Findings of the Wiscland-2 pilot project

Prepared by Wiscland-2 Team
Contact: Howard Veregin
veregin@wisc.edu

August 31, 2014

1. Introduction

The purpose of this document is to report on the findings of the pilot investigation covering the northern and southern test sites for the Wiscland-2 land cover mapping project. The pilot investigation was conducted to study input requirements, the classification scheme and associated accuracy, and to examine whether existing point/polygon-based datasets, such as Forest Inventory Analysis (FIA) and Continuous Forest Inventory (CFI), are useful in remote sensing based map-making. The preliminary findings reported here address these issues in order with the supporting evidence in the form of figures and tables.

2. Methodology

2.1 Data

Landsat

The original WISCLAND product (Wiscland-1) was derived from Landsat Thematic Mapper (TM) data ca. 1992 at 30-meter nominal spatial resolution. It is recommended that the new update to the land cover layer continue to be made from Landsat data, using the follow-up Landsat mission data such Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor and/or newly launched (Feb 11, 2013) Landsat 8 Operational Land Imager (OLI). Both of these sensors continue to provide observations at 30-meter spatial resolution, but with much improved signal-to-noise ratio. The Landsat 7 ETM+ sensor experienced an unrecoverable technical problem in May 2003 that rendered all images acquired after that date to have data gaps. These data gaps cover an estimated 22 percent of any given Landsat 7 scene and are most pronounced along the edge of the scene and gradually diminish toward the center. The maximum width of the data gaps along the edge of the image would be equivalent to one full scan line, or approximately 390 to 450 meters (13 to 15 pixels). The precise location of the missing scan lines will vary from scene to scene. An area approximately 22 kilometers wide in the middle of the scene contains no data loss, and this region of each image is very similar in quality to images acquired before the technical problem occurred. In fact, data quality in non-gap areas are of science quality, without any loss in radiometric fidelity, assessed by both NASA and USGS engineers.

During the period between May 2003 and December 2011, Landsat 7 acquisitions with missing data were complemented with Landsat 5 TM acquisitions to alleviate the data gap issue but because of diminished on-board recording capabilities of Landsat 5, acquisitions were somewhat limited. Between December 2011 and May 2013, Landsat 7 ETM+ data with gaps were the only source of observations. Since May 2013 (with some acquisition going back to March 2013), the Landsat 8 OLI instrument has been operational and the data archive is growing. Both of these sensors continue to provide the similar but improved spectral capabilities of the TM sensor used in Wiscland-1. In light of these new developments, generating a minimum 30-meter spatial resolution land cover product for the proposed update is technically feasible, and practically possible. However, the pilot study was conducted using Landsat 5 TM and Landsat 8 OLI instrument data only, spanning several years centered around year 2010. The footprints of interest were Landsat path/row 25/28, located in the north-central part of the state and Landsat path/row 24/30 in the south-central part of the state.

SPOT

The DNR Wisconsin Land cover business needs document highlighted the desire for a map at 10-meter spatial resolution, at least for select locations such as urban areas and small vegetation patches. We have identified a free-of-charge limited data source, the Satellite Pour l'Observation de la Terre (SPOT), a French satellite system that has been in existence since the early 1980s. It provides data in several spectral bands at 5-20 meter spatial resolution, depending on the satellite (SPOT-6 is the latest satellite in the series). As a private entity, SPOT satellite data are not free and are purchased through various channels in the US. However, as part of the *North America Data Buy Program*, USGS and NASA have purchased a large archive of SPOT data for every location in the US from January 1, 2010 through June 20, 2013 and are making these data available to federal and state government agencies at no cost (https://lta.cr.usgs.gov/SPOT_NADB).

In terms of cost, it is possible that the entire archival data available for Wisconsin is available for free. However, free availability of SPOT data does not mean it is the best (or first) choice for updating Wiscland-1. In our view, the success of the update will depend highly on the temporal availability of images throughout the growing season, including at least one early, one mid, and one late season image.

Figure 1 shows the spatial distribution of SPOT footprints across Wisconsin. First to notice is that they are smaller than Landsat footprints, covering an area about an eighth of the area covered by a Landsat scene. Second, approximately 100 individual SPOT footprints are required to cover the entire state as opposed to less than 20 individual Landsat scenes. Figure 2 depicts the temporal availability of SPOT data between 2009 and 2012 over Wisconsin. One interpretation of this figure is that the temporal availability of SPOT data is not uniform across the state; some locations have greater than 15 images across three years while other locations have less than five, which may make it hard for accurate classification.

Additionally, this unequal temporal distribution is also reflected in the distribution of images across the growing season. Figure 3 shows that histogram distribution of all 500 SPOT images across a full year, combining images from 2009 to 2012. As shown, the frequency is loaded towards peak summer (July) images

To summarize, although we have identified a data source, free-of-charge, to be French SPOT system, a detailed investigation of the temporal availability (i.e. repeat coverage) of SPOT data does not appear to be sufficient, as delineating many of the vegetated classes require multiple views of the surface through the year to capture the differences in their phenological cycle. This assessment was upheld by the Science Advisory Committee, where it was determined that 30m resolution (such as Landsat) is typically sufficient for as statewide project, and the lack of consistent temporal and spectral coverage of Wisconsin by the SPOT system would make it difficult to utilize on a statewide basis.

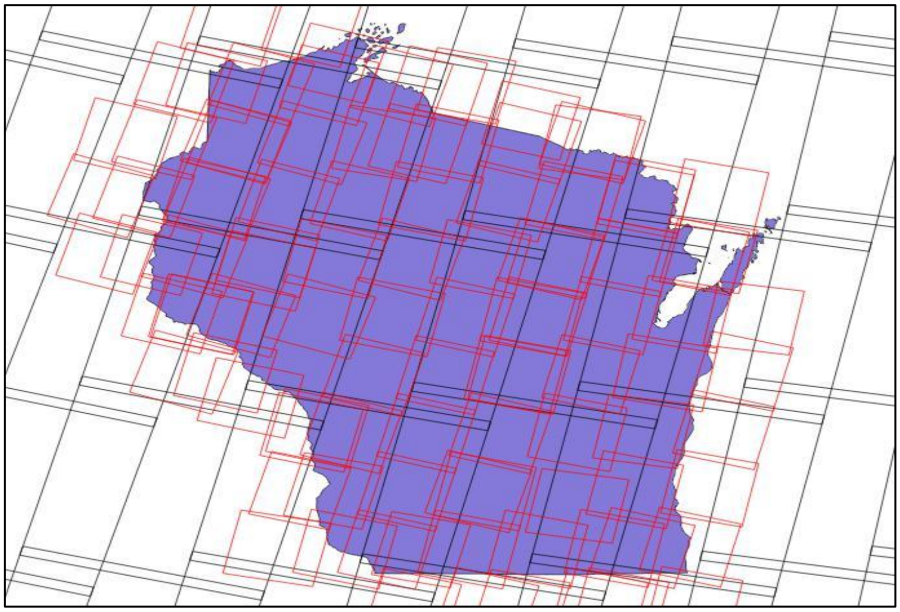


Figure 1. Landsat (black) and SPOT (red) individual image footprints over Wisconsin.

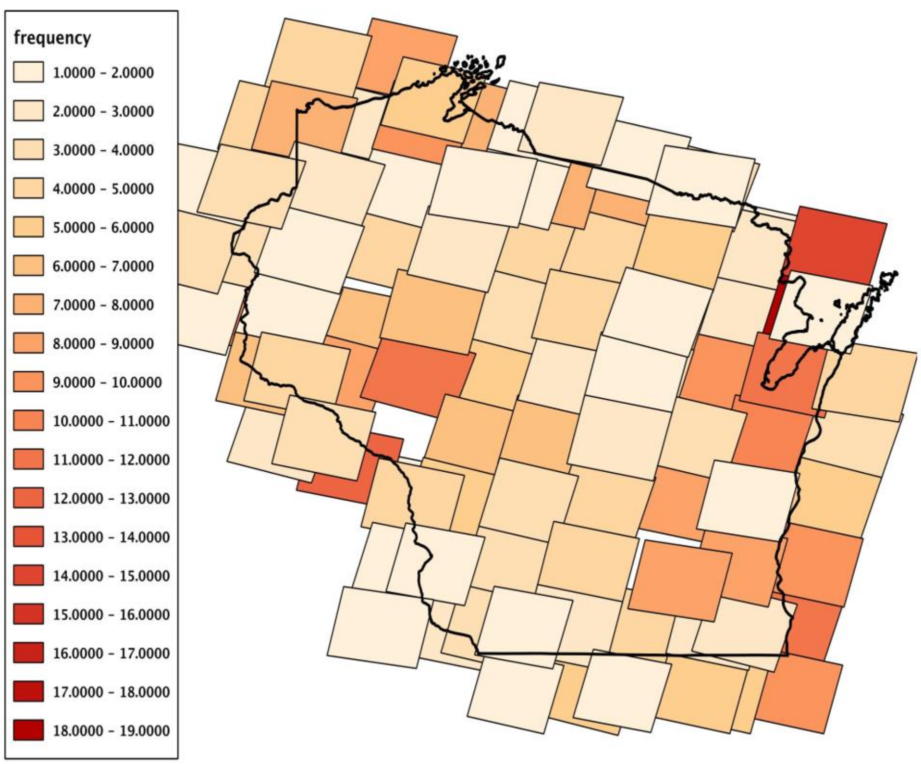


Figure 2. Temporal distribution of SPOT data for every footprint across Wisconsin. The legend and the color indicate the number of individual images available for that footprint between 2009 and 2012. Note that this information only reflects the data available from USGS as part of their data buy program. It is likely that there are more images available but these images are not found in the USGS archives.

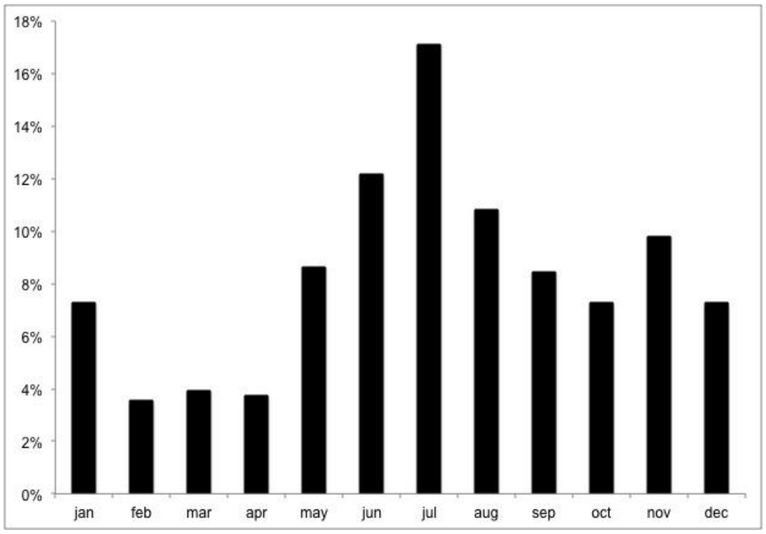


Figure 3. Histogram distribution of SPOT images by month across three years (2009-2012).

In terms of data used in the pilot, Tables I and II list all the sources used. A total of 142 inputs were used for the northern site and 166 inputs were used for the southern site in the pilot study, majority of which was from satellite observations.

Table I. Inputs used in the pilot study for the northern test site.

Dataset	Sensor	Features	Date	Source	Res (m)
Landsat	LT5	1-5, 7, NDVI	04/16/09	USGS	30
Landsat	LT5	1-5, 7, NDVI	06/03/09	USGS	30
Landsat	LT5	1-5, 7, NDVI	09/07/09	USGS	30
Landsat	LT5	1-5, 7, NDVI	09/23/09	USGS	30
Landsat	LT5	1-5, 7, NDVI	01/29/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	03/02/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	03/18/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	04/19/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	07/08/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	08/09/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	09/10/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	09/26/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	10/12/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	12/15/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	05/08/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	05/24/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	06/25/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	07/11/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	08/28/11	USGS	30

Landsat	LC8	NDVI	12/07/13	derived[1]	30
Landsat	LT5	NDVI maximum	2009-2011	derived[1]	30
Landsat	LT5	NDVI minimum	2009-2011	derived[1]	30
Landsat	LT5	NDVI mean	2009-2011	derived[1]	30
Landsat	LT5	NDVI standard deviation	2009-2011	derived[1]	30
Landsat	LC8	NDVI	12/07/13	derived[1]	30
Elevation (DEM)			2010	DNR	30
Slope			2010	derived[2]	30
Topographic Wetness Index			2010	derived[2]	30
2011 Crop Data Layer			2011	NASS	30
Soil organic carbon			various	SSURGO	30
Soil water holding capacity			various	SSURGO	30

[1]*derived from the USGS data*
[2]*derived from the DEM data*

Table II. Inputs used in the pilot study for the southern test site.

Dataset	Sensor	Features	Date	Source	Res(m)
Landsat	LT5	1-5, 7, NDVI	04/28/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	05/14/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	07/01/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	07/17/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	08/18/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	10/05/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	10/21/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	11/06/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	12/08/10	USGS	30
Landsat	LT5	1-5, 7, NDVI	01/09/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	03/14/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	03/30/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	05/01/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	05/17/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	06/02/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	07/04/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	07/20/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	08/05/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	08/21/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	09/06/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	09/22/11	USGS	30
Landsat	LT5	1-5, 7, NDVI	10/08/11	USGS	30

Landsat	LT5	1-5, 7, NDVI	10/24/11	USGS	30
Landsat	LC8	NDVI	02/02/14	derived[1]	30
Landsat	LT5	NDVI maximum	2009-2011	derived[1]	30
Landsat	LT5	NDVI minimum	2009-2011	derived[1]	30
Landsat	LT5	NDVI mean	2009-2011	derived[1]	30
Landsat	LT5	NDVI standard deviation	2009-2011	derived[1]	30
Elevation (DEM)			2010	DNR	30
Slope			2010	derived[2]	30
Topographic Wetness Index			2010	derived[2]	30
2011 Crop Data Layer			2011	NASS	30
Total organic carbon			various	SSURGO	30
Total water holding capacity			various	SSURGO	30

[1]*derived from the USGS data*
[2]*derived from the DEM data*

The satellite data were chosen based on image availability and cloud cover, where all available images over 2009-2011 were acquired, cloud-masked, and processed. Spectral metrics and indices were chosen and calculated based on extensive previous work emphasizing the importance of vegetation indices in distinguishing vegetated land cover. Other work has shown significant accuracy increases for classifications incorporating ancillary spatial datasets over spectral data alone (e.g. Watanchaturaporn, 2008). The five ancillary datasets used in the pilot were particularly chosen with attention to the environmental conditions and characteristics of the target land cover classes. For example, while upland and lowland shrub may appear similar in spectral data, elevation information should logically help separate the two classes. Similarly, soil properties such as total organic carbon and water holding capacity should help delineate tree species. Note that this while this list represents all the features used for the pilot study, feature selection process is ongoing and any new, usable satellite observations that become available will be acquired and integrated.

2.2 Classification algorithm

Simply put, an image classification exercise is to relate input data, either derived from satellites or other sources, to predetermined output categories (or classes). In this pilot, the goal is to decide whether a pixel belongs to a particular category defined as a part of four different classification schemes described below. One way to arrive at these decisions is to use expert opinion in which observed environmental data are related to known ecological, phenological, and environmental characteristics of a category by an expert. However, this approach tends to be more costly and requires involvement of experts. The alternative is to use a classification algorithm, particularly those based on machine learning domain, to help develop these "rules" to assign the appropriate class to a pixel. In this investigation, we used two supervised classification algorithms, Support Vector Machines (SVM) and decision trees (DT).

SVMs are a supervised nonparametric statistical learning technique that is increasingly being used by the remote sensing community (Huang et al., 2002; Mantero et al., 2005; Mountrakis et al 2011). The heart of an SVM training algorithm lies in the concept of a linear hyperplane – an optimal boundary found through an iterative learning procedure that separates the training set into a discrete predefined number of classes while minimizing misclassifications errors (Vapnik, 1979; Zhu and Blumberg, 2002). Several approaches have been developed to improve SVM predictive accuracies using multispectral remote sensing data. These include the soft margin approach and kernel-based learning that lead to SVM optimization, although the kernel functions often result in more expensive parameterization.

Prior research has identified at least three benefits of SVMs that make them particularly suitable for remote sensing applications. First, regardless of the size of the learning sample, not all the available examples are used in the specification of the hyperplane. This allows SVMs to successfully handle small training data sets because only a subset of points – the support vectors – that lie on the margin(s) are used to define the hyperplane. Second, unlike many statistical classifiers, SVMs do not make prior assumptions on the probability distribution of the data, which leads to reduction in classification errors when input data do not conform to a required distribution (e.g. Gaussian). Third, SVM-based classification algorithms have been shown to produce generalizable models from a set of input training data, eliminating the notion of overfitting.

To perform the SVM-based classification, we used the LIBSVM implementation that provides linear, polynomial (cubic) and radial-basis kernels (Chang C-C and Lin J-C, 2011). The LIBSVM software is an open source tool and can be downloaded from http://www.csie.ntu.edu.tw/~cjlin/libsvm/. This implementation includes C-support vector classification (C-SVC), n-support vector classification (n-SVC), distribution estimation (one-class SVM), e-support vector regression (e-SVR), and n-support vector regression (n-SVR) formulations. All SVM formulations supported in LIBSVM are quadratic minimization problems. Using the radial-basis kernel classification option, the LIBSVM required only two parameters to be defined: the kernel parameter c and the cost parameter C. Both of these parameters are data dependent and are identified separately for each footprint/date-pair combination using the grid search option over log-transformed hyper-parameters as suggested by (Hsu C-W et al., 2001). Note that SVMs have been shown to perform well given a certain level of noise (i.e. mislabeled training data), which is also considered as another advantage of SVMs.

Decision trees are becoming widely utilized in large area mapping of land cover where large and complex spectral datasets require robust algorithms. Decision trees are constructed through the recursive partitioning of training data according to a statistical test applied to the training features (here, value test of a spectral feature). After they are built, unlabeled pixels are sorted down the tree according to the decision rules and eventually terminate in a class assignment. Ten trees are estimated using boosting, a technique that improves class discrimination by iteratively training classifiers based on different weightings of the training data. Because a class label is assigned with each iteration of the boosting algorithm, together the ensemble of trees provides an estimate of conditional probabilities for each class at every pixel. The decision tree algorithm utilized for this study is the C4.5 (Quinlan, 1993). The splitting criterion used in C4.5 is information gain, a statistical measure related to entropy which measures the reduction in uncertainty related to assigning a class label after the classes are partitioned according to the attribute test. Because the learning process depends on a heuristic search approach, DT are often considered nonmetric rather than nonparametric although like SVM they do not require assumptions about the distribution of the data. The source code for C4.5 is also open source and can be downloaded from http://www2.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/c4.5/tutorial.html.

Numerous studies have demonstrated the utility of both SVM and DT, and several have also compared the performance of the two for land cover mapping applications (Huang, 2002; Giri, 2012; Schneider, 2012). While map accuracy produced from both SVM and DT is often high relative to other algorithms, SVM consistently produces more accurate results than decision trees. Moreover, SVM has been shown to perform well with large amount of missing or cloud-contaminated data and have less tendency to overfit on the training data (and thus perform better when classifying unseen pixels). These findings have been confirmed on the training data (detailed below) used for this pilot, where initial tests showed a slight but notable (average 3%) higher accuracy for SVM. For these reasons, the Phase II work will likely move forward with the classification using SVM. However, because SVM are processing- and time-intensive, the map products produced for the pilot have been produced with the C4.5 decision tree algorithm.

2.3 Training data and the classification schemes

Since we take a supervised classification approach in the pilot study, the training (learning) data are extremely important. Training data refer to a set of input features or exemplars with known class identities. Depending on the nature of desired classes, they can be generated from ground surveys, image interpretation, or derived from other datasets. In this investigation, the training data is derived and synthesized from a variety of existing sources, including national-and state-level forest and wetland inventories, regional ground surveys, and ground truth data from Wiscland-1 project. While the reference data are spatially distributed throughout the state, many of the data sources emphasize a specific land cover (e.g. forest types) so not all sources provide points in each footprint.

Perhaps more important information about the training data is contained in its class-specific counts. While there is variation in the source of the training data, there is much more variation in class specific training data. With respect to class specific data, we need to define the classification schemes used in the pilot study. To better understand the effect of categorical detail on classification accuracy, we took the draft Wiscland-2 classification scheme (Appendix A) and generated four different sets of class schemes, each with increasing complexity. These complexities are labeled as levels in the following discussion where level 1 refers to the most generalized and level 4 refers to the most specific scheme.

The level 1 classes can be considered generalized land cover categories, separating uplands from lowlands, forests from other vegetation types, and non- or sparsely-vegetated (e.g. barren, urban, water) areas. Level 2 categories add more depth, separating forest types (i.e. deciduous and coniferous), three different types of lowlands (e.g. meadows, shrubs, and forested lowlands), and two grassland categories. Levels 3 and 4 add much more categorical detail to the classification, going down to species level, and are targeting specific stand-based forest categories that have both forestry and deer management importance. Note that these levels correspond to the levels identified in the draft Wiscland-2 classification scheme. In the final product, the goal is to have the classes across the 4 levels 'nested', with areas that are classified in level 1 divided into further detail of species-level in level 4 and minimized shift between categories among the map levels. However, time constraints on creating the maps for this pilot program prevented this forced nesting from being performed on the pilot sites. Individual classifications were run at each level without 'nesting', resulting in some shift in categories between maps as the inputs (training data) shifted slightly with each classification scheme.

Because the training data are extracted from several projects designed to serve diverse needs, synthesis and standardization of the training data was an integral component of the classification process. With the exception of two training data sources (Forest Inventory Analysis (FIA) and Continuous Forest Inventory (CFI)), all reference data were provided in the form of polygons with a range of associated attributes related to the land properties. To render these data more functional for training in the classification step, where each sample is associated with an individual 30m pixel, points were randomly sampled from each polygon. This sampling procedure has the added advantage of increasing the sample size, which is important for rare land cover classes. Points were generated at the same density for each polygon. The relevant attributes from the polygon are maintained in the point file and sources are synthesized within a central database.

Another necessary step in developing the training database was translating each classification scheme used in the source data into the Wiscland-2 land cover scheme. This is done in a procedure called *cross-walking* wherein each land cover class is cross-referenced from a land cover class in one scheme to an analogous class in another scheme. While some cases are simply one-to-one matches between the schemes, others involve refining definitions and applying thresholds to ensure that each sample given a label under the draft Wiscland-2 scheme meets the same criteria. Most notably, several forestry sources provide stand-level information (e.g. FIA, CFI, WisFIRS reconnaissance (Recon)). For these, a purity ratio was calculated and a threshold applied to determine whether the sample fits into the definition of a Wiscland-2 forest class. The threshold is chosen based on definitions provided by the original data sources and range from 65% stand purity (for FIA, CFI) to 75% stand purity (WisFIRS Recon). Any samples not meeting these thresholds were not included in the samples database.

In all other data sources, all samples were included in the database and cross-walked to the most appropriate class or flagged to revisit. There are several cases where the land cover label for a sample was provided, but not at the requisite level of detail to match a level 4 class in the Wiscland-2 scheme (e.g. “Emergent/Wet Meadow” without sub-categorization). In these cases, the most appropriate and detailed class label possible was assigned. At the pilot stage, these samples are included in the database and are used for training, even though in absence of labels for all levels they cannot be used for the higher level classifications. For this reason, the training sample will likely be limited to only those sites that can be categorized down to level 4 during Phase II. In the following sections, any data counts refer to the anticipated training sample, where sites without detailed (level 4) information are removed. However, the classifications discussed in the Results section of this paper are based on classifications utilizing all of the available training data, not just limited to the ‘ideal’ reference points with detailed information.

Relatedly, several data sources include sample points falling in agricultural classes, which are all currently included in the database. Agricultural classes are a special case in this workflow for several reasons. First several, more accurate sources of agricultural land cover information are already available at the state level. The Cropland Data Layer (CDL) produced by the National Agricultural Statistics Service produces annual updates a 30 m spatial dataset containing information on crop type with high accuracy for the target classes (85-97%). Secondly, while the Wiscland-2 land cover map aims to provide a static snapshot of the land, agricultural landscape tend to be more dynamic, rotating over varying intervals. A statewide layer containing crop rotation information derived from the CDL and local knowledge and defined for the Common Land Unit (CLU) is available to further refine the cropland categories. Given the dynamism and availability of current agricultural data, these classes will not be classified directly from satellite data like the other land cover classes. At this point, neither the

CDL nor crop rotation information has been integrated for the pilot map results. All agricultural samples are currently maintained in the database and are included in the classification, while the final product will utilize CDL crop data and will not classify agricultural areas using collected reference data.

Training data: Northern site

Training data for the northern site were extracted from four separate sources, resulting in 149,900 unique samples (Table III).

Table III. The source and the counts of training data for the northern site

Source	**Count**	**Notes**
Recon	135691	
Wiscland-1	11278	Mostly non-forest categories
FIA	1707	
CFI	1224	

Based on this arrangement, the source of training data is very much unbalanced: a large majority of the training data is generated from the Recon forest inventory data. The impact of this distribution is clear in Table IV, which provides information the four classification schemes, along with training data counts for each class in each level. Levels 1 and 2 are heavily weighted towards forestry categories, partially because the majority of the data sources (Recon, FIA, and CFI) are forest-related inventory sources. In Levels 3 and 4, there is more representation in counts between categories, although the most common classes are still forest-related. Several classes have few training sites available, which may be due to a lack of collection sites or that specific land cover being rare on the landscape. A review of the reference data available compared to the landscape composition is discussed in the following section. It is not clear at this time how or if this imbalance in the training samples affect classification or class-specific accuracy, but some preliminary information can be extracted and is discussed below.

Table IV. Description and the counts of four classification levels used in the northern pilot

Level 1	Level 1 Label	Count	Level 2	Level 2 Label	Count	Level 3	Level 3 Label	Count	Level 4	Level 4 Label	Count
1	Urban/Developed	52	1	Urban/Developed	52	1	Urban/Developed	52	1	High Intensity Urban	1
2	Agriculture	2116	2	Agriculture	2116	2	Agriculture	2116	2	Low Intensity Urban	51
3	Grassland	1918	3	Grassland	1225	3	Grassland	1225	3	Corn	1424
4	Forest	120266	4	Hay	693	4	Hay	693	4	All Other Crops	634
5	Open Water	1262	5	Coniferous Forest	20222	5	Jack Pine	3563	5	Cranberries	58
6	Wetland	24206	6	Broad-leaved Deciduous Forest	100044	6	Red Pine	13555	6	Grassland	1225
7	Barren	48	7	Open Water	1262	7	White Pine	1615	7	Hay	693
8	Shrubland	32	8	Emergent/Wet Meadow Wetland	245	8	Fir Spruce	813	8	Jack Pine	3563
			9	Lowland Shrub	661	9	Hemlock	676	9	Red Pine	13555
			10	Forested Wetland	23300	10	Aspen	89529	10	White Pine	1615
			11	Barren	48	11	Paper Birch	975	11	Fir Spruce	813
			12	Shrubland	32	12	Oak	5291	12	Hemlock	676
						13	Red Maple	2559	13	Aspen	89529
						14	Northern Hardwoods	1690	14	Paper Birch	975
						15	Open Water	1262	15	N. Pin Oak, Black Oak	95
						16	Emergent/Wet Meadow Wetland	245	16	Red Oak	5196
						17	Broad-leaved Deciduous Lowland Shrub	528	17	Red Maple	2559
						18	Broad-leaved Evergreen Lowland Shrub	133	18	Sugar Maple	1582
						19	Coniferous Forested Wetland	7	19	White Ash	108
						20	Swamp Conifer Forested Wetland	2134	20	Open Water	1262
						21	White Cedar Forested Wetland	12319	21	Emergent/Wet Meadow Wetland	245
						22	Black Spruce Forested Wetland	4897	22	Broad-leaved Deciduous Lowland Shrub	528
						23	Bottomland Hardwood	30	23	Broad-leaved Evergreen Lowland Shrub	133
						24	Swamp Hardwood	3913	24	Coniferous Forested Wetland	7
						25	Barren	48	25	Swamp Conifer Forested Wetland	2134
						26	Shrubland	32	26	White Cedar Forested Wetland	12319
									27	Black Spruce Forested Wetland	4897
									28	Silver Maple Bottomland Hardwoods	30
									29	Black Ash Swamp Hardwoods	3913
									30	Barren	48
									31	Shrubland	32

Training data: Southern site

Training data for the southern site consisted of 57,483 unique samples from 16 separate data sources (Table V). The source data for these samples are listed in Table V below. A variety of data sources were available for this pilot location. Wiscland-1 and Recon were the most prevalent data sources, at approximately 16% of test sites per source each, but several sources focused on grassland and wetland habitat were also highly utilized. FIA and CFI data figured much less prominently in this site.

Table V. The source and the counts of training data for the southern site

Source	Count	Notes
Wiscland-1	9359	Mostly non-forest categories
Recon	8980	
DNR Grassland Bird land cover map	7786	
FWS Horicon Marsh land cover	5806	
DNR Bird Conservation Areas	5681	
DNR Badger home ranges land cover	3831	
DNR Glacial Habitat Restoration Area reference sites	3121	
DNR Duck Habitat study	2831	
DNR Pheasant Unit land cover map	2580	
DNR Biomass land cover map	1842	
FWS Waterfowl Production Area land cover map	1496	
DNR Seeded Grasslands	1489	
DNR Southwest Grasslands reference sites	1122	
DNR Farming Agreements	1009	Grassland habitat data
FIA	444	
CFI	106	

The class specific counts for the southern site are listed in Table VI. With respect to counts per class, Levels 1 and 2 are weighted towards grassland-type categories at this pilot site. This is not surprising as a large number of the data sources were focused on grassland habitat project areas. Agriculture and forestry sites also had a significant amount of training points available. In Levels 3 and 4, there is more homogeneity in counts between categories, although the samples appear to be weighted toward pasture and agricultural land cover classes. The southern site does not have as large of a bias toward forest classes as was present in the northern site, but that is to be expected with the differences in land use between these two sites, agriculture is much more prevalent in the southern part of the state compared to the north.

Table VI. Description and the counts of four classification levels used in the southern pilot

Level 1	Level 1 Label	Count	Level 2	Level 2 Label	Count	Level 3	Level 3 Label	Count	Level 4	Level 4 Lable	Count
1	Urban/Developed	726	1	Urban/Developed	726	1	Urban/Developed	726	1	High Intensity	92
2	Agriculture	10662	2	Agriculture	10662	2	Agriculture	10662	2	Low Intensity	576
3	Grassland	22990	3	Grassland	5502	3	Grassland	5502	3	Golf Course	58
4	Forest	10850	4	Pasture	9662	4	Pasture	9662	4	Corn	6619
5	Open Water	3620	5	Hay	7826	5	Hay	7826	5	Soybeans	1723
6	Wetland	7984	6	Coniferous	5683	6	Jack Pine	551	6	Winter Wheat	160
7	Barren	167	7	Deciduous	5167	7	Red Pine	3681	7	All other crops	2160
8	Shrubland	484	8	Open Water	3620	8	White Pine	1447	8	Warm-Season Grass	4019
			9	Emergent/Wet Meadow	5788	9	Fir Spruce	4	9	Cool-Season Grass	1483
			10	Lowland Shrub	253	10	Aspen	1200	10	Pasture	9662
			11	Forested Wetland	1943	11	Paper Birch	4	11	Hay	7826
			12	Barren	167	12	Oak	3573	12	Jack Pine	551
			13	Shrubland	484	13	Red Maple	125	13	Red Pine	3681
						14	Northern Hardwoods	189	14	White Pine	1447
						15	Central Hardwoods	76	15	Fir Spruce	4
						16	Open Water	3620	16	Aspen	1200
						17	Emergent/Wet Meadow	5788	17	Paper Birch	4
						18	Broad-leaved Deciduous Lowland Shrub	244	18	N. Pin Oak, Black Oak	965
						19	Broad-leaved Evergreen Lowland Shrub	9	19	Red Oak	2024
						20	Broad-leaved Deciduous Forested Wetland	215	20	White Oak	499
						21	Coniferous Forested Wetland	5	21	Burr Oak	85
						22	Swamp Conifer Forested Wetland	59	22	Red Maple	125
						23	Black Spruce Forested Wetland	1000	23	Sugar Maple	133
						24	Bottomland Hardwood	619	24	White Ash	56
						25	Swamp Hardwoods	45	25	Walnut	76
						26	Barren	167	26	Open Water	3620
						27	Shrubland	484	27	Floating Aquatic Herbaceous Vegetation	5
									28	Reed Canary (lowland and upland)	89
									29	Phragmites	239
									30	Cattails	5455
									31	Broad-leaved Deciduous Lowland Shrub	244
									32	Broad-leaved Evergreen Lowland Shrub	9
									33	Broad-leaved Deciduous Forested Wetland	215
									34	Coniferous Forested Wetland	5
									35	Swamp Conifer Forested Wetland	59
									36	Black Spruce Forested Wetland	1000
									37	Green Ash Bottomland Hardwood	467
									38	Silver Maple Bottomland Hardwood	152
									39	Black Ash Swamp Hardwoods	45
									40	Barren	167
									41	Shrubland	484

2.4 Class accuracy vs. sample size

An important step in finalizing the samples database includes determining whether each class is sufficiently represented. While there is no direct way to test this, there are a few measures than can be used to assess the effect of sample size on classification accuracy. Note that it is clear from the test samples listed so far that certain categories may have unacceptably low sample sizes. To evaluate the sample database and to test the effect of sample size on classification accuracy, several tests were performed. First, we performed an initial test to evaluate the general relationship between class size and accuracy. Secondly, we looked at whether scaling down the entire sample had an effect on accuracy. Thirdly, we compared the class frequencies of the training sample to the landscape composition approximated from existing land cover data.

To better understand the relationship between sample size and class-specific accuracy, we first ran a *cross-validation* exercise. Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. To do this, we took the observations falling within the northern pilot site and randomly split it into *known data* (60%) on which training is run (*training dataset*), and a dataset of *unknown data* (40%) against which the classification model is tested (*testing dataset*) for the northern site (Table VII). To get a complete picture of the expected accuracy, we repeated this procedure 100 times, each time partitioning the original data into different complementary subsets of training and validation sets.

To test the effect of sample size on classification accuracy, we count the number of correct and incorrect results as:

True positive (TP): A class is correctly identified as that class
False positive (FP): Other categories are falsely identified as the category of interest
True negative (TN): A category is correctly identified as not being another category
False negative (FN): The category of interest is not identified correctly but it should have been

Then, we plot the individual class accuracies, measured by different metrics against the samples in the training data used to train the classifier:

Precision (PREC): describes the fraction of the test samples, for each class, correctly classified, that are relevant to the user's information need. Precision is analogous to positive predictive value. It is calculated as TP/(TP + FP) and can be interpreted as the rate of 1 – the omission error.

Recall (RECALL): describes the fraction of the test samples, for each class, that are relevant to the inquiry and are successfully classified. Recall is often called sensitivity and can be interpreted as the probability that correct test samples are retrieved by the classification. Note that it is trivial to achieve recall of 100% by returning all test samples as the correct class. Therefore recall alone is not enough but one needs to measure the number of non-relevant test samples that were classified as the category of interest also, for example by computing the precision. Recall is calculated as TP/(TP + FN) and can be interpreted as the rate of 1 – the commission error.

Although precision and recall can be used to evaluate and algorithms outputs, these metrics often have an inverse relationship in a classification, increases in precision may be correlated with decreases in recall. Therefore it is difficult to set a numeric goal for these metrics to designate a classification as 'good' without taking context and goals of the classification into account. These metrics are useful in developing additional evaluation metrics for a classification, such as the F-measure.

<u>F-Measure (F(1.0)):</u> describes the weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is: (2 * precision * recall)/(precision + recall). This is also known as the F1 measure, because recall and precision are evenly weighted. In other words, the user measures the effectiveness of retrieval with respect to a user who attaches as much importance to recall as precision. An evaluation of f-measure also requires consideration of the context and goals of the classification, but for this pilot a score of greater than 0.5 would be considered good.

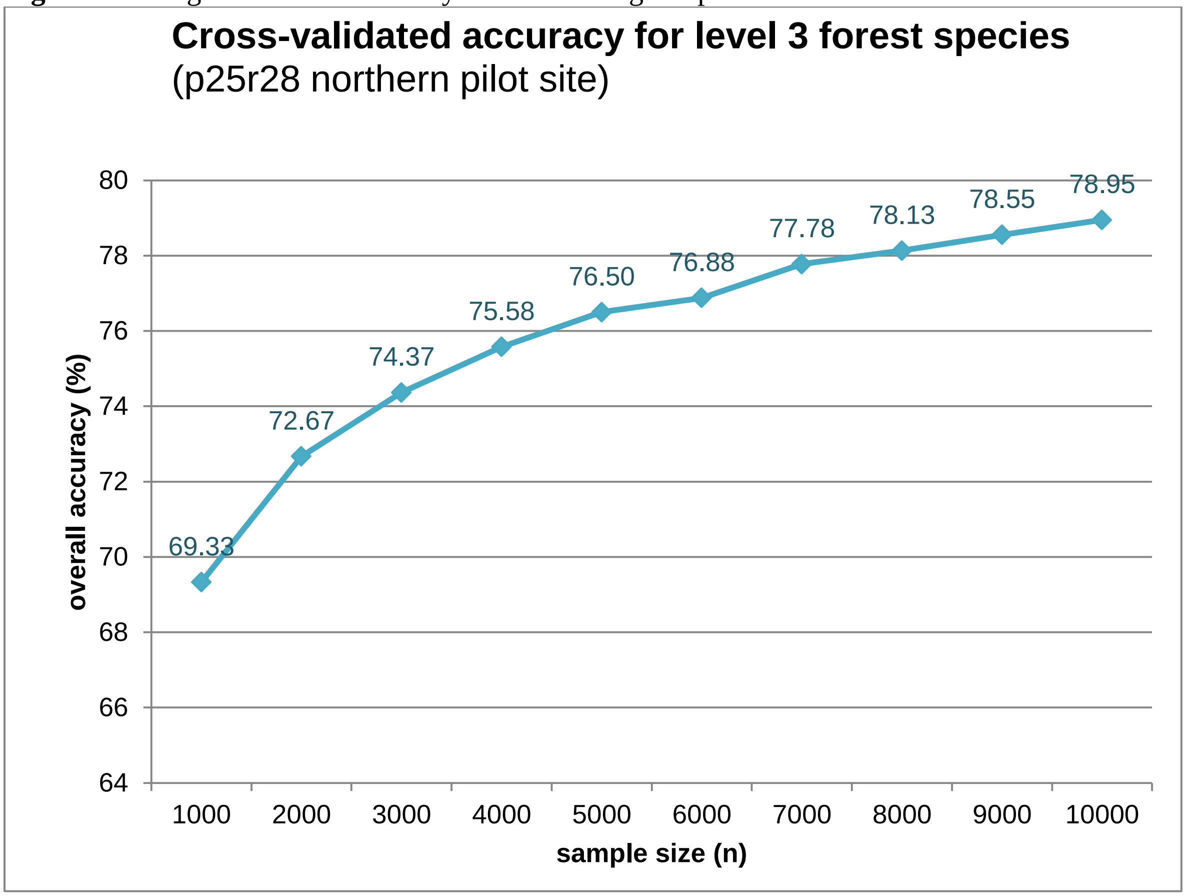
<u>Area Under the ROC curve (AUC):</u> The accuracy of a classification problem depends on how well it separates the class of interest from those that are not of interest. One way to measure the accuracy is to calculate the area under the Receiver Operating Characteristic (ROC) curve. An area of 1 represents a perfect classification; an area of .5 represents an OK classification. The AUC metric simply represents this measure.

Table VII. Accuracy metrics and sample size per class for level 2 classification for northern site.

Class Description	**Sample Size**	**TP**	**FP**	**TN**	**FN**	**PREC**	**RECALL**	**F(1.0)**	**AUC**
Emergent/Wet Meadow	25	5	4	19971	20	0.55556	0.2	0.29412	0.5999
Lowland Shrub	89	15	7	19904	74	0.68182	0.16854	0.27027	0.58409
Forested Wetland	3015	2371	494	16491	644	0.82757	0.7864	0.80646	0.87866
Barren	9	1	0	19991	8	1	0.11111	0.2	0.55556
Shrubland	1	0	0	19999	1	0	0	0	0.5
Low Intensity Urban	1	1	1	19998	0	0.5	1	0.66667	0.99997
Cropland	273	143	57	19670	130	0.715	0.52381	0.60465	0.76046
Grassland	168	89	48	19784	79	0.64964	0.52976	0.58361	0.76367
Hay	104	45	34	19862	59	0.56962	0.43269	0.4918	0.71549
Coniferous Forest	2749	1958	465	16786	791	0.80809	0.71226	0.75715	0.84265
Broad-leaved Deciduous Forest	13390	12683	1400	5210	707	0.90059	0.9472	0.92331	0.8677
Open Water	176	172	7	19817	4	0.96089	0.97727	0.96901	0.98846

Because of the potential accuracy gains from increasing sample size, a 100-crosshold validation exercise was conducted to test whether a revised sampling technique could provide sufficient training data and improve classification results. For efficiency, tests were performed for one classification level for the northern site. Given the importance of the forest categories to the Wiscland-2 project, the test was performed for the 10 forest classes included in the Level 3 classification (Figure 4). The number of training samples iteratively increased by 1000 samples across distributed between all 10 classes. The results reiterate the trend shown in the previous tests, accuracy increases with increasing sample size, but with diminishing returns.

Figure 4. Changes in overall accuracy with increasing sample size.



While these tests indicate that there is a relationship between sample size and accuracy, because land cover is distributed unevenly across space and training data are somewhat limited by cost and accessibility, there is not necessarily potential to gather hundreds or thousands of samples for every land cover class. Moreover, classes with smaller sample sizes may have lower representation in the sample simply because they are rare in the landscape. If so, these classes are expected to have lower class accuracies than more predominant classes. Thus, an important step in finalizing the samples database includes determining whether each class is sufficiently represented.

As indicated above, the training dataset is quite imbalanced, particularly in the northern site where forest classes constitute the majority of the training exemplars and other classes are poorly represented. To evaluate whether the sample class distribution is representative of the landscape, class frequencies from the samples database were compared against the landscape composition as characterized in Wiscland-1. While Wiscland-1 is not expected to exactly reflect the current landscape because of the age of data, it does provide a general indication of the landscape composition appropriate for this exploratory analysis. Wiscland-1 land cover composition was calculated for both the northern and southern sites and the per-class composition was compared to the distribution of reference data points generated. The results of these comparisons are shown in tables VIII and IX.

Overall, the distribution of sample points is quite similar to the Wiscland-1 composition, indicating that the sample is representative of the landscape. Some categories like “Aspen” are oversampled. Oversampling is not a significant problem as the sample can simply be reduced by sub-setting the sample points that better reflect the composition of the landscape. Reducing the sample in these overrepresented classes will also have the effect of boosting the relative proportions of some underrepresented classes. Some urban and agriculture classes appear underrepresented in the current database in each pilot scene. Given the proposed methods of utilizing CDL for agricultural classes and manual collection of additional urban reference data through interpretation of imagery, neither is anticipated to present an issue. First, although agriculture sites are present in both the database and in the classification at this pilot stage, the final product will be classified according to existing cropland data rather than image classification. While urban cover will be classified using the image data, the class representation will be augmented through photo-interpretation of the Landsat data or Google Earth high resolution imagery.

More pertinently, several categories lack sufficient sites in the current sample. For example, Table VIII shows that while only 0.72% of the samples fall into the “Mixed/Other Broadleaf Deciduous Forest” class, it composes approximately 15% of the landscape. Part of this issue simply relates to difficulties in cross-walking between two classification schemes (e.g. the “Mixed/Other Broadleaf Deciduous Forest” is not a category in the Wiscland-2 scheme). However, if some misrepresentation is assumed to be true, there are two potential solutions. First, more data sources may be acquired in the underrepresented classes and integrated into the database. Alternatively, the classification may require the use of all samples in the database rather than only those that have the level of detail required for the level 4 scheme. In this case, the “Mixed/Other Broadleaf Deciduous Forest” category could be boosted from 1,083 up to 16,222 sites if the sites that fit into the category but do not provide species-level detail are included.

Table VIII. Landscape composition vs. sample composition for the northern site

Class Description	Percent Cover (Wiscland-1)	Percent of Samples
High Intensity Urban	0.10	0.00
Low Intensity Urban	0.20	0.03
Golf Course	0.02	0.00
Agriculture[1]		
Primary Row Crops[1]		
Corn	0.27	0.95
Other Row Crops	0.56	0.42
Forage Crops	0.87	0.46
Cranberries	0.06	0.04
Grassland[2]	5.23	0.82
Jack Pine	2.32	2.38
Red Pine	1.91	9.05
White Spruce	0.01	0.54
Mixed Coniferous/Other Forest	1.94	1.53

Aspen	19.41	59.80
Oak	0.10	3.53
Maple	6.47	1.71
Sugar Maple	3.97	1.06
Mixed/Other Broad-leaved Deciduous Forest[3]	15.57	0.72
Mixed Deciduous/Coniferous Forest[4]	6.72	
Open Water	5.13	0.84
Emergent/Wet Meadow[2]	1.78	0.16
Lowland Shrub Wetland	2.17	
Broad-leaved Deciduous Lowland Shrub	6.19	0.35
Broad-leaved Evergreen Lowland Shrub	0.93	0.09
Needle-leaved Lowland Shrub	0.16	0.00
Broad-leaved Deciduous Forested Wetland	2.82	0.00
Coniferous Forested Wetland	5.69	12.67
Mixed Deciduous/Coniferous Forested Wetland	4.89	0.00
Barren	1.32	0.03
Shrubland[2]	1.48	0.02
	100.00	100.00

[1]Indicates a class that is further subdivided. The total is equal to the sum of the children so are not counted in the total.
[2]While these categories are further subdivided in the proposed Wiscland2 class scheme, no more detailed information exists in the current sample db. The most detailed available data were used.
[3]More samples are available for this level 2 category (n=16222), but they do not provide full level 4 detail information so were not counted. This larger class includes broad leaf deciduous (4200), as well as paper birch, white ash, and central hardwoods/walnut.
[4]This class was removed from the Wiscland2 classification scheme.

Table IX. Landscape composition vs. sample composition for the southern site

Class Description	Percent Cover (Wiscland-1)	Percent of Samples
High Intensity Urban	0.98	0.21
Low Intensity Urban	0.85	1.34
Golf Course	0.13	0.13
Agriculture[1]		
Primary Row Crops[1]		
Corn	21.47	15.39
Other Row Crops	8.91	5.02
Forage Crops	21.12	18.20
Cranberries	0.00	0.00
Grassland	12.39	12.79
Jack Pine	0.33	1.28
Red Pine	0.49	8.56
Mixed/Other Coniferous Forest	0.37	3.36

Oak	4.99	8.31
Maple	0.00	0.60
Mixed/Other Broad-leaved Deciduous Forest[2]	12.73	0.32
Mixed Deciduous/Coniferous Forest[3]	1.40	
Open Water	2.51	8.42
Emergent/Wet Meadow	4.69	13.46
Lowland Shrub[1]	0.19	
Broad-leaved Deciduous Lowland Shrub	1.42	0.57
Broad-leaved Evergreen Lowland Shrub	0.02	0.02
Needle-leaved Lowland Shrub	0.02	0.00
Broad-leaved Deciduous Forested Wetland	2.45	0.50
Coniferous Forested Wetland	0.33	0.01
Mixed Deciduous/Coniferous Forested Wetland	0.03	0.00
Barren	1.06	0.39
Shrubland	0.09	1.13
	100.00	100.00

[1]Indicates a class that is further subdivided. The total is equal to the sum of the children so are not counted in the total.
[2]More samples are available for this level 2 category (n=12502), but they do not provide full level 4 detail information so were not counted. This larger class includes broad leaf deciduous (4200), as well as paper birch, white ash, and central hardwoods/walnut.
[3]This class was removed from the Wiscland2 classification scheme.

3. Results for the Northern Footprint

Results for the northern footprint were generated using a subset of training data from the database of samples falling with the Landsat footprint. To estimate overall and per-class accuracy, a 100-fold cross-validation was conducted using 60% of the data for training (n=30,000) and 40% for testing (n=20,000). For this pilot study, each level of classification was run independently of the others. A significant implication of this processing approach is that the classifications are *not* forced to be nested within the more generalized class scheme. For example, it is possible that a pixel is classified as “Burr Oak” at level 4 but as “Emergent/Wet Meadow” at level 2. It is very important to note that in the proposed methodology, the classifications are conducted hierarchically from level 1 to 4 to prevent these inconsistencies. For the pilot study, both time constraints and an interest in identifying potential class confusion rendered it more advantageous to run each classification separately.

3.1 Classification accuracy at each level

In general, there is an inverse relationship between categorical detail and accuracy of a map. As categorical detail increases (e.g. from Level 1 to Level 4), the accuracy of the classification would be expected to decrease. Figure 5 illustrates the results of increasing the classification level on four accuracy statistics evaluated for the southern site. The median value of the 100 trials is reported for each. The overall accuracy, precision, and recall decline at similar rates, with accuracy decreasing by ~5% moving from level 1 to level 2, and from level 2 to level

3. The difference in accuracy between level 3 and level 4 is quite low, dropping only 1.3-1.5% for each metric. Overall accuracy was 92.3% at level 1 and 80.2% for level 4 classifications in the northern site.

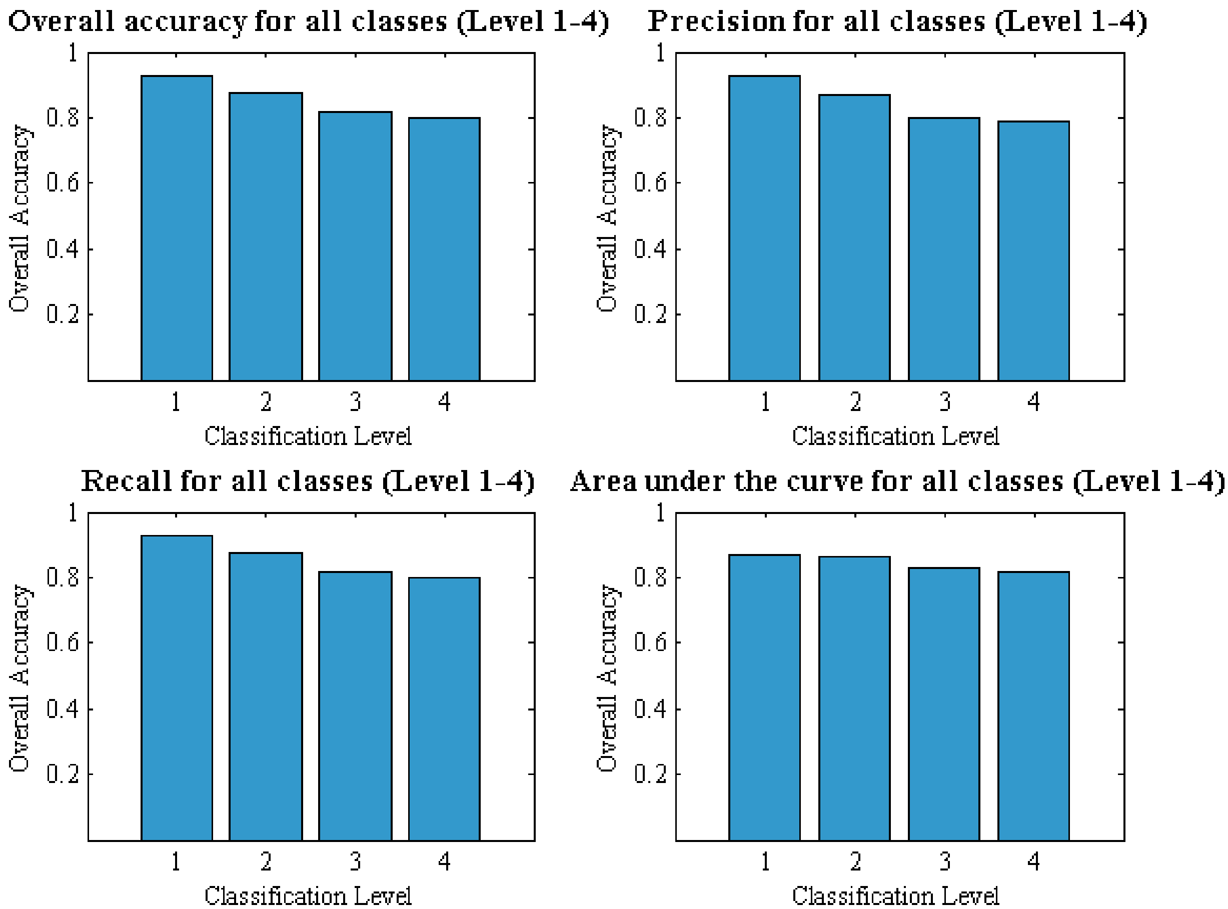


Figure 5. Accuracy assessments for four classification levels on the northern site

Another way to illustrate the relationship between accuracy and class detail is by plotting the median accuracy against the number of categories in the classification scheme (Figure 7). While this plot reiterates the relationship described above, it also reveals that the decline in accuracy occurs more rapidly when the category count is low. While the accuracy declines by the same amount between the first three levels, only six categories are added between level 1 and 2, whereas there is a difference of 15 classes between level 2 and 3. Note that as there are only 4 data points in Figure 6, a statistically significant conclusion is not possible. Nevertheless, these results confirm the expected decline in accuracy as the categorical detail increases.

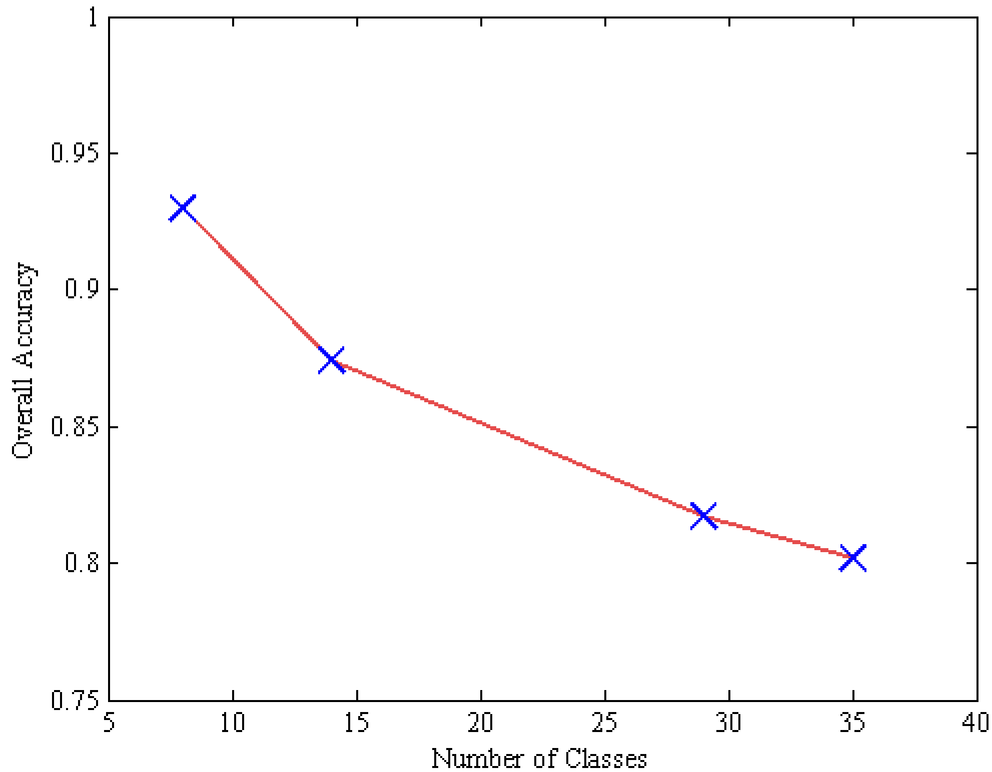


Figure 6. Changes in overall accuracy over number of classes. The overall accuracy is calculated as the mean of 100 rounds of cross-validation procedure. The data points on the X-axis correspond to the class numbers from level 1 (left) to level 4 (right).

While the forgoing discussion informs the expected changes in accuracy as the categorical detail is increased, it does not reveal much information about the expected accuracy of individual categories, which is more relevant in the Wiscland-2 project. To evaluate how well individual classes fair across different levels of classification, we also looked at additional variables such as class-specific accuracies, class-specific ability to recall, and class-specific precision of classification. A confusion matrix, also known as a contingency table or an error matrix, is a specific table layout that allows visualization of the performance of a classification algorithm, typically a supervised learning algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two or more classes. In Table X below we also provide an example of Level 1 classification confusion matrix as the first point of discussion.

Table X. Confusion matrix for Level 1 classification of the northern site.

		Predicted								
		Urban	Ag	Grassland	Forest	Water	Wetland	Barren	Shrub	Sum
Truth	Urban	0	1	0	5	0	0	0	0	6
	Ag	0	134	39	105	6	9	0	0	293
	Grassland	0	18	188	57	0	0	0	0	263
	Forest	0	13	6	15637	3	365	0	0	16024
	Water	0	0	0	5	179	0	0	0	184
	Wetland	0	1	1	757	2	2458	0	0	3219
	Barren	0	0	0	4	0	0	2	0	6
	Shrub	0	0	0	5	0	0	0	0	5
	Sum	0	167	234	16575	190	2832	2	0	20000

In the example confusion matrix (Table X), the diagonal elements for each class indicate the correctly classified samples in the test set while the off-diagonal elements represent the class-specific errors of omission (rows) or commission (columns). Another way of summarizing the class-specific accuracies and errors is to look different error rates, which are summarized in Tables XI to XV below:

Table XI. Per class accuracy statistics for level 1 classification of the northern site

	TP	FP	TN	FN
Urban[1]	0	0	19994	6
Agriculture[1]	134	33	19674	159
Grassland[1]	188	46	19691	75
Forest[1]	15637	938	3038	387
Open Water	179	11	19805	5
Wetland[1]	2458	374	16407	761
Barren	2	0	19994	4
Shrubland	0	0	19995	5

[1]This class includes all training data that falls into any of its subcategories (for example, “Urban” includes “High Intensity”, “Low Intensity”, “Golf Course”, as well as urban that is not otherwise specified).

Table XII. Per class accuracy statistics for level 1 classification of the northern site

	PREC	RECALL	F(1.0)	AUC
Urban[1]	0	0	0	0.5
Agriculture[1]	0.8024	0.45734	0.58261	0.72783
Grassland[1]	0.80342	0.71483	0.75654	0.85625
Forest[1]	0.94341	0.97585	0.95935	0.86997
Open Water	0.94211	0.97283	0.95722	0.98614

Wetland[1]	0.86794	0.76359	0.81243	0.87065
Barren	1	0.33333	0.5	0.66667
Shrubland	0	0	0	0.5
[1]This class includes all training data that falls into any of its subcategories (for example, “Urban” includes “High Intensity”, “Low Intensity”, “Golf Course”, as well as urban that is not otherwise specified).				

Looking back at the Level 1 example confusion matrix and all the derived accuracy measures, the following conclusions can be derived. First, the largest confusion in the Level 1 classification scheme occurs between grassland and agriculture categories and the forest and lowland/wetland categories. Given the ecological and phenological characteristics of these classes, this is not very surprising. Note that since the agricultural category will be imported from the NASS CDL database, the first form of confusion is not very relevant.

The true/positive error rates and the other accuracy measures also reflect these findings. For example, the forest (category 4) and lowland (category 6) categories have higher commission (False positive) rates than omission errors, typically with each other. This reflects the confusing nature of these two classes.

Similar results for the level 2 through 4 categories are given in the following tables:

Table XIII. Per class accuracy statistics for level 2 classification of the northern site

	PREC	**RECALL**	**F(1.0)**	**AUC**
Emergent/Wet Meadow	0.55556	0.2	0.29412	0.5999
Lowland Shrub	0.68182	0.16854	0.27027	0.58409
Forested Wetland	0.82757	0.7864	0.80646	0.87866
Barren	1	0.11111	0.2	0.55556
Shrubland	0	0	0	0.5
Low Intensity Urban	0.5	1	0.66667	0.99997
Cropland[1]	0.715	0.52381	0.60465	0.76046
Grassland	0.64964	0.52976	0.58361	0.76367
Hay	0.56962	0.43269	0.4918	0.71549
Coniferous Forest	0.80809	0.71226	0.75715	0.84265
Broad-leaved Deciduous Forest	0.90059	0.9472	0.92331	0.8677
Open Water	0.96089	0.97727	0.96901	0.98846
[1]This class only includes any training data without any higher level of detail (for example, “Urban” only includes urban sites whose intensity was not otherwise specified).				

The level 2 statistics show high accuracy for the dominant land cover types (“Broad-leaved Deciduous Forest”, “Coniferous Forest”, “Forested Wetland”) while rare classes generally have high rates of omission (“Lowland shrub”, “Emergent/Wet Meadow”). In each case, the rare classes are mostly mistakenly ascribed to one of the dominant classes. For example, 10/25 “Emergent/Wet Meadow” pixels were labeled (“Broad-leaved Deciduous Forest” and 6/25 as “Forested Wetland”. While lowland confusion may need to be addressed through the addition of training data or spectral features, confusion between forested and lowland classes can be largely mitigated by integrating the hierarchical classification method (e.g. only allowing lowland pixels to be classified within level 2 lowland categories).

Table XIV. Per class accuracy statistics for level 3 classification of the northern site

	PREC	**RECALL**	**F(1.0)**	**AUC**
Jack Pine (JP)	0.75943	0.64789	0.69924	0.82133
Red Pine (RP)	0.73371	0.71944	0.72651	0.84681
White Pine (PW)	0.44	0.05446	0.09692	0.52687
Fir Spruce (FS)	0.58824	0.08621	0.15038	0.54293
Hemlock Hardwoods (H)	0.41667	0.09434	0.15385	0.54682
Broad-leaved Deciduous Forest	0	0	0	0.5
Aspen (A)	0.84608	0.93935	0.89028	0.84316
Paper Birch (BW)	0.11111	0.0082	0.01527	0.5039
Oak (O)	0.76285	0.59896	0.67104	0.79576
Red Maple (MR)	0.76	0.27299	0.40169	0.63573
Low Intensity Urban	0	0	0	0.5
Northern Hardwoods (NH)	0.69677	0.47162	0.5625	0.73462
Open Water	0.94054	0.97753	0.95868	0.98849
Emergent/Wet Meadow	0.68421	0.61905	0.65	0.80937
Broad-leaved Deciduous Lowland Shrub	0.7619	0.19512	0.31068	0.59744
Broad-leaved Evergreen Lowland Shrub	1	0.15385	0.26667	0.57692
Coniferous Forested Wetland	0.79604	0.81181	0.80384	0.89088
Bottomland Hardwoods (BH)	1	0.125	0.22222	0.5625
Swamp Hardwoods (SH)	0.725	0.51277	0.60069	0.75385
Barren	0.83333	0.55556	0.66667	0.77775
Shrubland	0.5	0.5	0.5	0.74997
Corn	0.75862	0.43564	0.55346	0.71711
All Other Crops	0.78667	0.71084	0.74684	0.85502
Cranberries	1	0.88889	0.94118	0.94444
Grassland	0.65868	0.75342	0.70288	0.87528
Hay	0.50794	0.4	0.44755	0.69922

Table XV. Per class accuracy statistics for level 4 classification of the northern site

	PREC	**RECALL**	**F(1.0)**	**AUC**
Jack Pine (PJ)	0.74011	0.61215	0.67008	0.80372
Red Pine (PR)	0.72921	0.69598	0.71221	0.8347
White Pine (PW)	0.4386	0.13089	0.20161	0.56464
Fir Spruce (FS)	0.52632	0.09615	0.1626	0.54785
Hemlock	0.5	0.13043	0.2069	0.56492
Broad-leaved Deciduous Forest	0	0	0	0.49997
Aspen (A)	0.83507	0.94705	0.88754	0.83564
Paper Birch (BW)	0.71429	0.03704	0.07042	0.51847
Oak (O)	0.64286	0.28125	0.3913	0.6405
N. Pin Oak, Black Oak	0.33333	0.09091	0.14286	0.5454
Low Intensity Urban	0	0	0	0.5

Red Oak	0.79737	0.60714	0.68938	0.80077
Red Maple (MR)	0.68852	0.26752	0.38532	0.63279
Sugar Maple	0.69811	0.47234	0.56345	0.73496
White Ash	0	0	0	0.5
Open Water	0.92347	0.98907	0.95515	0.99416
Emergent/Wet Meadow	0.72222	0.40625	0.52	0.703
Broad-leaved Deciduous Lowland Shrub	0.61905	0.19697	0.29885	0.59828
Broad-leaved Evergreen Lowland Shrub	0.57143	0.23529	0.33333	0.61757
Swamp Conifer (SC)	0.60889	0.5249	0.56379	0.76022
White Cedar (CW)	0.74744	0.75015	0.7488	0.86363
Black Spruce (SB)	0.65227	0.44543	0.52936	0.71855
Silver Maple	1	0.16667	0.28571	0.58333
Black Ash	0.77863	0.57955	0.6645	0.78754
Barren	0	0	0	0.5
Shrubland	0	0	0	0.5
Corn	0.78704	0.45213	0.57432	0.72548
All Other Crops	0.9	0.71053	0.79412	0.85511
Cranberries	0.7	0.63636	0.66667	0.81811
Grassland	0.63842	0.69325	0.66471	0.84501
Hay	0.62857	0.54321	0.58278	0.77095
Coniferous Forest	0	0	0	0.5

As shown in the level 3 and 4 results, some of the forest categories, which are dominant in this footprint and important for the Wiscland-2 project, have high omission rates (such as the “Oak” classes, “Paper Birch”, and “Spruce”). Again, the analysis of the landscape composition from Section 2.4 indicates that these are rarer classes which are inherently more difficult to detect. More dominant classes and more distinct classes, such as “Aspen” and “Pine” classes, are classified with very high accuracy. A large number of the omissions for each class are miscategorized as aspen. This indicates that the overrepresentation of “Aspen” sites in the training sample may be amplifying the class size in these results and higher accuracies may be achieved if the “Aspen” sample is reduced.

3.2 Map results

The final set of results concern the spatial patterns of classification outcomes. There are four 1.2 km b 1.2 km areas being shown in Figure 7. These refer to different levels of classification, Wiscland-1, and NLCD results.

The maps were produced by mapping each data point to a class label based on the SVM model developing during the training step. As a result, each pixel is associated with one of the class labels for each classification level 1 through 4. To reduce noise in the map, a post-processing procedure was applied to reclassify spurious pixels. Assuming some contiguity of the landscape, an eight-neighbor window was applied to all pixels and any pixel not sharing the same class label with at least four neighbors were filled. The new label was also derived from a clump operator where similar classified areas were clumped together according to a three by three pixel window. In the future image segmentation will be utilized as a post-processing method of eliminating some of this 'speckle' associated with the classification process. Segmentation harnesses the spectral similarity of surrounding pixels to identify objects in the imagery. This is particularly important for land cover, where segmentation can delineate natural boundaries that can then be assigned a uniform land cover when combined with the classification results. At this point, current tests of the segmentation process have not provided a desirable output, either over- or under-simplifying the landscape. Figure 8 gives an example of some of the outputs that were creating during the segmentation testing. However, it is likely that through adjustments to the input data or algorithm selection, image segmentation will provide a valuable method to refine the pixel-based results.

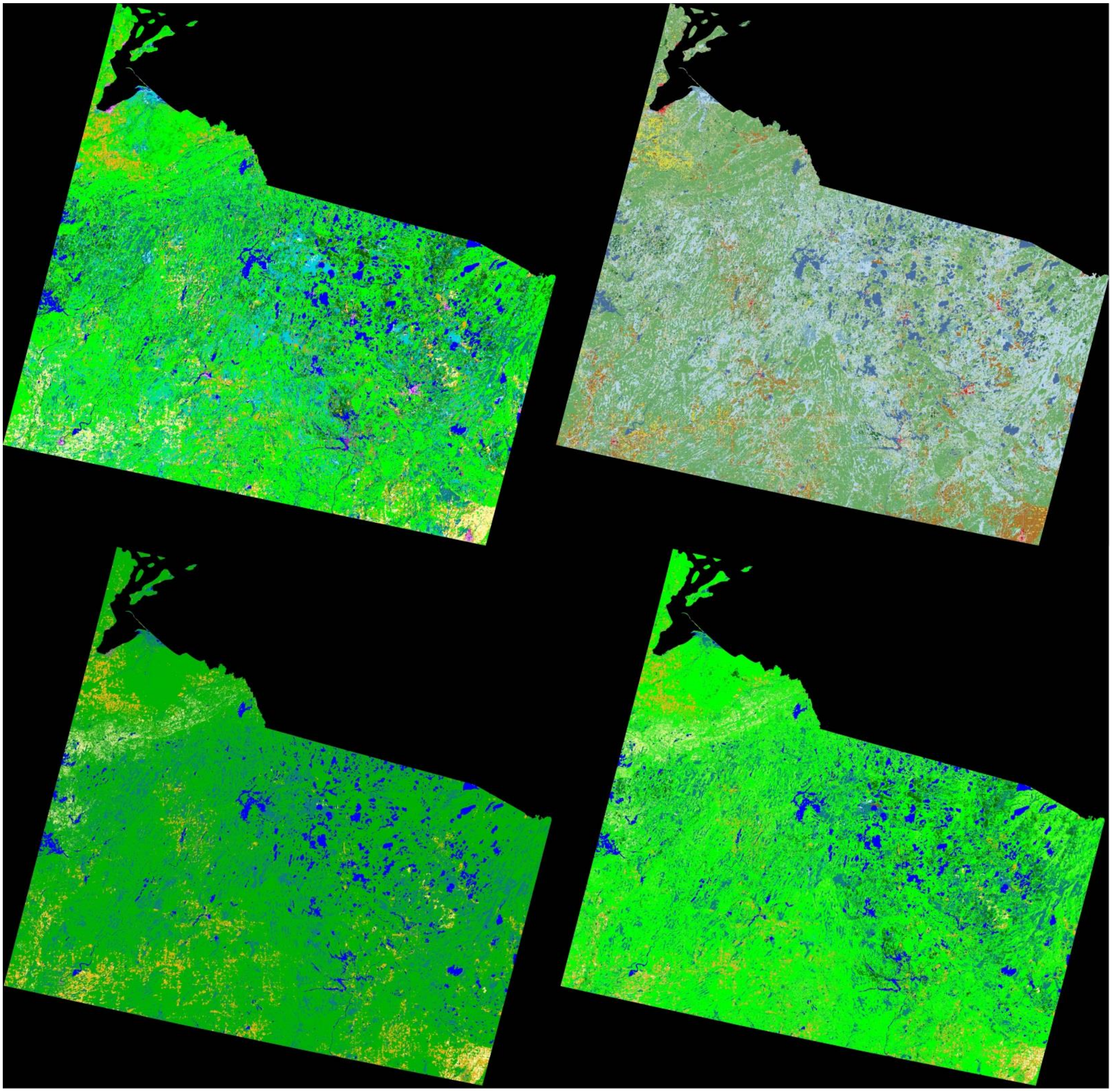


Figure 7. Results of different levels of classification compared to the Wiscland1 and NLCD for the northern site. Wiscland1 (upper left), NLCD (upper right), Level 1 classification (lower left), and level 4 classification (lower left).

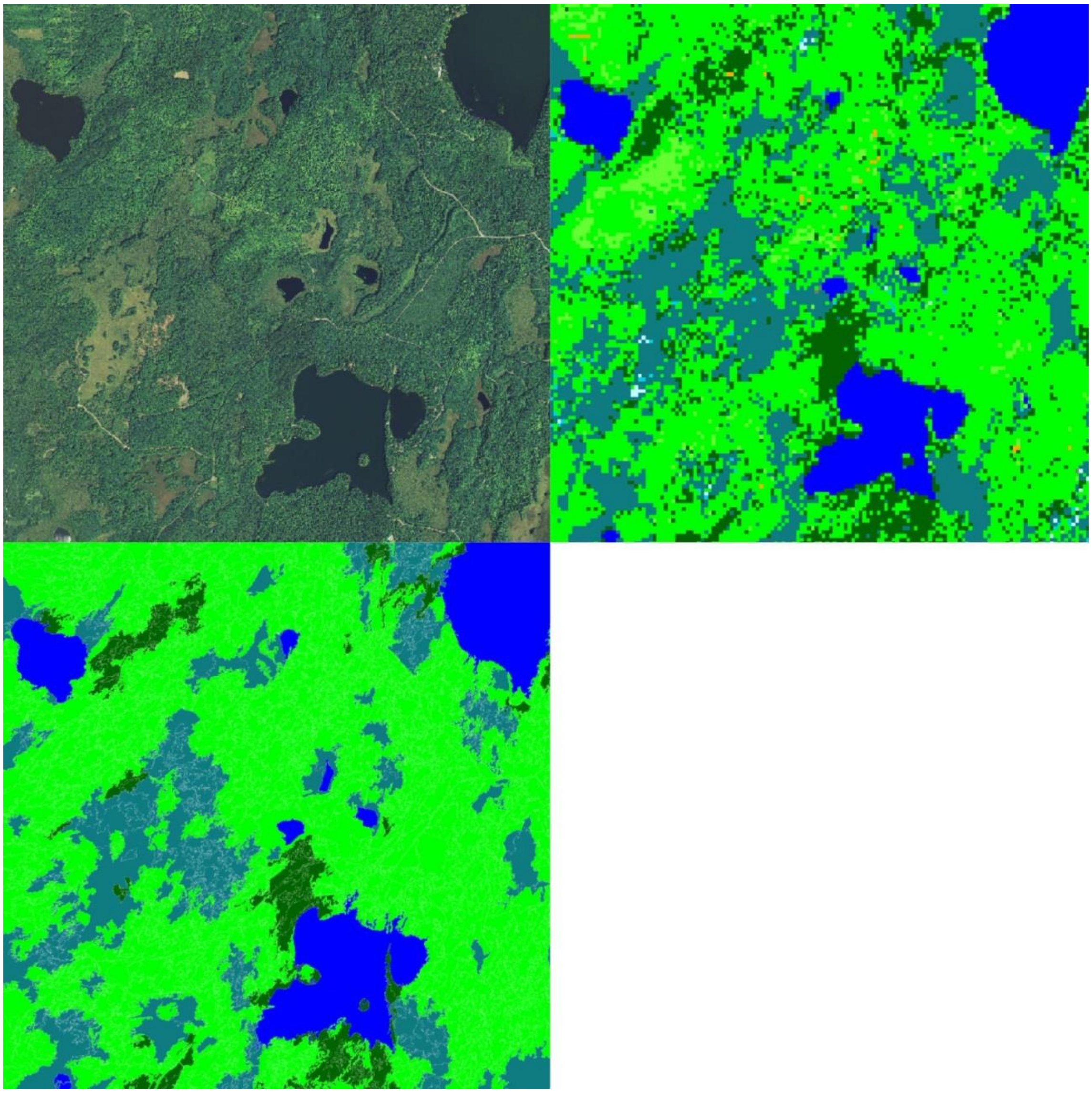


Figure 8. Trial segmentation results for the north site. 1 meter resolution NAIP image (upper left), level 4 classification result (upper right), and segmented post-processed classification result (lower left).

4. Results for the southern footprint

Results for the southern footprint were generated using the full database of samples falling with the Landsat footprint. To estimate overall and per-class accuracy, a cross-validation was conducted using approximately 66% of the data for training (n=80,000) and 33% for testing (n=43341). Similarly to the northern study site, each level 1-4 classification was run independently of the others.

4.1 Classification accuracy at each level

Overall, it is noteworthy that the overall classification accuracy achieved at each level is lower than those of the northern site. Figure 9 illustrates the results of increasing the number of classes on the four accuracy statistics evaluated for the southern site. While AUC remained fairly stable across the various classifications, PREC, RECALL, and overall accuracy declined from 81.6% at level 1 to 65.7% at level 4.

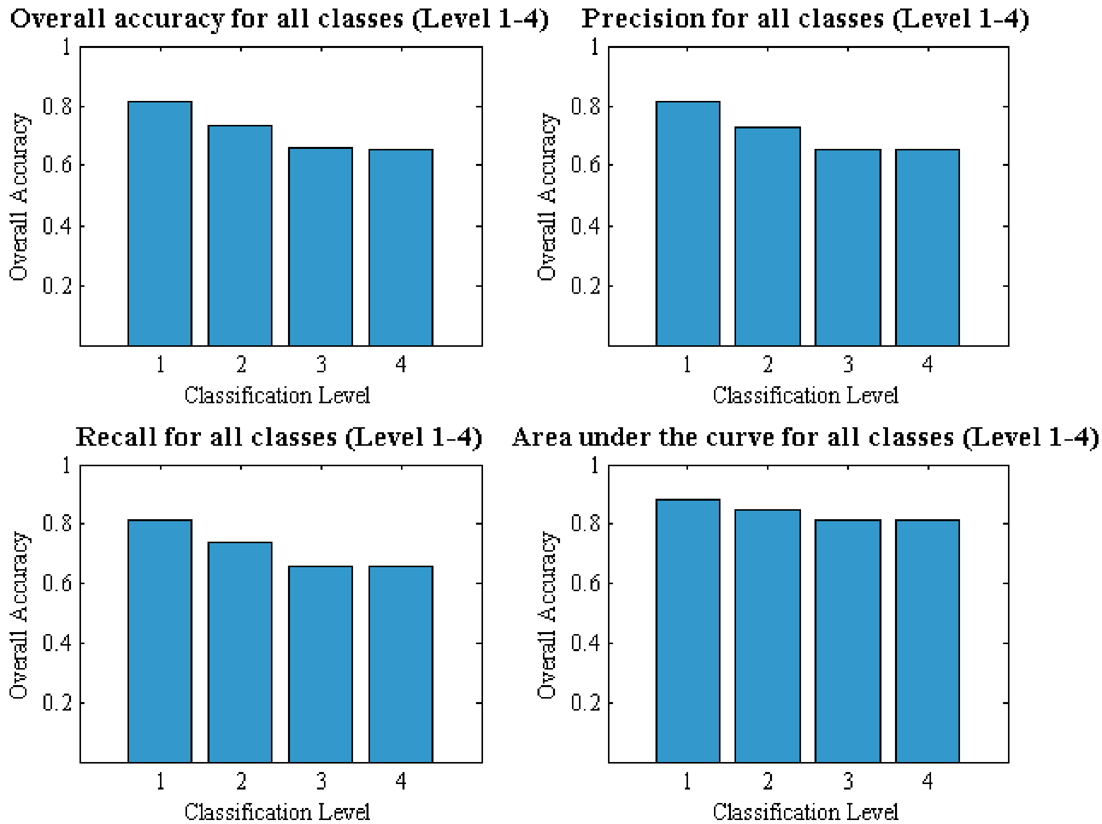


Figure 9. Accuracy assessments for four classification levels on the southern site

Figure 10 provides more insight into some of the potential reasons for the relatively lower accuracy. The range and number of classes present in the southern pilot site are much broader than that of the northern site: the level 4 classification includes 53 categories as opposed to 35. Like the northern site, the changes in the various accuracy statistics between levels 3 and 4 were very minor as the cost of adding more classes is decreased when there are already numerous categories. A more dramatic shift in accuracy is evident when comparing level 2 (19 classes with 73.7% accuracy) vs. level 4 (53 classes with 65.7% accuracy).

Overall accuracy is helpful in getting a general overview of the map output, but it is also important to evaluate the per-class accuracy metrics. Lower than desired accuracy on a class may be associated with a lack of training data for that particular class (as discussed above), high amounts of interclass variability, or a variety of other confounding factors. Per-class accuracy metrics can also help evaluate whether it is feasible to attempt classification on a class. Tables XVI through XX show the per-class results for classification levels 1 through 4 for the southern site. PREC and RECALL for the level 1 classification was lowest on the class with the fewest training sites (barren), and tended to be higher on classes that had either the most training sites available (agriculture, forest, and grassland classes) or classes that have a unique, easy-to-identify signature (“Water”). AUC for all of the classes at level one was over 0.5 (the threshold for an ‘ok’ classification), although “Barren” and “Shrubland” classes were close to 0.5 while “Forest” and “Wetland” were over 0.9.

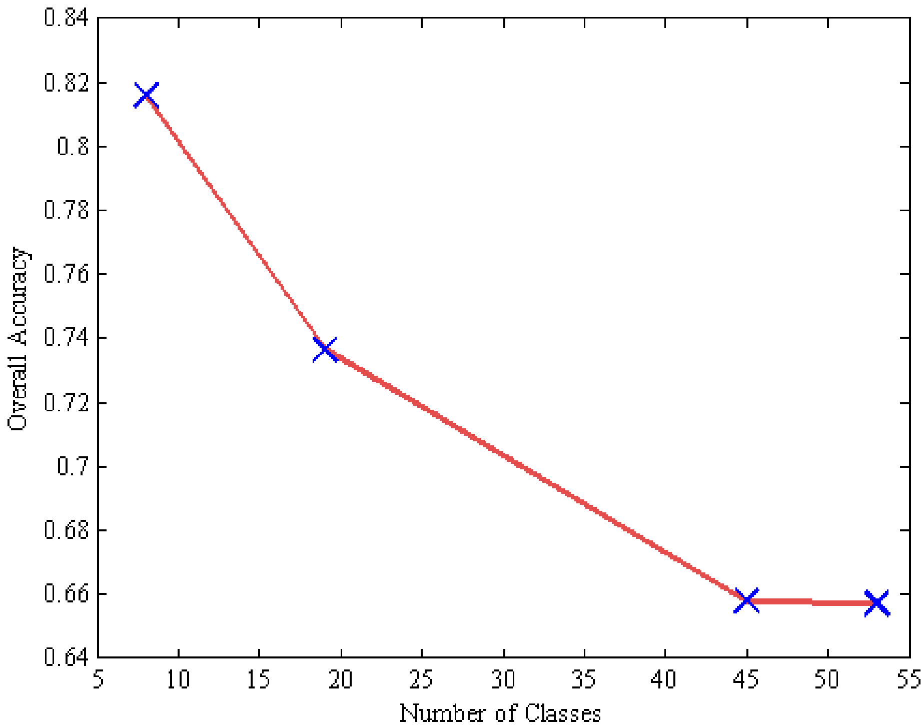


Figure 10. Overall map accuracy compared to number of classes for the southern site.

Evaluating the confusion matrix (Table XVII) can also lead to insight regarding confusion between classes. At level 1, confusion between the “Urban” and “Barren” classes is understandable given that both of these areas have sparse vegetation present. “Shrubland” at level 1 shows confusion between several different classes (grassland, forest, and wetland classes) which may be indicative of a spectral similarities between these various classes, high variability within the “Shrubland” training data itself, or a response to the lower number of available training data for this class, for example.

Table XVI. Per class accuracy statistics for level 1 classification of the southern site

	PREC	**RECALL**	**F(1.0)**	**AUC**
Urban[1]	0.67326	0.59937	0.63417	0.79185
Agriculture[1]	0.8232	0.82569	0.82444	0.88525
Grassland[1]	0.75888	0.78586	0.77213	0.84335
Forest[1]	0.87368	0.89004	0.88179	0.92399
Open Water	0.88475	0.77875	0.82838	0.88787
Wetland[1]	0.8536	0.84447	0.84901	0.90978
Barren	0.625	0.16667	0.26316	0.58326

Shrubland	0.8	0.04598	0.08696	0.52297
[1]This class includes all training data that falls into any of its subcategories (for example, “Urban” includes “High Intensity”, “Low Intensity”, “Golf Course”, as well as urban that is not otherwise specified).				

Table XVII. Confusion matrix for level 1 classification of the southern site.

Truth	Predicted								
	Urban	Ag	Grassland	Forest	Water	Wetland	Barren	Shrub	Sum
Urban	1327	134	491	233	4	20	5	0	2214
Ag	133	8493	1424	151	1	84	0	0	10286
Grassland	316	1444	9681	544	8	325	1	0	12319
Forest	125	158	625	9503	6	259	0	1	10677
Water	1	10	52	19	975	195	0	0	1252
Wetland	30	72	407	373	106	5370	0	1	6359
Barren	32	2	8	4	2	2	10	0	60
Shrub	7	4	69	50	0	36	0	8	174
Sum	1971	10317	12757	10877	1022	6291	16	10	43341

Per-class accuracy at higher levels of classification covers a large range, from 93.1% for winter wheat to 0% PREC and RECALL for several classes at level 4. It is important to note that the classification output shown in these tables utilized all of the training data available, including training sites that did not have a specified level 4 classification (forest areas that did not have a species designated, for example). The impact of including all of the training data vs. only level 4 training data is unclear, but future processing may only include the refined ‘ideal’ training sites. For example, PREC, RECALL, and AUC for the general “Wetland” category in level 4 were relatively low while “Cattail”, a specific wetland category had higher PREC, RECALL, and AUC. It may be the refining the training data to include only areas that are specified at/near level 4 would increase the accuracy of the classification itself.

Table XVIII. Per class accuracy statistics for level 2 classification of the southern site

	PREC	**RECALL**	**F(1.0)**	**AUC**
Urban[1]	0.62714	0.62714	0.62714	0.8046
Forest[1]	0.5783	0.38068	0.45913	0.68197
Coniferous Forest	0.84919	0.82637	0.83762	0.90899
Broad-leaved Deciduous Forest	0.71	0.81595	0.75929	0.88194
Open Water	0.86706	0.81912	0.8424	0.90767
Wetland[1]	0.64773	0.36774	0.46914	0.68315
Emergent/Wet Meadow	0.79305	0.85705	0.82381	0.91573
Lowland Shrub	0.65351	0.27644	0.38853	0.6373
Forested Wetland	0.68073	0.63636	0.6578	0.81406
Barren	0.84211	0.26667	0.40506	0.6333
Shrubland	0.72	0.09836	0.17308	0.5491
High Intensity Urban	0.68966	0.57143	0.625	0.78561
Low Intensity Urban	0.47561	0.19697	0.27857	0.59799
Golf Course	0	0	0	0.49993
Cropland[1]	0.79614	0.86733	0.83021	0.8994
Grassland[1]	0.65025	0.18723	0.29075	0.59278

Idle Grass	0.72378	0.76622	0.7444	0.86212
Pasture	0.65158	0.67308	0.66215	0.80946
Hay	0.88571	0.21934	0.35161	0.60953
[1]This class only includes any training data without any higher level of detail (for example, "Urban" only includes urban sites whose intensity was not otherwise specified).				

Table XIX. Per class accuracy statistics for level 3 classification of the southern site

	PREC	RECALL	F(1.0)	AUC
Urban[1]	0.6105	0.62371	0.61703	0.80206
Grassland[1]	0.58657	0.2368	0.3374	0.61703
Idle Grass[1]	0.61657	0.65153	0.63357	0.80525
Warm-season Grass	0.46681	0.45349	0.46005	0.72097
Cool-season Grass	0.61662	0.42125	0.50054	0.70895
Pasture	0.62594	0.70437	0.66284	0.81955
Hay	0.78947	0.39788	0.5291	0.69847
Forest[1]	0.56314	0.44205	0.4953	0.71052
Coniferous Forest	0.47603	0.38292	0.42443	0.68968
Jack Pine (PJ)	0.78395	0.58796	0.67196	0.79358
Red Pine (PR)	0.77739	0.84119	0.80803	0.91682
High Intensity Urban	0.625	0.45455	0.52632	0.72717
White Pine (PW)	0.7	0.54545	0.61314	0.77147
Broad-leaved Deciduous Forest	0.66974	0.82751	0.74031	0.89377
Aspen (A)	0.65257	0.53465	0.58776	0.76599
Paper Birch (BW)	0	0	0	0.5
Oak (O)	0.69186	0.71687	0.70414	0.85339
Red Maple (MR)	0.64286	0.20455	0.31034	0.60221
Northern Hardwoods (NH)	0.75676	0.41176	0.53333	0.70578
Central Hardwoods	0.66667	0.18182	0.28571	0.59089
Open Water	0.86187	0.80339	0.8316	0.89971
Low Intensity Urban	0.47475	0.2156	0.29653	0.6072
Wetland[1]	0.62814	0.36982	0.46555	0.68405
Emergent/Wet Meadow	0.59142	0.66526	0.62617	0.81938
Floating Aquatic Herbaceous Vegetation	0	0	0	0.49999
Reed canary (lowland and upland)	0.5625	0.28125	0.375	0.64054
Phragmites	0.71642	0.57831	0.64	0.78894
Cattails	0.82191	0.86931	0.84495	0.93036
Lowland Shrub	0.5259	0.32754	0.40367	0.66239
Broad-leaved Deciduous Lowland Shrub	0.73684	0.56	0.63636	0.77983
Broad-leaved Evergreen Lowland Shrub	0	0	0	0.5
Forested Wetland	0.48601	0.43311	0.45803	0.7142
Golf Course	0.44444	0.22222	0.2963	0.61105
Broad-leaved Deciduous Forested	0.44681	0.24138	0.31343	0.62039

Wetland				
Coniferous Forested Wetland	0.85067	0.85067	0.85067	0.92468
Bottomland Hardwoods	0.76023	0.63107	0.68966	0.81506
Swamp Hardwoods	0.72727	0.57143	0.64	0.78568
Barren	0.70833	0.26984	0.3908	0.63484
Shrubland	0.50909	0.1573	0.24034	0.57834
Cropland[1]	0.66167	0.77493	0.71383	0.8522
Corn	0.72216	0.59398	0.65183	0.79061
Soybeans	0.7113	0.28053	0.40237	0.63946
Winter Wheat	0.8125	0.41935	0.55319	0.70961
All other crops	0.80159	0.56583	0.66338	0.78174

[1]This class only includes any training data without any higher level of detail (for example, "Urban" only includes urban sites whose intensity was not otherwise specified).

Table XX. Per class accuracy statistics for level 4 classification of the southern site

	PREC	RECALL	F(1.0)	AUC
Urban[1]	0.61136	0.61257	0.61196	0.79676
Grassland[1]	0.61255	0.2338	0.33843	0.61567
Idle Grass[1]	0.61215	0.65225	0.63157	0.80565
Warm-season Grass	0.46776	0.45435	0.46096	0.72143
Cool-season Grass	0.6015	0.45113	0.51557	0.72371
Pasture	0.63859	0.7112	0.67294	0.82431
Hay	0.79082	0.38272	0.51581	0.69088
Forest[1]	0.53899	0.47947	0.5075	0.72753
Coniferous	0.49798	0.31538	0.38619	0.65625
Jack Pine (PJ)	0.62673	0.66341	0.64455	0.83077
Red Pine (PR)	0.76354	0.85168	0.8052	0.92174
High Intensity Urban	0.69231	0.34615	0.46154	0.67303
White Pine (PW)	0.7106	0.496	0.58422	0.74682
Fir Spruce (FS)	0	0	0	0.49998
Broad-leaved Deciduous Forest	0.68174	0.81913	0.74415	0.89055
Aspen (A)	0.65416	0.57277	0.61076	0.78488
Paper Birch (BW)	0	0	0	0.5
Oak (O)	0.75309	0.51261	0.61	0.75607
N. Pin Oak, Black Oak	0.72107	0.64456	0.68067	0.82119
Red Oak	0.6085	0.58616	0.59712	0.78995
White Oak	0.59756	0.26064	0.36296	0.62994
Burr Oak	0.75	0.10345	0.18182	0.55171
Low Intensity Urban	0.36522	0.23204	0.28378	0.61518
Red Maple (MR)	0.54545	0.31579	0.4	0.65778
Sugar Maple	0.78571	0.45833	0.57895	0.7291
White Ash	0.42857	0.15789	0.23077	0.5789
Walnut	0.33333	0.04	0.07143	0.51998
Open Water	0.84137	0.82518	0.8332	0.91029
Wetland[1]	0.53769	0.34628	0.42126	0.67207

Emergent/Wet Meadow	0.60604	0.6887	0.64473	0.83129
Floating Aquatic Herbaceous Vegetation	0	0	0	0.5
Reed canary (lowland and upland)	0.8	0.21053	0.33333	0.60524
Phragmites	0.7377	0.56962	0.64286	0.78463
Golf Course	0.5	0.38462	0.43478	0.69225
Cattails	0.84623	0.86643	0.85621	0.92947
Lowland shrub	0.53333	0.32458	0.40356	0.6609
Broad-leaved Deciduous Lowland Shrub	0.80303	0.65432	0.72109	0.82701
Broad-leaved Evergreen Lowland Shrub	1	0.25	0.4	0.625
Forested Wetland	0.53488	0.44421	0.48535	0.72
Broad-leaved Deciduous Forested Wetland	0.48649	0.22222	0.30508	0.61089
Coniferous Forested Wetland	0	0	0	0.5
Swamp Conifer Forested Wetland	0.4	0.16667	0.23529	0.5833
Black Spruce (SB) Forested Wetland	0.8509	0.90685	0.87798	0.95275
Green Ash Bottomland Hardwood	0.70779	0.68987	0.69872	0.84442
Cropland[1]	0.65603	0.7839	0.71429	0.8562
Silver Maple Bottomland Hardwood	0.86538	0.69231	0.76923	0.84607
Black Ash Swamp Hardwoods (SH)	0.7	0.53846	0.6087	0.7692
Barren	0.66667	0.27586	0.39024	0.63784
Shrubland	0.48148	0.14286	0.22034	0.5711
Corn	0.7086	0.59465	0.64665	0.79065
Soybeans	0.65759	0.28027	0.39302	0.6391
Winter Wheat	0.93103	0.41538	0.57447	0.70767
All Other Crops	0.84086	0.5138	0.63785	0.75603

[1]This class only includes any training data without any higher level of detail (for example, “Urban” only includes urban sites whose intensity was not otherwise specified).

4.2 Map results

Maps were produced in the same way as described above. An example of the spatial distribution of the various classification levels for the southern site is shown in Figure 11.

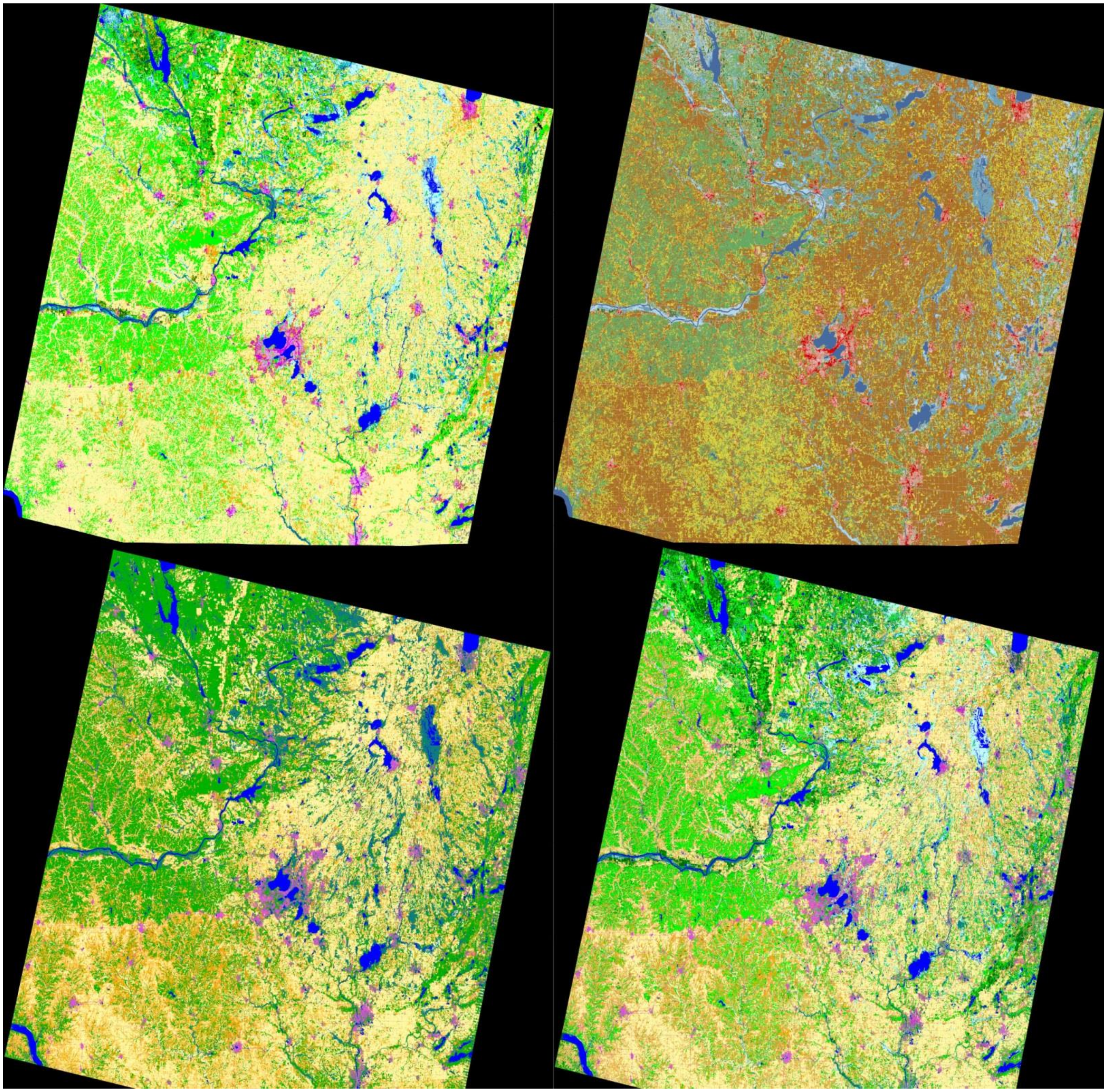


Figure 11. Results of different levels of classification compared to the Wiscland1 and NLCD for the southern site. Wiscland1 (upper left), NLCD (upper right), Level 1 classification (lower left), and level 4 classification (lower left).

5. Overall conclusions

The overall map accuracy achieved for this pilot ranged from 92.9 to 80.2% for the northern site and 81.6 to 65.7% for the southern site for the level 1 through 4 classification schemes. As would be expected, as the number of classes increased (through a more complex classification scheme) the resulting map accuracy decreased. Some of this decrease may be mitigated through refinement of the reference data, collection of additional reference points to supplement low-sample size classes, and forced 'nesting' of the higher level classes within the level 1 class areas. The higher accuracy achieved in the northern site can largely be attributed to: 1) the fewer number of land cover categories present, and 2) the large amount of tree species data contained in the database from sources such as FIA, CFI, and recon.

Both pilot sites, however, had classes that were rare and/or lacked a sufficient number of reference points. In particular, several tree species were poorly classified in the north:

1) Hemlock
2) Spruce
3) Paper birch
4) Pin/Black oak
5) Silver maple

However, given that these species also only constitute a small percentage of the landscape, and the accuracies of the northern site are quite high at the species level given the current inputs and training data, additional efforts at data acquisition may be more efficiently targeted toward the dominant species in the southern portion of the state. Namely, several wetland categories were not well captured, including:

1) Swamp Conifer
2) Broad-leaf deciduous forested wetland
3) Broad-leaf evergreen lowland shrub
4) Reed Canary

In addition, while "Grassland" was reasonably classified in the level 1 classification, distinguishing cool- from warm-season grasses is problematic. Here, the issue does not necessarily appear to be related to the training sample however as the sample proportion mirrors that of the landscape (~12% in each) and the count is sufficient (>1000 for each). Because the confusion is predominantly with agricultural lands, the integration of cropland information from CDL should increase accuracy, as will the implementation of a more hierarchical classification procedure.

Relatedly, classes that are structurally and/or spectrally similar are often committed to the more common land cover type, resulting in lower accuracy for those rare and variable classes. One way to mitigate this issue is to cut back the samples in classes that are over-classified in the results, such as "Aspen" (north) and "Agriculture" (south). A finalized classification scheme will also aid in developing a more certain evaluation of the per-class reference points available and a final determination as to whether/where additional reference data collection is needed.

The use of existing datasets for reference data collection allows for minimal field collection, but considerations related to data standardization and spatial coverage will continue to be reviewed. Perhaps the most significant concern is the level of detail provided by each source, as many do not have enough information to assign the sample to a level 4 category. For the pilot study, all data were included in the training, whether the sample had up to a level 4 category or not. This was advantageous for many classes where the sample size would otherwise be low. However, because of the considerable variability between classification schemes of the training data sources and Wiscland-2, utilizing these sample points may introduce a potentially significant amount of noise into the training data. Going forward, it appears to be advantageous to eliminate these sites from the classification procedure and obtain additional samples with the requisite level of detail to ensure a high-quality dataset that conforms to the class definitions developed for Wiscland-2.

Due to time and processing constraints, several processes and methods that are intended for use in the final product were not utilized for this pilot study. The differences between the proposed methodology and pilot methodology are as follows:

1) **Training database.** Only samples able to be labeled with a level 4 class will be used, rather than the full available set. The database will also be augmented where necessary.
2) **Feature selection.** A formal feature selection process will be conducted to assess the relative importance of each input feature and additional spectral and/or ancillary data will be added if accuracy gains are achieved.
3) **Classification algorithm.** The classifications will be performed using the Support Vector Machine (SVM) algorithm rather than C4.5 decision tree algorithm.
4) **Agricultural classes.** The NASS CDL and crop rotations data will be used to identify agricultural cropland rather than the image classification.
5) **Hierarchical classification scheme.** The classification will be run hierarchically, rather than separately, so that each pixel will fall into the hierarchical classification scheme (level 1-4). With each subsequent classification, only subcategories will be classified so the labels are fully nested.
6) **Image segmentation.** An image segmentation algorithm will be used to serving multiple purposes: a) to improve map accuracy by identifying natural features more accurately than spectral data alone, b) to smooth out spurious pixels for a more refined cartographic produce, and c) to implement a minimum mapping unit of 2 acres.
7) **Dataset delivery.** Map data will be delivered in a format similar to that of Wiscland-1, with the information relevant to each level of classification contained within a single file, rather than four separate files.

Overall, the viability of using existing datasets to create an updated land cover map for Wisconsin is quite feasible. Several concerns have been identified regarding the spatial distribution of the data sets, the number of samples per desired class, and the level of detail available across the various data sets need to be resolved in the future refinements of the process. However, the overall accuracies and many of the per-class accuracy outputs achieved with the current processes are reasonably high, especially given the stage of the project. The methodology development is an iterative process and the methodology for the pilot deviated from the proposed methodology for Phase II in the ways outlined above. The additional of CDL data and image segmentation, nesting of the classes, and the use of the SVM classification algorithm are all anticipated to have significantly positive impact on the final product.

References

Chang, C.-C., and C.-J. Lin. 2011. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2:27:1-27:27.

Giri, C.P. 2012. Remote sensing of land use and land cover: principles and applications. Taylor and Francis Group: Boca Raton, FL.

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 23:725-749

Hsu, C.-W., Chang, C-.C, Lin, C-.J. 2001. A practical guide to support vector classification.

Mantero, P., G. Moser, and S. Serpico. 2005. Partially Supervised Classification of Remotes Sensing Images through SVM-based Probability Density Estimation. IEEE Transactions on Geoscience and Remote Sensing, 43(3): 559-570.

Mountrackis, G., J. Im, and C. Ogole. 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 66:247-259.

Quinlan, J. R. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

Vapnik, V. 1979. Estimation of Dependences Based on Empirical Data [in Russian]. Nauka, Moscow, 5165-5184, 27 (English translation: Springer Verlag, New York, 1982).

Watanachaturaporn, P., M.K. Arora, and P.K. Varshney. 2008. Multisource classification using support vector machines: an empirical comparison with decision tree and neural network classifiers. Photogrammatic Engineering and Remote Sensing, 74(2): 239-246.

Zhu, G., and D.G. Blumberg. 2002. Classification Using ASTER Data and SVM Algorithms; the Case Study of Beer Sheva, Israel, Remote Sensing of Environment, 80(2): 233-240

Appendix A

Draft Wiscland-2 classification scheme

Wiscland	Proposed #	Description
100	1	**URBAN/DEVELOPED**
101	1.1	**High Intensity**
104	1.2	**Low Intensity**
105	1.3	**Golf course**
110	2	**AGRICULTURE**
	2.1	**Cropland**
	2.1.1	Corn
	2.1.2	Soybeans
	2.1.3	Alfalfa
	2.1.4	Winter Wheat
	2.1.5	Potatoes
	2.1.6	All other Crops
	2.1.7	Cranberries
150	3	**GRASSLAND**
PROPOSED	3.1	**Idle grass (unmanaged)**
PROPOSED	3.1.1	warm
PROPOSED	3.1.2	cool
PROPOSED	3.x	**Pasture (managed, forage)**
PROPOSED	3.x	**Hay (managed, forage)**
160	4	**FOREST**
161	4.1	**Coniferous**
162	4.1.1	Jack Pine (PJ)
163	4.1.2	Red Pine (PR)
PROPOSED	4.1.3	White Pine (PW)
166	4.1.4	Fir Spruce (FS)
PROPOSED	4.1.6	Hemlock Hardwoods (H)
PROPOSED	4.1.6.1	*Hemlock*
175	4.2	**Broad-leaved Deciduous**
176	4.2.1	Aspen (A)
PROPOSED	4.2.2	Paper Birch (BW)
177	4.2.3	Oak (O)
COMBO	4.2.3.1	*N. Pin Oak, Black Oak*
180	4.2.3.2	*Red Oak*
PROPOSED	4.2.3.3	*White Oak*
PROPOSED	4.2.3.4	*Burr Oak*
PROPOSED	4.2.4	Red Maple (MR)
PROPOSED	4.2.5	Northern Hardwoods (NH)

185	4.2.5.1	*Sugar Maple*
PROPOSED	4.2.5.2	*White Ash*
PROPOSED	4.2.6	Central Hardwoods
PROPOSED	4.2.6.1	*Walnut*
200	5	**OPEN WATER**
210	6	**WETLAND**
211	6.1	**Emergent/Wet Meadow**
212	6.1.1	Floating Aquatic Herbaceous Vegetation
PROPOSED	6.1.2	Reed canary (lowland & upland)
PROPOSED	6.1.3	Phragmites
PROPOSED	6.1.4	Cattails
PROPOSED	6.1.5	Buckthorn/honeysuckle (lowland)
217	6.2	**Lowland Shrub**
218	6.2.1	Broad-leaved Deciduous
219	6.2.2	Broad-leaved Evergreen
220	6.2.3	Needle-leaved
222	6.3	**Forested**
223	6.3.1	Broad-leaved Deciduous
229	6.3.2	Coniferous
PROPOSED	?	Swamp Conifer (SC)
PROPOSED	?	White Cedar ©
PROPOSED	?	Black Spruce (SB)
PROPOSED	?	Tamarack (T)
234	6.3.4	Mixed Deciduous/Coniferous
PROPOSED	6.3.5	Bottomland Hardwoods (BH)
PROPOSED	6.3.5.1	*Green Ash*
PROPOSED	6.3.5.2	*Silver Maple*
PROPOSED	6.3.6	Swamp Hardwoods (SH)
PROPOSED	6.3.6.1	*Black Ash*
240	7	**BARREN**
250	8	**SHRUBLAND**
PROPOSED	8.1	Buckthorn/honeysuckle (upland)