Detecting changes in the extent of mountain top removal in southwestern West Virginia

Calla Little
Environmental Studies 372 – Final Project

University of Wisconsin-Madison
May 2015

Introduction

Coal removal in Appalachia using a technique called mountain top removal (MTR) has been a controversial topic for many years due to its economic and environmental impacts. MTR is a type of surface mining in which explosives are used to remove top soil and expose the coal beneath (Palmer et al. 2010). The blasted topsoil is pushed into surrounding valleys causing the landscape to be flattened and streams to be filled (Palmer et al. 2010). These practices have damaging effects on the surrounding ecosystem. The forests of these regions are some of the most diverse in North America (Palmer et al. 2010) and often times the same species are unable to repopulate the mined area due to compacted soils or the establishment of non-native, quickly growing, herbaceous species (Wilson-Kokes and Skousen 2014). Watersheds are also affected by this form of mining. The compacted soil and lack of vegetation result in decreased infiltration rates and higher run-off rates (Negley and Eshleman 2006). Surface mining has not only proven dangerous for the environment. The increased run-off rates, release of toxic material, and higher levels of airborne dust all have effects on human health (Palmer et al. 2010). Furthermore, the economic benefits of MTR in the region are questionable (Perdue and Pavela 2012). For these reasons, more insight into the extent and location of surface mining in Appalachia is necessary for a variety of applications ranging from ecology to disaster management.

At this time, the best data on the extent of mines for the West Virginian portion of Appalachia can be obtained as a shapefile from the West Virginia Department of Environmental Protection. Unfortunately, the shapefile does not accurately reflect the extent of mined area. Through examination of the data and satellite imagery, it became apparent that not all permitted areas were mined and that not all mined areas were permitted. Due to the impact of mining on the environment and human health and the inaccuracy of current data sets, the goals of this study are to map the extent of open and reclaimed mines and measure the area of mined land from 1984 to 2013 for a portion of southwestern West Virginia.

In order to assess mined area and changes in extent, I created classification scheme, masked out areas of non-interest, and employed a supervised classification technique to determine landcover for four dates (1984, 1991, 2003, 2013). The first three images were collected from Landsat5 TM and the last date was taken from Landsat8 OLI. After classification I performed an accuracy assessment on each classification and calculated area of open and reclaimed mines for each date.

Background

There have been a variety of studies that have attempted and successfully used remote sensing to study the landcover and changes in extent of open mines. A number of studies have successfully mapped mines using a variety of supervised classification techniques ranging from parallel-piped (Anderson et al. 1977) to maximum likelihood (Schmidt and Glaesser 1998, Irons and Kennard 1986).

Later studies have focused on characterizing the landscape as well as detecting changes in landcover throughout time. Prakash and Gupta (1998) used image differencing and the normalized difference vegetation index (NDVI) with Landsat TM data to identify mining extent changes in India. Demirel et al. (2011) were able to successfully use high resolution, 1m to 4m data and the Support Vector Machine (SVM) classification method to quantify the changes in landcover surrounding an open cast mine in Goynuk, Turkey. Latifovic et al. (2005) quantified landcover change in the Athabaska Oil Sands of Canada by examining vegetation trends in the region using data from Landsat5 TM and Landsat7 ETM+ and a postclassification change detection method. Townsend et al. (2009) determined changes in land cover from MTR within a portion of the Central Appalachians in order to understand the effect mining has on hydrology.

Most studies have had success in separating mined areas from unmined areas, except for in regions with spectrally similar classes (e.g. urban or bare agricultural fields). There has been less attention paid to reclaimed areas and certainly more difficulty in accurately detecting them. Reclaimed grasslands are often confused with pastures and reclaimed forests are very similar in spectral signatures to stable forest. Townsend et al. (2009) found that masking unmined areas and using images from multiple dates to identify formerly mined areas was a successful approach to overcoming these issues.

Study Area

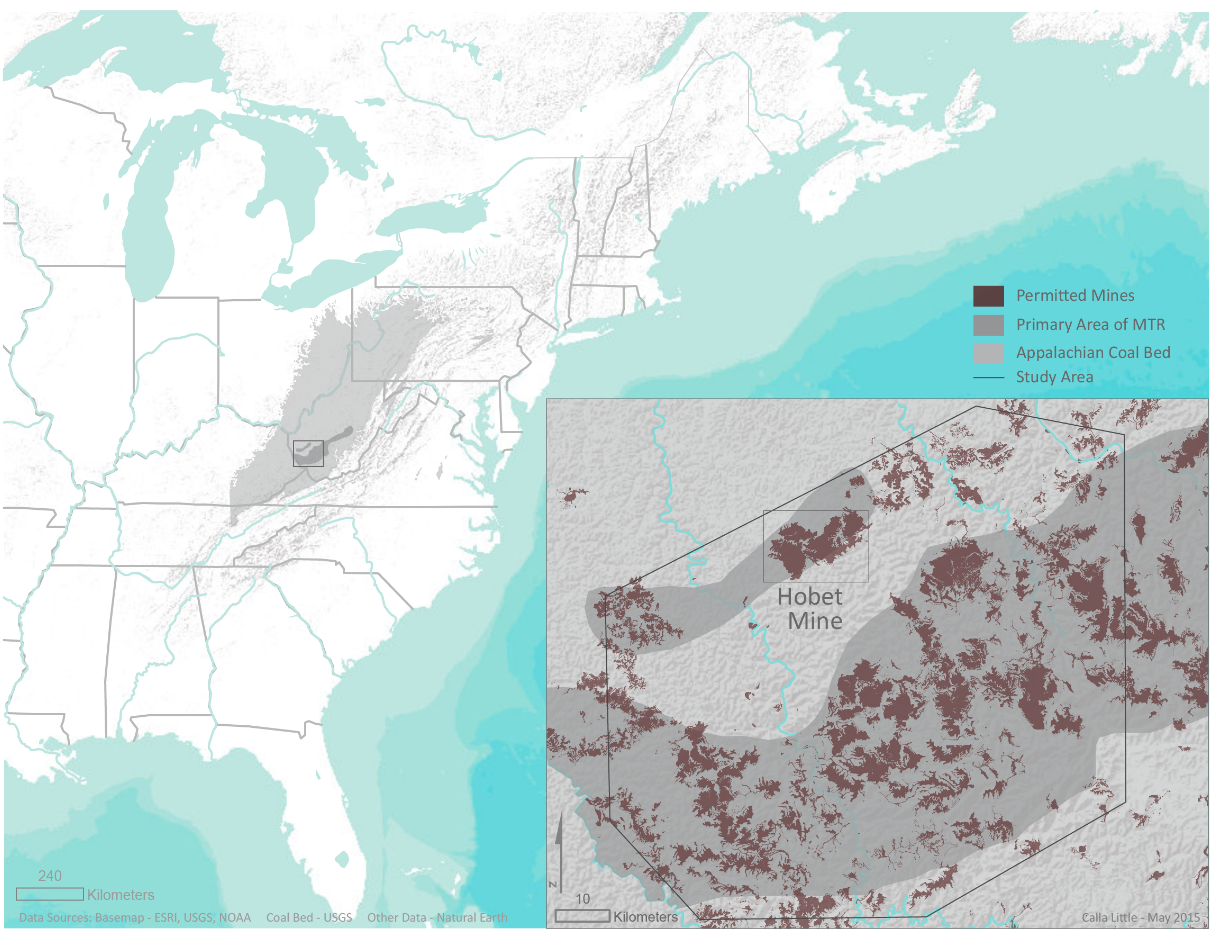

Figure 1 – Map of the study area and primary coal removal areas

The study area is a portion of a Landsat image (path 18, row 34) covering seven counties in the southwestern portion of West Virginia. This portion of the state is a well-known region for MTR and is situated over the Pond Creek coal zone. The area is mountainous and mostly forested. There are no major urban areas and the economy relies heavily on coal mining (Perdue and Pavela 2012). Figure 1 shows the extent of my study area. This area covers a large portion of West Virginia where MTR mining is most intensive.

Methods

I acquired imagery for four different dates spanning from 1984 to 2013. In order to separate bare mines from vegetated areas it was necessary to choose dates when vegetation was present. September seemed to be the best month for acquiring cloud free images and I was able to obtain anniversary dates in September for the 1984, 1991, and 2013 images. Due to the size of my study area it was difficult to find acceptable cloud free images, so for 2003 it was necessary to use an image from June.

In order to map the mined and reclaimed land present in the study area, I first needed to separate it from the other landcover present in the scene. I inspected each of the Landsat images in conjunction with high resolution Google Earth data and created a classification scheme to represent the area. My seven classes were (i) urban (ii) water (iii) bare mines (iv) forest (v) reclaimed grassland (vi) reclaimed forest and (vii) pasture.

Since it was only my goal to map current mines and reclaimed mines, I first created an unmined area mask to eliminate the areas of noninterest. I achieved this by choosing training sites to represent each of my classes and then using a supervised classification machine learning algorithm called support vector machine (SVM). SVM classifies data by finding the optimal place to split two classes. It determines the best place by maximizing the distance between the nearest data points (the support vectors) and the line that splits the data (the hyperplane). I chose SVM because it can handle complex data by splitting non-linear boundaries. It does this by applying a user specified kernel function, which transforms the data into a higher dimension where linear splitting is possible. After SVM uses the training data to determine class definitions, it iterates through the remaining pixels in the image and classifies them based on their probability of membership. The user can manipulate the probability threshold, which determines the lowest probability a pixel may have in order to be assigned to a class. One can also manipulate the penalty parameter, which determines how much misclassification is permissible when creating the hyperplanes. I chose to use the radial basis function kernel, a penalty parameter of 100, and a classification probability threshold of 0.

With these parameters, I ran SVM on an image stack consisting of all the dates. I did this hoping that it would help differentiate reclaimed forest from stable forest by taking into account the stable temporal signature of the consistently forested areas. This output a general classification of the area from which, due to confusion between other spectrally similar classes (e.g. bare mine and urban), I chose to use only the stable forest class to build a mask. I eliminated other areas of non-interest by combining the stable forest class with an urban vector layer and county roads vector layers (acquired from the U.S. Census Bureau) with 30m buffers applied to them. I then applied this mask to each one of the image dates.

After masking out areas of non-interest I was able to refine my class scheme. My new class scheme varied somewhat for each date and consisted of (i) water (ii) reclaimed grassland (iii) reclaimed forest (except 1984) (vi) bare mine and (v) forest (except 2013). I then classified each date by choosing training sites for each of the classes in each of the images and applying the SVM algorithm using the same parameters as described above.

After classification I performed an accuracy assessment on each date's classification and on the mask I created. The accuracy assessment on the mask consisted of 98 randomly selected pixels and the assessments of the classifications consisted of 107 randomly selected pixels within the portion of the image that was unmasked. I labeled each of the 107 locations and generated a confusion matrix. From the confusion matrix I calculated overall accuracy, the kappa statistic, user's accuracy, and producer's accuracy.

Results

Figure 2 is a graph showing the change in the area of bare mines for each of the four image dates. The greatest increase occurred between 1991 and 2003, which was expected since these image dates are furthest apart. When looking at the slope of the line, we can see that the area of bare mines between these two dates also showed the fastest growth rate. Even though the growth rate has slowed since 2003, the total area of bare mines is still increasing and is at its greatest extent with the total area more than doubling since 1984 to 106km^2.

Figure 3 is a graph showing the area of mined and reclaimed land for each of the four study dates. Not surprisingly the area of reclaimed and total mined land increases each year. We can see that the area of reclaimed grassland is increasing much quicker than the area of open mines. This may indicate that open mines exist on the landscape for a relatively short period, but that it takes a considerably longer amount of time for the vegetation to return to a spectrally indistinct state (i.e. reclaimed forest). We can also note the total area affected by mining for each year, with the area in 2013 being nearly 5 times greater than the total area in 1984.

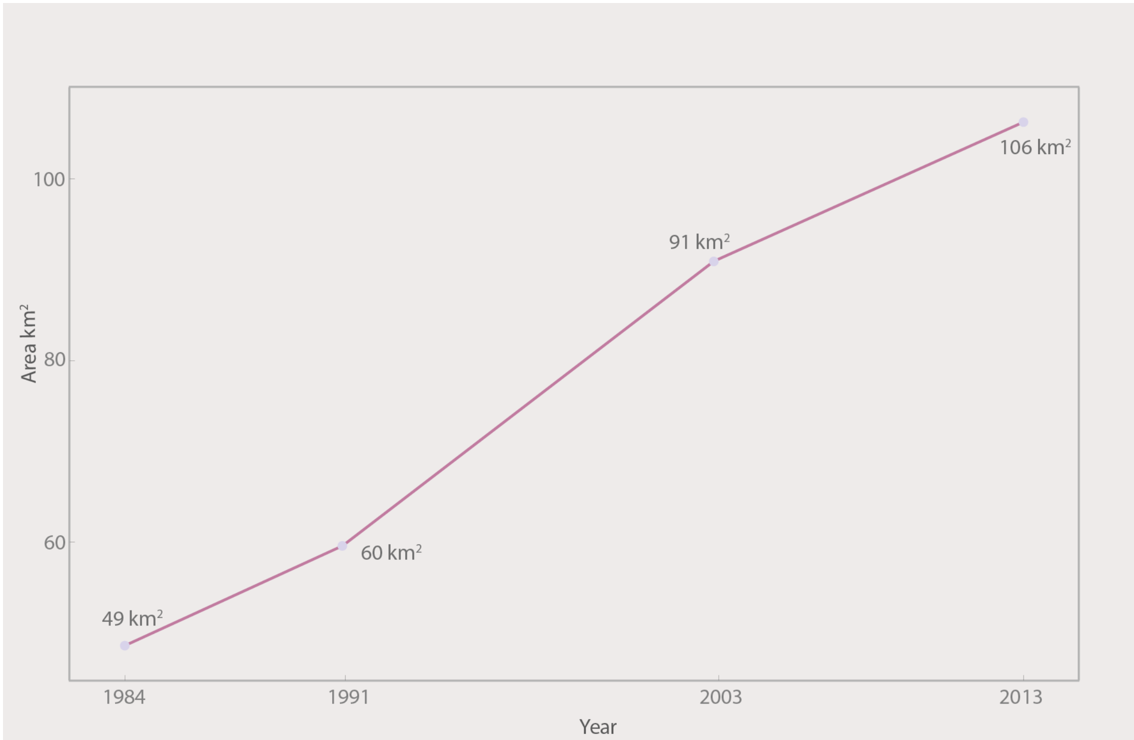

Figure 2 – The growth trend and amount of land covered by bare mines.

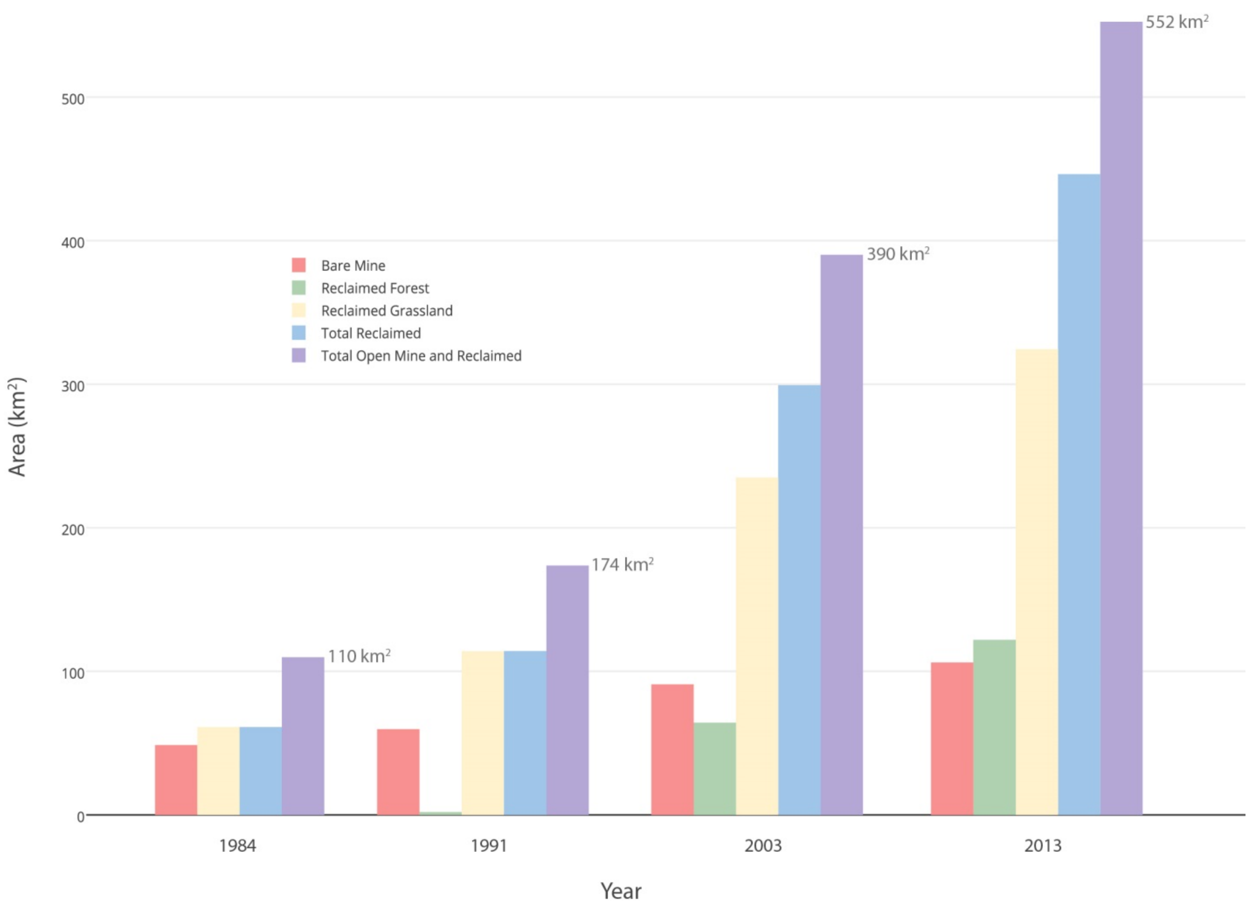

Figure 3 – The area covered by bare mines and reclaimed mines in km^2

Figure 4 shows the classification results of Hobet Mine (see Figure 1 map of study area for location), one of the largest mines in the study area. We can see the expansion of the mine throughout the study period, where in general an active mine will move to reclaimed grassland in the following date, then to reclaimed forest at the next point in time. It appears that the active mine almost always becomes reclaimed grassland, but that the reclaimed grassland is not as predictable. Sometimes it will move back to bare mine and other times it may remain grassland instead of moving to reclaimed forest. When comparing this classification to the Landsat imagery and Google Earth imagery, it seems to represent the area well. The main issues occurring are in the form of speckle throughout the unmined regions where the unmined area mask failed. For example, running north-south through the center of the image a line is visible, most noticeable as reclaimed grassland in the 2013 image, where a power line was unable to be separated from the temporal signature of a mine during the masking process. From the visibility of the access roads running throughout the reclaimed areas in the later images, we can see that the classification seemed to perform rather well when separating between reclaimed grassland and bare areas.

Table 5 shows the accuracy assessment for the unmined area mask. I achieved an overall accuracy of 98%. It appears that my mask performed rather well, but that I included some mined areas in the mask. It should also be noted that my study area was quite large, yet my accuracy assessment used only 98 points to determine the accuracy, and so these results are likely somewhat inflated.

Table 6 shows the accuracy assessments for each classification. 2013 had the lowest accuracy with only 66% of the pixels being correctly classified. 2003 and 1991 had better success with around 75% accuracy and 1984 showed the best results with 86% accuracy. It

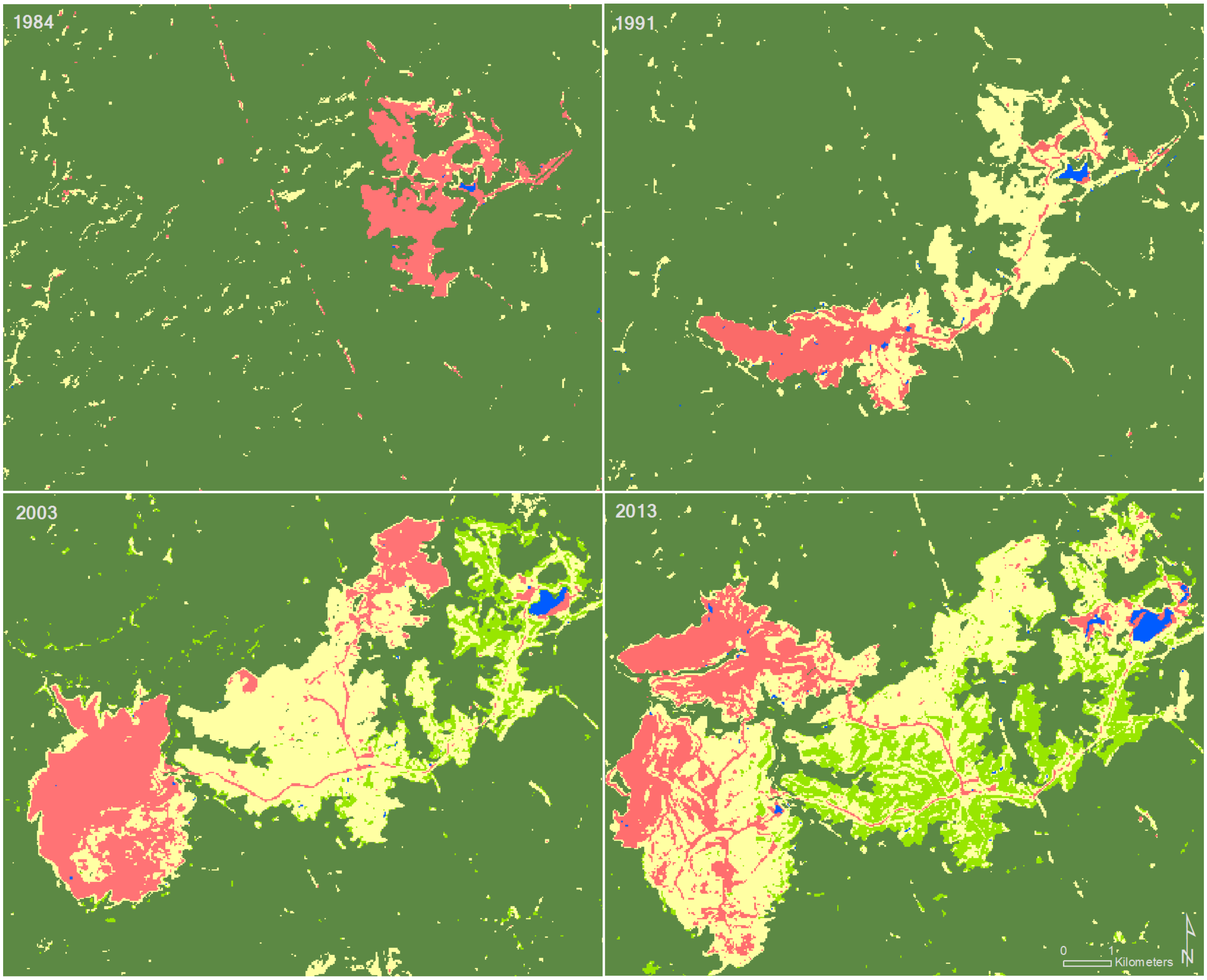

Figure 4 – The final map created from the classification of the study area. This is a portion of the study area called Hobet mine.

appears that the most error occurred when trying to classify reclaimed forest. The overall accuracy is also being affected by a number of unmined pixels that should have been masked out. Even though 1984 showed the highest overall accuracy, it performed the poorest in

	Unmined	Mined	Total	User's
Unmined	81	2	83	96%
Mined	0	15	15	100%
Total	81	17	98	
Producer's	100%	88%		
Overall Accuracy:		98%		
Kappa Coefficient:		0.925		

Table 5 – Confusion matrix for unmined area mask

the class of most interest, bare mine. All other dates had excellent user's accuracy for bare mines, and somewhat lower producer's accuracy, indicating that in general most errors were errors of omission.

Discussion

Estimates for total bare mined area appear to be relatively trustworthy, if not somewhat underestimated. The low errors of commission indicate that almost all of the area that was classified as bare mine was actually bare mine; therefore, the area that was calculated from these classifications is not an inflated estimate. Most errors occurring within the bare mines class are errors of omission, which means that the area estimates may actually be lower than reported by this study. In addition to the errors of omission, the many tailing ponds throughout the study area (classified as water) were also omitted from bare mine area calculations. While these are not technically bare mines, they do fall into the category of active mines and, thus, mined area. If these were included in the mine statistics, there would be approximately 30 more square kilometers of mined area present in the study area.

Area estimates for reclaimed grassland and reclaimed forest are not nearly as accurate. Reclaimed forest, in particular, is inaccurate, especially in the earlier dates. This was due to the fact that reclaimed forest training sites were difficult to create without earlier imagery. I struggled to find good training sites to represent the class. In the later dates, where I could use past image dates to identify reclaimed forest, it performed much better, but there were still issues with separating it from forested areas. There may have been more success with separating the two if I had used multi-year image stacks to capture temporal signature of the transition from grassland to forest.

1984	Water	Reclaimed Grassland	Bare Mine	Forested	Unmined	Total	User's
Water	0	0	0	0	0	0	0%
Reclaimed Grassland	0	10	0	0	0	10	100%
Bare Mine	1	0	3	0	2	6	50%
Forested	0	1	3	78	8	90	87%
Unmined	0	0	0	0	0	0	
Total	1	11	6	78	10	106	
Producer's	0%	91%	50%	100%			
Overall Accuracy	85.90%						
Kappa Coefficient	0.6093						

1991	Water	Reclaimed Grassland	Reclaimed Forest	Bare Mine	Forested	Unmined	Total	Users
Water	1	0	0	1	0	0	1	50%
Reclaimed Grassland	0	15	0	4	2	3	44	63%
Reclaimed Forest	0	0	0	0	0	0	9	0%
Bare Mine	0	0	0	15	0	0	14	100%
Forested	0	2	3	0	54	7	39	82%
Unmined	0	0	0	0	0	0	0	
Total	1	17	3	20	56	10	107	
Producer's	100%	88%	0%	75%	96%			
Overall Accuracy	79.44%							
Kappa Coefficient	0.6658							

2003	Water	Reclaimed Grassland	Reclaimed Forest	Bare Mine	Forested	Unmined	Total	Users
Water	1	0	0	0	0	0	1	100%
Reclaimed Grassland	0	30	4	8	0	2	44	68%
Reclaimed Forest	0	0	6	0	2	1	9	67%
Bare Mine	0	0	0	14	0	0	14	100%
Forested	0	0	2	1	29	7	39	94%
Unmined	0	0	0	0	0	0	0	
Total	1	30	12	23	31	10	107	
Producer's	100%	100%	50%	61%	74%			
Overall Accuracy	74.77%							
Kappa Coefficient	0.6597							

2013	Water	Reclaimed Grassland	Reclaimed Forest	Bare Mine	Unmined	Total	Users
Water	0	0	0	0	2	2	0%
Reclaimed Grassland	1	35	12	7	5	60	58%
Reclaimed Forest	0	7	11	0	2	20	55%
Bare Mine	0	0	0	25	1	26	96%
Unmined	0	0	0	0	0	0	
Total	1	42	23	32	10	108	
Producer's	0%	83%	47%	78%			
Overall Accuracy	65.74%						
Kappa Coefficient	0.5034						

Table 6 – Confusion matrices for each of the classification dates, starting with 1984 on the top to 2013 on the bottom

Even though estimates for reclaimed grassland and reclaimed forest individually were poor, they were improved when combined. For example, the accuracy assessment for 2013 was the lowest of all the classifications, but when I recalculated the accuracy with reclaimed grassland and reclaimed forest merged, the overall accuracy became 83%. Furthermore, there was not much confusion between the mined classes (i.e. bare, reclaimed grassland, and reclaimed forest) and the unmined classes (i.e. forested and unmined). Most of the confusion occurring, with the exception of the 2013 classification, happened within those broad categories. With that in mind, I believe that estimates of total reclaimed area and total area mined are reasonable.

It seems that most the error in the classifications was caused by the failure of the unmined area mask. I think it is safe to assume that the 98% accuracy assessment was inflated. When we look at the individual classification accuracy assessments, which were taken from a much smaller area and so are more likely to be representative, we can see that 10 of the pixels that identified as unmined in the accuracy assessment (approximately 11%) were included into classes they shouldn't have been. Looking at the 2013 assessment, where confusion with mined areas and areas that should have been masked was most common, it seems that they most often fell into the reclaimed grassland class. This is likely due to the illumination angle effects of the terrain in this mountainous region. Prior to using a multidate image stack to create the mask I attempted to perform a mask on a single date. The biggest issue with it was the inability to separate out brightly reflecting forested ridges from reclaimed grassland. It appears that this was still the biggest issue in my multidate mask.

Conclusion

Most errors in the classifications were a result of mask failure, however, in general these errors were not detrimental to mined area calculations, especially the bare mine area estimates.

An improved mask would yield higher overall accuracy, but would likely not affect area estimates that greatly. The accuracy of bare mined area is if anything underestimated due to errors of omission and the area continues to increase each year, with 2013 being the greatest so far. The slow regeneration of forests on these mined areas indicate that the mines have a lasting impact on the land and are distinguishable for many years after their initial active period. As total mined area continues to increase, further studies should be conducted to determine the impacts these mines may have on the environment and human health.

Anderson, A.T., D. Schultz, N. Buchman, H.M. Nock, 1977, “Landsat imagery for surface-mine inventory,” *Photogrammetric Engineering and Remote Sensing*, 43 (8): 1027–1036

Demirel, N., M. K. Emil, and H. S. Duzgun, 2011, "Surface Coal Mine Area Monitoring Using Multi-temporal High-resolution Satellite Imagery," *International Journal of Coal Geology*, 86.1:3-11.

Irons, J. R., and R. L. Kennard, 1986 "The Utility of Thematic Mapper Sensor Characteristics for Surface Mine Monitoring," *Photogrammetric Engineering and Remote Sensing,* 52.3: 389-96.

Latifovic, R., K. Fytas, J. Chen, and J. Paraszczak, 2005, "Assessing Land Cover Change Resulting from Large Surface Mining Development," *International Journal of Applied Earth Observation and Geoinformation*, 7.1:29-48.

Negley, T. L., and K. N. Eshleman, 2006, "Comparison of Stormflow Responses of Surface-mined and Forested Watersheds in the Appalachian Mountains, USA," *Hydrological Processes,* 20.16: 3467-483.

Palmer, M. A., E. S. Bernhardt, W. H. Schlesinger, K. N. Eshleman, E. Foufoula-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* 327.5962: 148-49.

Perdue, Rt, G. Pavela, 2012, “Addictive Economies and Coal Dependency: Methods of Extraction and Socioeconomic Outcomes in West Virginia, 1997-2009”, *Organization & Environment,* 25.4:368-384

Prakash, A., and R. P. Gupta, 1998, "Land-use Mapping and Change Detection in a Coal Mining Area - a Case Study in the Jharia Coalfield, India," *International Journal of Remote Sensing,* 19.3: 391-410.

Schmidt, H., and C. Glaesser, 1998, "Multitemporal Analysis of Satellite Data and Their Use in the Monitoring of the Environmental Impacts of Open Cast Lignite Mining Areas in Eastern Germany," *International Journal of Remote Sensing,* 19.12: 2245-260.

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,* 113.1: 62-72.

Wilson-Kokes, L., and J. Skousen, 2014, "Nutrient Concentrations in Tree Leaves on Brown and Gray Reclaimed Mine Soils in West Virginia," *Science of* the *Total Environment,* 481: 418-24.