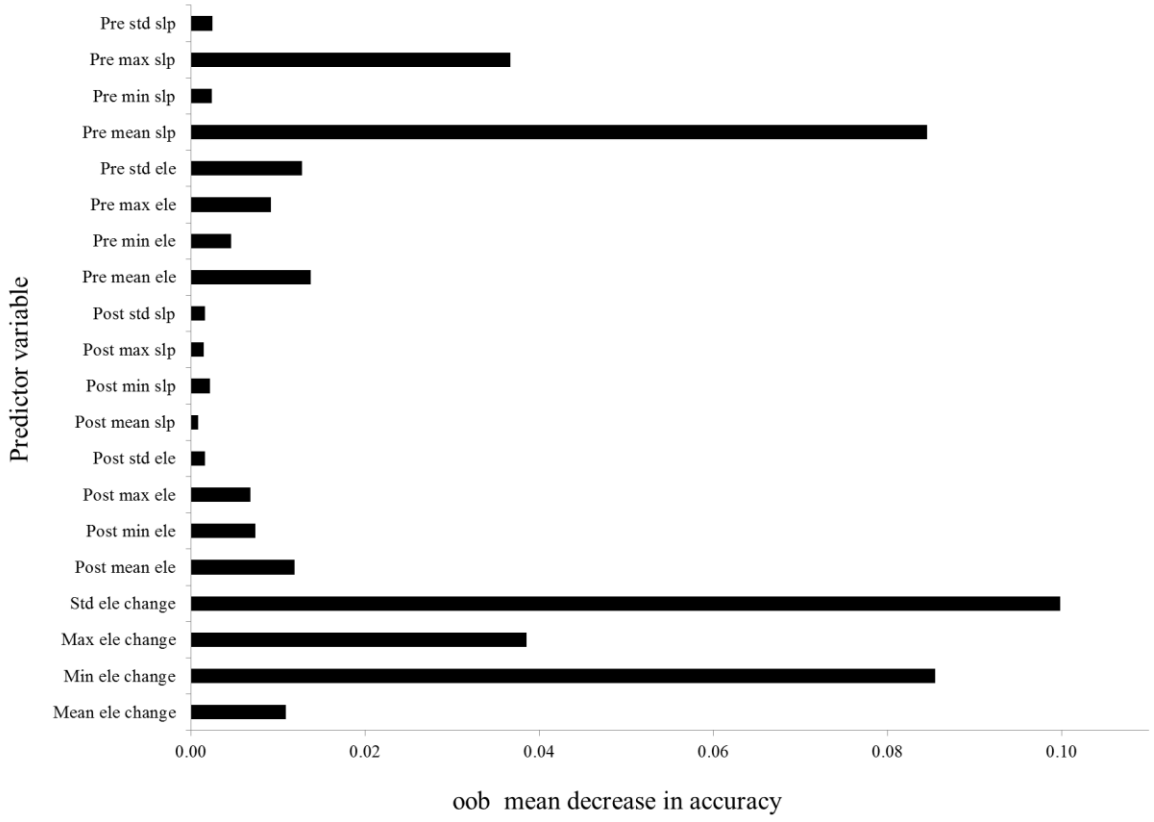

Figure 6: Variable importance as estimated by oob mean decrease in accuracy for model using all pre- and post-mining elevation (Ele) and slope (Slp) and elevation change (Ele Change) descriptive statistics.

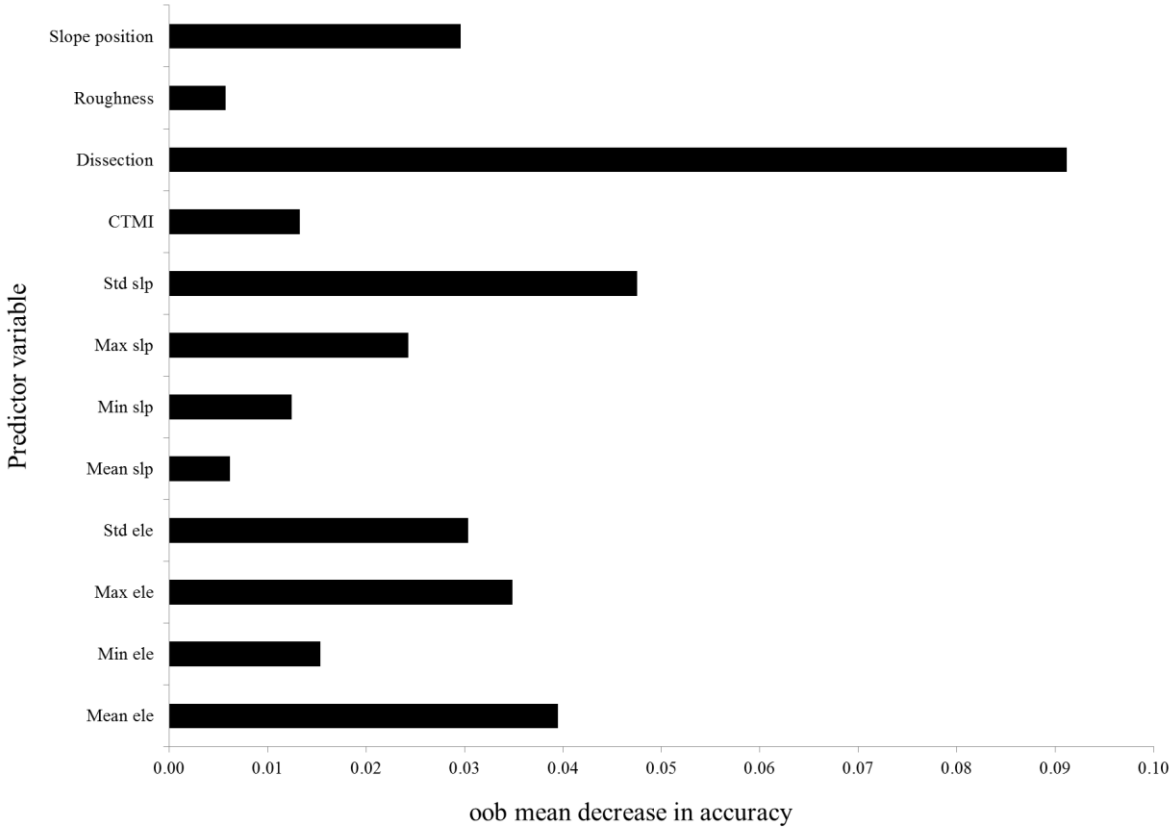

Figure 7: Variable importance as estimated by oob mean decrease in accuracy for model using all post-mining variables.

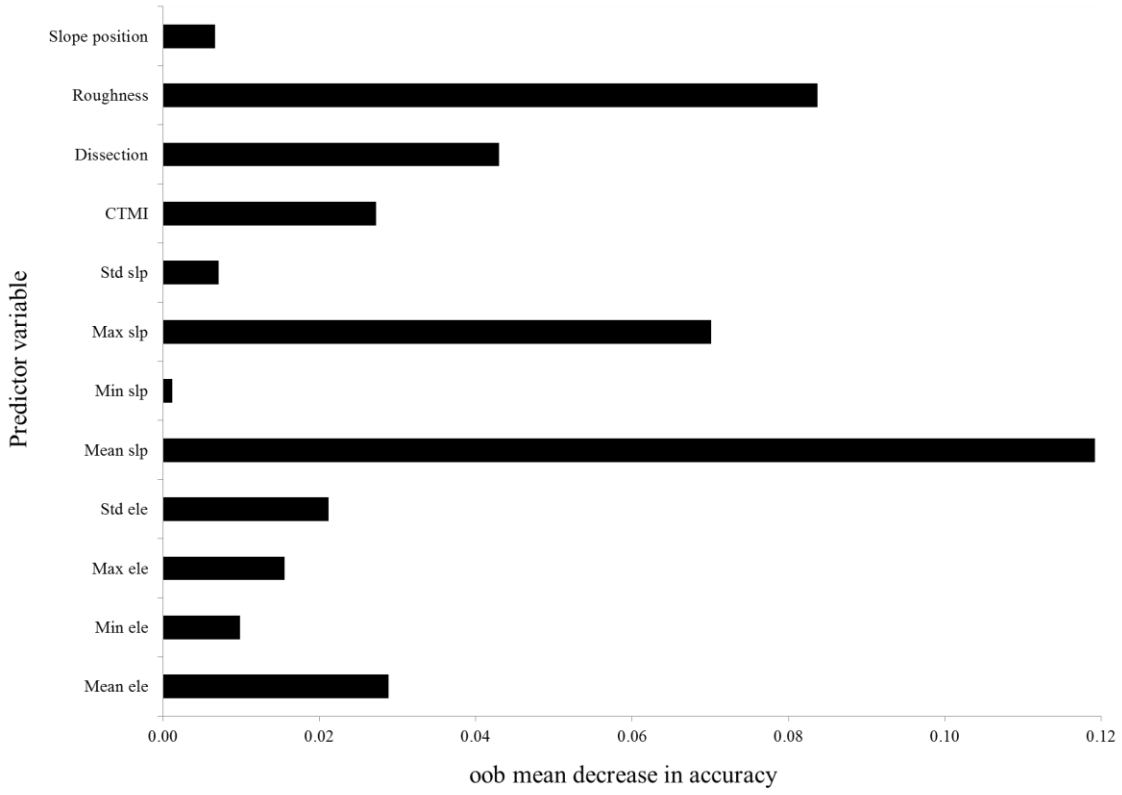

Figure 8: Variable importance as estimated by oob mean decrease in accuracy for model using all pre-mining variables.

CHAPTER 6

Conclusion

1. Synthesis of Results

The aim of this dissertation was to investigate high spatial resolution remotely sensed data (aerial- and satellite-based multispectral imagery, LiDAR, and terrain variables) and advanced classification methods (machine learning algorithms and GEOBIA classification) for mapping mine disturbance. The results of the research were presented in Chapters 2-5, in four separate articles, with overlapping themes. This concluding chapter will attempt to synthesize the research findings of those chapters.

In a comparison of NAIP orthophotography and RapidEye satellite imagery, NAIP data result in statistically lower classification accuracies in comparison to RapidEye data. This lower accuracy was attributed to inconsistent illumination and radiometric normalization, as the aerial data are acquired as individual scenes over an extended period of time (Chapter 3). Satellite data can offer radiometric consistency over large extents and repeat observations from similar viewing and illumination geometries. However, aerial imagery, specifically NAIP orthophotography, is not without benefits, including availability for multiple years, a general lack of cloud cover, contiguous coverage of large areas, availability, and low cost to the end user. Thus, although the classification of aerial imagery mosaics may be complicated by inconsistent illumination, the need for radiometric normalization, and low spectral resolution, NAIP data offer many favorable characteristics for mapping land cover.

The classification of NAIP imagery was generally improved by reducing the spatial resolution (e.g. resampling to a coarser cell size), decreasing the number of classes being mapped, and using the NIR band (Chapter 3). For example, NAIP imagery was found to provide accuracies above 90% when mapping only vegetation, barren areas, and water (Chapter 3) and

when separating herbaceous cover from other cover types (Chapter 5). This suggests that for certain classification tasks NAIP imagery may be an adequate, and even valuable, data source.

Combining LiDAR with multispectral imagery statistically improved the classification of mining and mine reclamation land cover in comparison to only using RapidEye satellite-based multispectral data, and this improvement was statistically significant for both pixel-based classification (Chapter 2) and GEOBIA classification (Chapter 4). For pixel-based classification, the nDSM data were found to be of particular importance (Chapter 2) while for GEOBIA classification both intensity and nDSM data were of importance (Chapter 4). This research suggests that the low spectral resolution and high heterogeneity inherent to high spatial resolution imagery can be mitigated by incorporating data from another sensor.

The incorporation of multi-temporal terrain attributes that characterized pre- and post-mining terrain characteristics calculated from DEMs allowed for accurate mapping of mine-reclaimed grasslands with user's and producer's accuracies above 80% and differentiation from spectrally similar non-mining grasslands (Chapter 5). This finding further supports the conclusion that combining disparate data sets is a means to improve classification accuracies in comparison to only using spectral data. This is especially true when attempting to map land use classes, which may be spectrally similar but may differ in other characteristics, such as terrain properties. Additionally, GEOBIA was a valuable tool for combining data collected using different sensors (i.e. the image and digital terrain data) and gridded at different cell sizes by providing a framework to characterize the terrain data using summary variables (such as mean, maximum, minimum, and standard deviation of the values) at the object scale (Chapter 5).

Generally, textural measures when used for pixel-based classification (Chapter 3) produced only a moderate increase in classification accuracy, and for GEOBIA classification

both first-order and second-order texture resulted in no improvement in accuracy (Chapter 4). Object-based geometric measures also did not improve the GEOBIA classification accuracy (Chapter 4). On the other hand, integrating information from LiDAR (i.e. another sensor) did result in a significant accuracy improvement. This lack of improvement using image spatial measures may not hold true for classification tasks of different landscapes. However, for the mapping of mining and mine reclamation using high spatial resolution multispectral imagery, incorporation of LiDAR was shown to be of greater value then the use of texture or GEOBIA geometry.

Generally, the SVM algorithm provided the best classification performance in comparison to RF and boosted CART for both pixel-based (Chapter 2) and GEOBIA (Chapter 4) classification. It also outperformed *k*-nn for GEOBIA classification (Chapter 4). However, SVM does have some challenges, for example the complexity of the optimization of the user-defined parameters. Although the other classifiers generally did not perform as well as SVM, they have other benefits. For example, boosted CART only requires a single user-defined parameter (the number of trees grown), which does not greatly affect the classification accuracy (DeFries and Chan, 2000; Muchoney et al., 2000; Friedl et al., 1999; Friedl et al., 2002; McIver and Friedl, 2002; Lawrence et al., 2004; Ghimire et al., 2012). The RF algorithm provides estimates of classification error (as used in Chapter 5) and an estimates of predictor variable importance (as used in all of the chapters) using the oob data (Breiman, 2001; Cutler et al., 2007), which was found to be of great value in this study for assessing the importance of predictor variables as image bands, LiDAR-derivatives, terrain variables, and summary statistics. Thus, it is important to consider other attributes of the classifier besides just classification accuracy. Generally, the input predictor variables used were of greater importance than the classifier chosen (Chapters 2

through 4), suggesting that data selection should be the prime focus in planning mining mapping projects.

The mapping methods outlined in this study may be applicable for other mapping projects and specifically the mapping of human-induced landscape disturbance. Although all mapping efforts are unique in terms of the input data being used, the spatial resolution of the data, the classes being mapped, and the intended use of the thematic classifications, many of the general findings of this dissertation may be applicable for improving land cover classification accuracy, such as the use of machine learning algorithms to leverage multiple datasets, combining high spatial resolution multispectral imagery with additional data (e.g. LiDAR), the value of GEOBIA for summarizing additional ancillary data for classification, and judicious pre-processing to potential improve the classification of aerial imagery. Also, this research generally suggests that remote sensing can be an accurate and valuable tool for mapping human-induced landscape change, which is important for documenting and confirming the Anthropocene as a geologic time period.

2. Practical Considerations and Limitations

Although the use of high spatial resolution remotely sensed data were shown to be of value for mapping mining and mine reclamation, especially when data from disparate sensors were combined, there are some practical considerations. First, satellite imagery is limited by availability and cloud cover. Cloud cover is less of a concern when using NAIP aerial imagery; however, this research suggests that satellite imagery is the optimal choice. Thus, data availability is of concern. Combining disparate data sets such as imagery and LiDAR improved the classification accuracy; however, there are practical limitations associated with combining data. For example, available data may not be temporally aligned (i.e. not collected at the same

time), which is of specific concern in a landscape in which land cover conditions can change rapidly as a result of mining and subsequent reclamation. Obtaining multiple data sets acquired within a short period may be difficult. In addition, implementation of machine learning algorithms and pixel-based and GEOBIA classification can be complicated by software limitations, data volume, and parameter optimization.

LULC mapping requires rigorous accuracy assessment procedures so that the end user has a reliable sense of the uncertainty of the classification. Rigorous accuracy assessment is also necessary in order to compare different classification approaches. Indeed, no LULC mapping project is complete without an accuracy assessment (Congalton, 1991). However, the accuracy assessment proved to be the most challenging phase of this research. Obtaining unbiased, accurate, and correctly proportioned validation data was difficult, especially considering that the goal of such assessments was to assess the map accuracy, not the average class accuracy. This requires that the training data be correctly proportioned so that the number of samples in each class is approximately proportional to the map area of each class. The use of GEOBIA classification further complicated the assessment as there is still active debate within the remote sensing community as to the correct assessment procedure for GEOBIA. At an even broader level, measures for assessing classification accuracy in general are also currently being debated. For example, the use of the Kappa statistic has recently been questioned and quantity and allocation disagreement have been suggested as an alternative (Pontius and Millones, 2011).

3. Final Remarks

High spatial resolution remotely sensed data are valuable for mapping and monitoring surface mining and mine reclamation, especially when combining data from multiple sensors, for example imagery and LiDAR. Although extracting information from such data can be

complicated by limited spectral resolution, high spatial resolution, and violation of multivariate normality assumptions, advanced classification techniques, including machine learning algorithms and GEOBIA, allow for such data to be classified. Using these classification techniques and data, remote sensing is a valuable tool for mapping and monitoring changing and complex landscapes, such as the southern coalfields of West Virginia.

4. Reference

Breiman, L., 2001. Random forests, *Machine Learning*, 54(1): 5-32.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications using remotely sensed data, *Remote Sensing of Environment*, 37: 35-46.

Cutler, D.R., T.C. Edwards, Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler, 2007. Random Forests for classification in ecology, *Ecology*, 88(11): 2783-2792.

DeFries, R.S., and J.C.W. Chan, 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data, *Remote Sensing of Environment*, 74(3): 503-515.

Friedl, M.A., C.E. Brodley, and A.H. Strahler, 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales, *IEEE Transactions on Geoscience and Remote Sensing*, 37(2): 969-977.

Friedl, M.A., D.K. McIver, J.C.F. Hodges, X.Y. Zhang, D. Muchoney, A.H. Strahler, C.E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf, 2002. Global land cover mapping from MODIS: Algorithms and early results, *Remote Sensing of Environment*, 83(1-2): 287-302.

Ghimire, B., J. Rogan, V. Rodríguez-Galiano, P. Panday, and N. Neeti, 2012. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA, *GIScience & Remote Sensing*, 49(5): 623-643.

Lawrence, R., A. Bunn, S. Powell, and M. Zambon, 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis, *Remote Sensing of Environment*, 90(3): 331-336.

McIver, D.K. and M.A. Friedl, 2002. Using prior probabilities in decision-tree classification of remotely sensed data, *Remote Sensing of Environment*," 81(2-3): 235-261.

Muchoney, D., J. Borak, H. Chi, M. Friedl, , S. Gopal, J. Hodges, N. Morrow, and A. Strahler, 2000. Applications of the MODIS global supervised classification model to vegetation and land cover mapping of Central America, *International Journal of Remote Sensing*, 21(6-7): 1115-1138.

Pontius, R.G., Jr., and M. Millones, 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment, *International Journal of Remote Sensing*, 32(15): 4407-4429.