Boreal Fen probability - technical documentation

ABMI Geospatial Centre

February, 2018

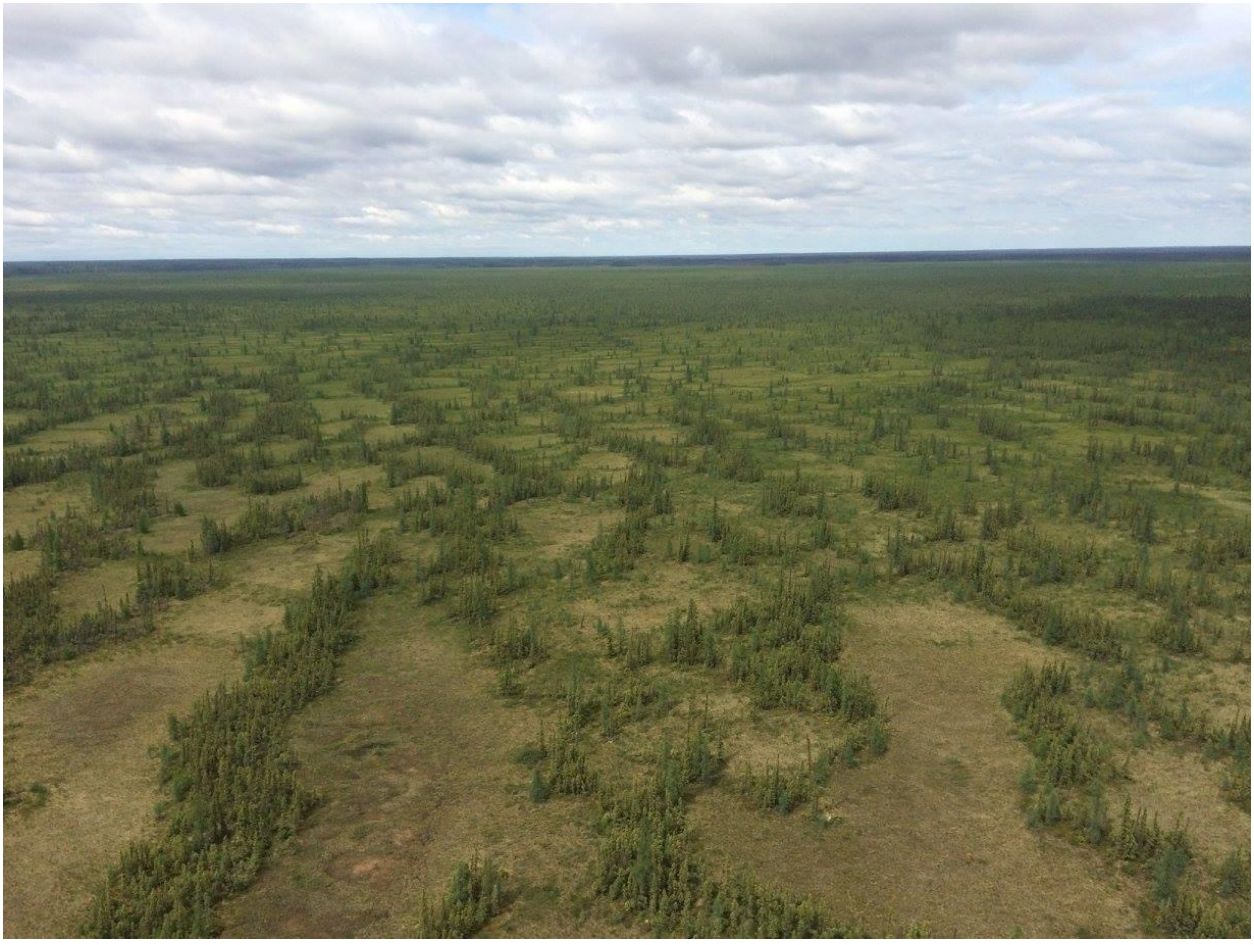


Photo credit – Emily Upham-Mills and Bayne Lab

Contents

1 Introduction 3
2 Methods 4
2.1 Study area 4
2.2 Data 4
2.3 Data exploration and variable selection 7
2.4 Wetland classification – machine learning algorithm 10
2.5 Cross validation accuracy assessment 10
2.6 Additional processing 11
3 Results 11
4 References 16

1 Introduction

Fen is one of the five wetland classes under Alberta Wetland Classification System (e.g., bog, fen, marsh, swamp, shallow open water). Spatial delineation of fen is a difficult task. As seen in the Biodiversity Management Framework, Lower Athabasca Region indicator "Fen Cover", spatial fen coverage differs greatly depending on which data sources are being used to calculate the indicator. Different photo-interpreted wetland inventories show vastly different results (Figure 1). Remotely sensed fen classification, such as the Alberta Merged Wetland Inventory (AMWI; Alberta Environment and Parks, 2017), proves to be even more difficult. Many studies have noted the difficulty in distinguishing fen with spectral data due to spectral mixing between adjacent classes (fen to bog) and spectral signature overlap (Belward *et al.,* 1990; Russel *et al.,* 1997; Ozesmi and Bauer, 2002). A study by Difebo *et al.,* (2015) in the Hudsons Bay lowlands highlighted the difficulty of distinguishing treed fen. The results showed that poor, non-treed fen resulted in 80%–90% accuracy, while conifer fen showed classification accuracies of 45%. This study also noted that fusion of optical remote sensing and a LiDAR-based topographic metrics increased overall wetland accuracy by about 10%. Another study, in Kananaskis, using Landsat, showed very low fen classification accuracy of 28.5% (Williams, 1992). In general, the classification of fen from solely optical remote sensing data is problematic. The optical signature for treed bogs, treed fens, and treed swaps are almost identical and therefore identifying these classes is challenging. More information is needed about subsurface hydrology to really distinguish fen from other wetland classes. The best remote sensing data for estimating hydrology is a high quality LiDAR-based Digital Elevation Model (DEM). The use of wetness indices and relative elevation derived from a DEM has been seen to be key for accurate mapping of wetland location and type (Hird *et al.,* 2017).

The best approach to classifying fen may be to use a fusion approach of data sources (Optical, SAR, and LiDAR based DEM) as suggested in Difebo *et al.* (2015). We intend to test this fusion approach by attempting to classify the probability of fen in wetland areas.

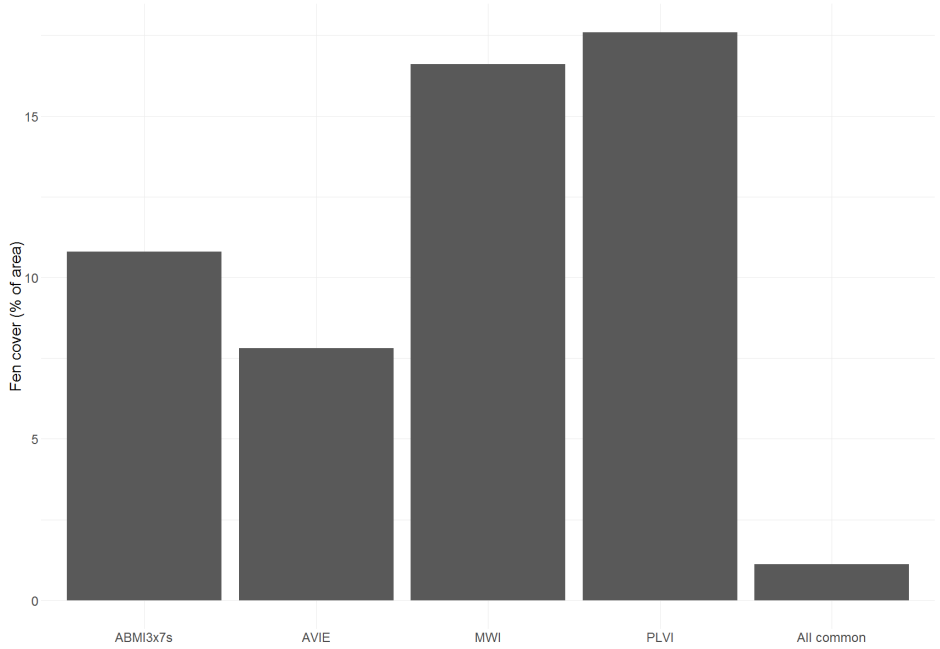


Figure 1: Fen cover calculated as percent of total area for four data sets. The All Common bar describes the percent of area which was commonly fen across the four data sets. AVIE and PLVI are Government of Alberta photo interpreted vegetation inventories, ABMI3x7s is a detailed photo interpreted land cover inventory, and MWI is a remotely sensed classification of wetland classes.

2 Methods

2.1 Study area

The study area primarily consists of the Boreal Natural Region of Alberta with small parts of the foothills, parkland, and Canadian Shield included. This study area makes up about 60% (397, 958 km^2) of the total area of Alberta.

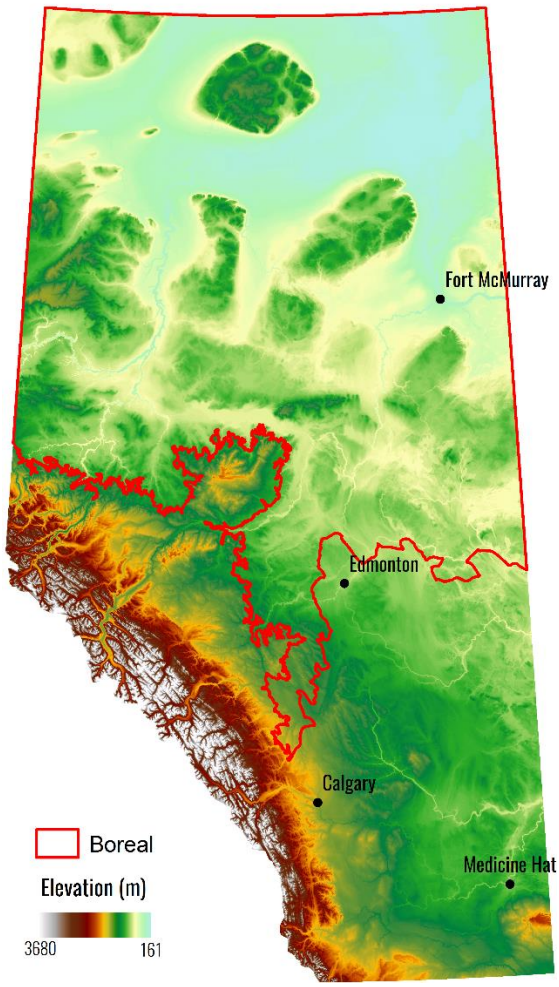


Figure 1: The spatial delineation of the boreal region.

2.2 Data

Sentinel-1, -2 (S1 and S2; Copernicus [2016, 2017]), LiDAR Digital Terrain Model (DTM) (Government of Alberta, 2006), and Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) (USGS, 2006) data were used to generate fen probability in the Boreal Natural Region. All Sentinel and SRTM data were acquired, processed, and downloaded through Google Earth Engine (GEE) (Google Earth Engine Team, 2015). GEE stores Sentinel-1 (SAR imagery) ground range detected scenes which have been pre-

processed with the Sentinel-1 Toolbox (Sentinel Application Platform – Sentinel-1 Toolbox). These pre-processing steps include thermal noise removal, radiometric calibration, and terrain correction (Google Earth Engine Team, 2015). Dual polarization (VV VH) Sentinel-1 (S1) images were further processed in the GEE environment by performing an incidence angle correction (Gauthier *et al.,* 1998) and smoothing with a 3x3 Sigma Lee filter (Lee *et al.,* 2009) (credit to Guido Lemoine for GEE code). Once all S1 images were processed, a normalized difference of polarization (NDPOL) was calculated (see Table 1) and added to the available bands. To generate a single composite image for the S1 variables the per pixel mean of the VH and NDPOL bands were calculated. A total of 478 S1 images were used in the calculation of the VH and NDPOL variables.

Sentinel-2 (optical imagery) top of atmosphere data was acquired through GEE. Clouds, shadows, snow, and ice were removed with the QA60 band (a quality control band used to identify bad pixels) and further cloud masking was done using bands 1 (aerosols) and 11 (cloud). S2 images intersecting with the Boreal region during 2016-2017 leaf-on season (May 15 – August 31) were used to generate vegetation indices, and Principal component 1 and 2. PC1/2 metrics were generated with the 10m S2 bands (B2, B3, B4, and B8) in a principal component analysis. The merged S2 data was generated using a median composting algorithm where the median time series value for each pixel was selected as the most representative pixel. A total of 3,148 S2 images were used in the calculation of the vegetation indices and PC bands.

The DEM/DTM data used for modelling came from three sources: 1m bare earth (BE) LiDAR for the forest regions of Alberta(Government of Alberta, 2006), 15m BE LiDAR from the prairie regions of Alberta(Government of Alberta, 2017), and 30m SRTM DEM data for anywhere without LiDAR (USGSm 2006). The 1m BE LiDAR was mean aggregated to 10m to match the S1 and S2 data and the 15m BE LiDAR was resampled to 10m using cubic convolution method. The SRTM data was turned into a floating point raster, then resampled to 10m resolution using cubic convolution and then subsequently smoothed using a 7x7 pixel mean filter. Two topographic indices (TWI and TPI, Table 1) were calculated separately for each DEM data set and then merged when complete. All topographic indices were calculated in SAGA version 5.0.0 (Conrad *et al.*, 2015). All the input variables can be seen in Figure 3 and the equations and description can be seen in Table 1.

Training data was taken from the Alberta Biodiversity Monitoring Institute 3x7km Land Cover Photoplots (hereafter 3x7s) (ABMI, 2016). These photoplots are derived from high resolution 3D image interpretation and give detailed attribution of land cover information. They are typically very accurate with less than 1% of features possessing errors (ABMI, 2016).

Table 1: List of possible input variables used in the Fen probability model

Variable	Data source	Equation	Description
ARI	Sentinel-2	$\left(\frac{Band\ 8}{Band\ 2}\right) - \left(\frac{Band\ 8}{Band\ 3}\right)$	Anthocyanin Reflectance Index. An index sensitive to anthocyanin pigments in plant foliage (Gitelson *et al.*, 2001).
NDVI	Sentinel-2	$\frac{(Band\ 8 - Band\ 4)}{(Band\ 8 + Band\ 4)}$	Normalized Difference Vegetation Index. Index for estimating photosynthetic activity, and leaf area (Rouse *et al.,* 1973).
NDWI	Sentinel-2	$\frac{(Band\ 3 - Band\ 8)}{(Band\ 3 + Band\ 8)}$	Normalized difference Water Index from Mcfeeters (1996)
NDPOL	Sentinel-1	$\frac{(VH - VV)}{(VH + VV)}$	Normalized Difference of Polarization.
PC1	Sentinel-2	-	The first principal component of variation of Bands 2, 3, 4, and 8 of Sentinel-2 data
PC2	Sentinel-2	-	The second principal component of variation of Bands 2, 3, 4, and 8 of Sentinel-2 data

PSRI	Sentinel-2	$\frac{(Band\ 4 - Band\ 2)}{(Band\ 5)}$	Plant Senescence Reflectance Index. A ratio used to estimate the ratio of bulk carotenoids to chlorophyll (Hatfield and Prueger, 2010).
REIP	Sentinel-2	$702 + 40\left(\frac{\left(\frac{Band\ 4 + Band\ 7}{2}\right) - Band\ 5}{(Band\ 6 - Band\ 5)}\right)$	Red Edge Inflection Point. An approximation on a hyperspectral index for estimating the position (in nm) of the NIR/red inflection point in vegetation spectra (Herrmann, *et al.,* 2011).
TPI	LiDAR, SRTM DEMs	-	Topographic Position Index (TPI) generated in SAGA (Conrad *et al.*, 2015). An index describing the relative position of a pixel within a valley, ridge top continuum calculated in a given window size. TPI was calculated with a 500m moving window for this purpose (Weiss, 2001).
TWI	LiDAR, SRTM DEMs	-	Saga Wetness Index. A SAGA (Conrad *et al.*, 2015) version of the Topographic Wetness Index. Potential wetness of the ground based on topography (Böhner *et al.,* 2002).
VH	Sentinel-1	-	Vertical polarization sending horizontal polarization receiving SAR backscatter in decibels.

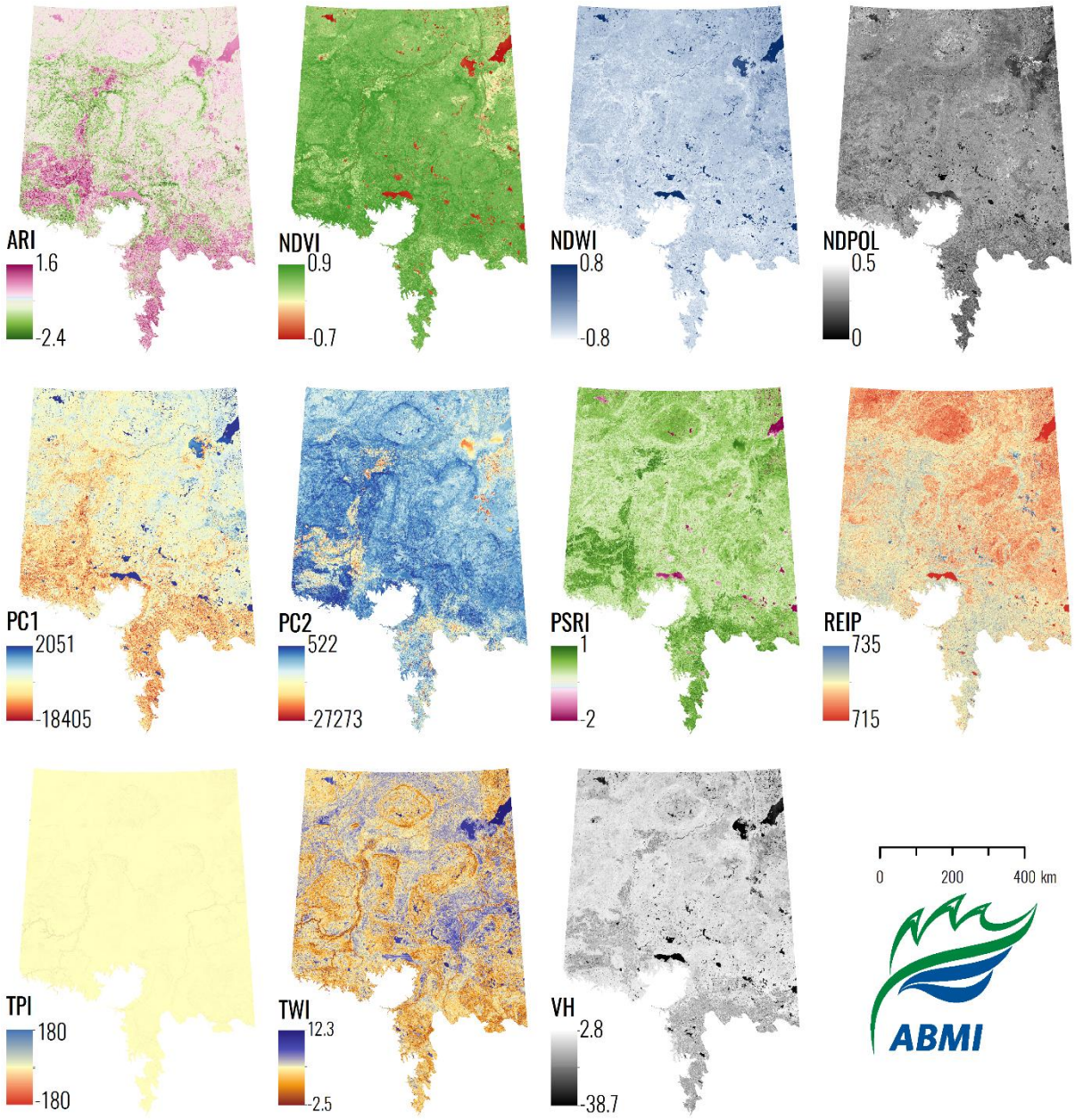
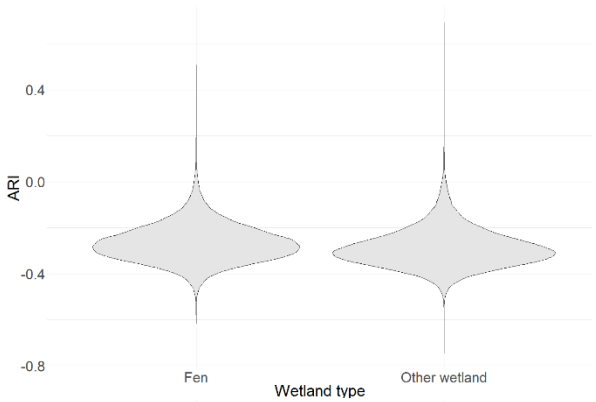
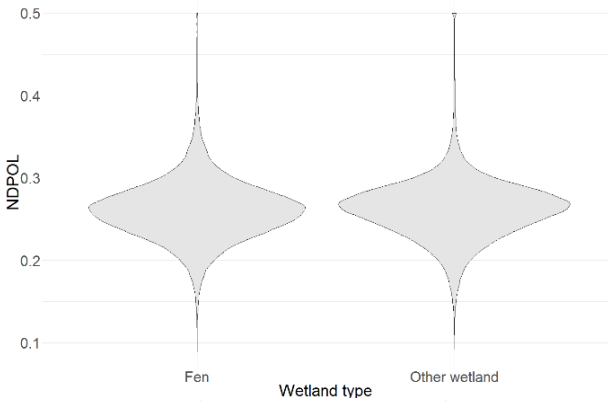


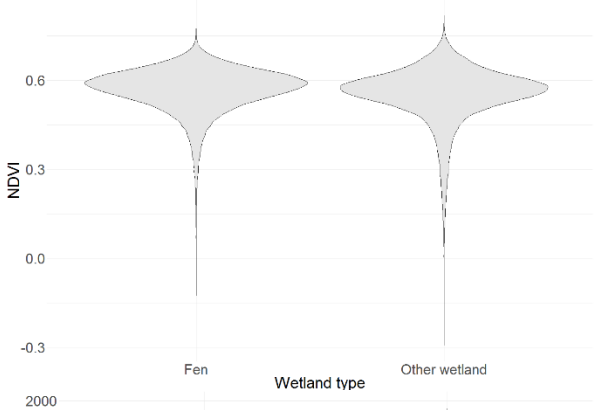
Figure 3: The 11 potential input raster inputs for the fen probability model. NDPOL and VH are generated from Sentinel-1, TPI and TWI are generated from DEM data, and the rest are generated from Sentinel-2 data.

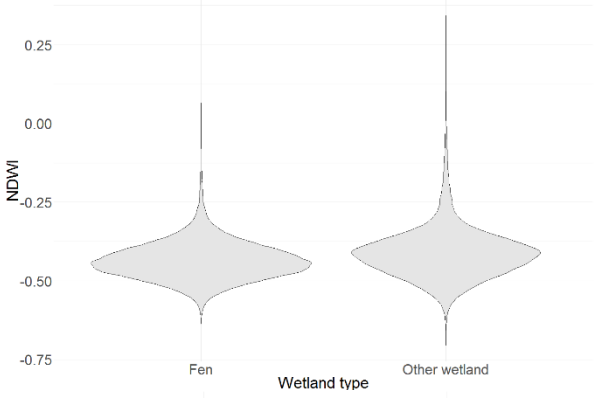
2.3 Data exploration and variable selection

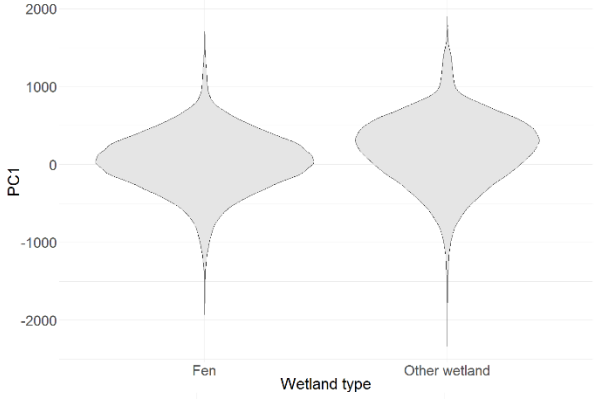
Figure 4 shows the distribution of values for all possible input variables divided by fen and other wetland classes. The fen training data comes from the 3x7s. Most variables show very little difference between the fen and other wetland classes. PC1 and TWI probably shows the largest difference between fen and other wetland classes. With the variables used here there seems to be very little distinction between fen and other wetland classes which further highlights the difficulty in correctly identifying fen.

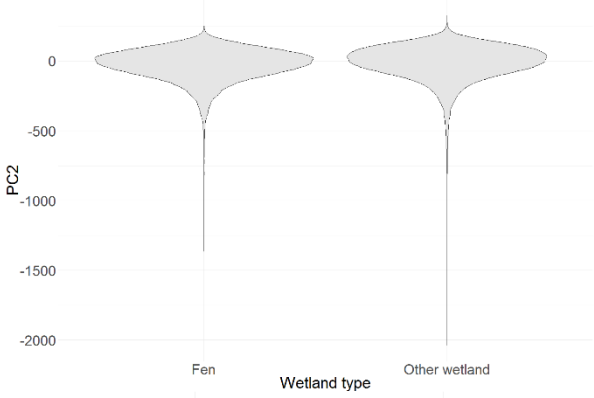

ARI
0.4
0.0
-0.4
-0.8
Fen
Other wetland
Wetland type

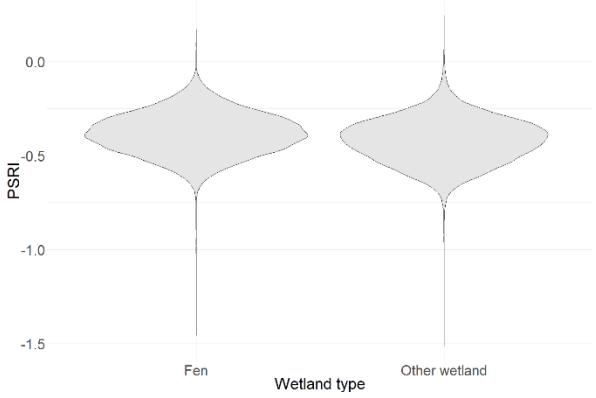

NDPOL
0.5
0.4
0.3
0.2
0.1
Fen
Other wetland
Wetland type

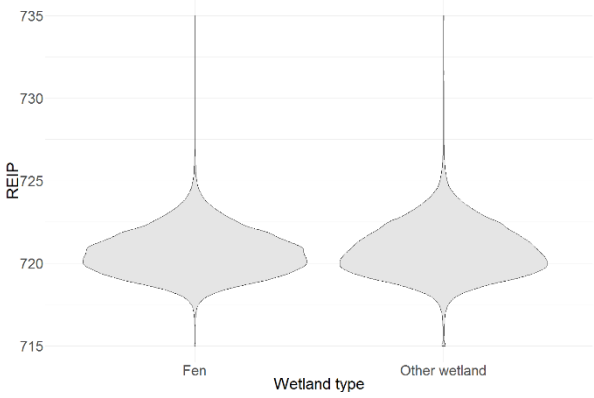

NDVI
0.6
0.3
0.0
-0.3
Fen
Other wetland
Wetland type


NDWI
0.25
0.00
-0.25
-0.50
-0.75
Fen
Other wetland
Wetland type


PC1
2000
1000
0
-1000
-2000
Fen
Other wetland
Wetland type


PC2
0
-500
-1000
-1500
-2000
Fen
Other wetland
Wetland type


PSRI
0.0
-0.5
-1.0
-1.5
Fen
Other wetland
Wetland type


REIP
735
730
725
720
715
Fen
Other wetland
Wetland type

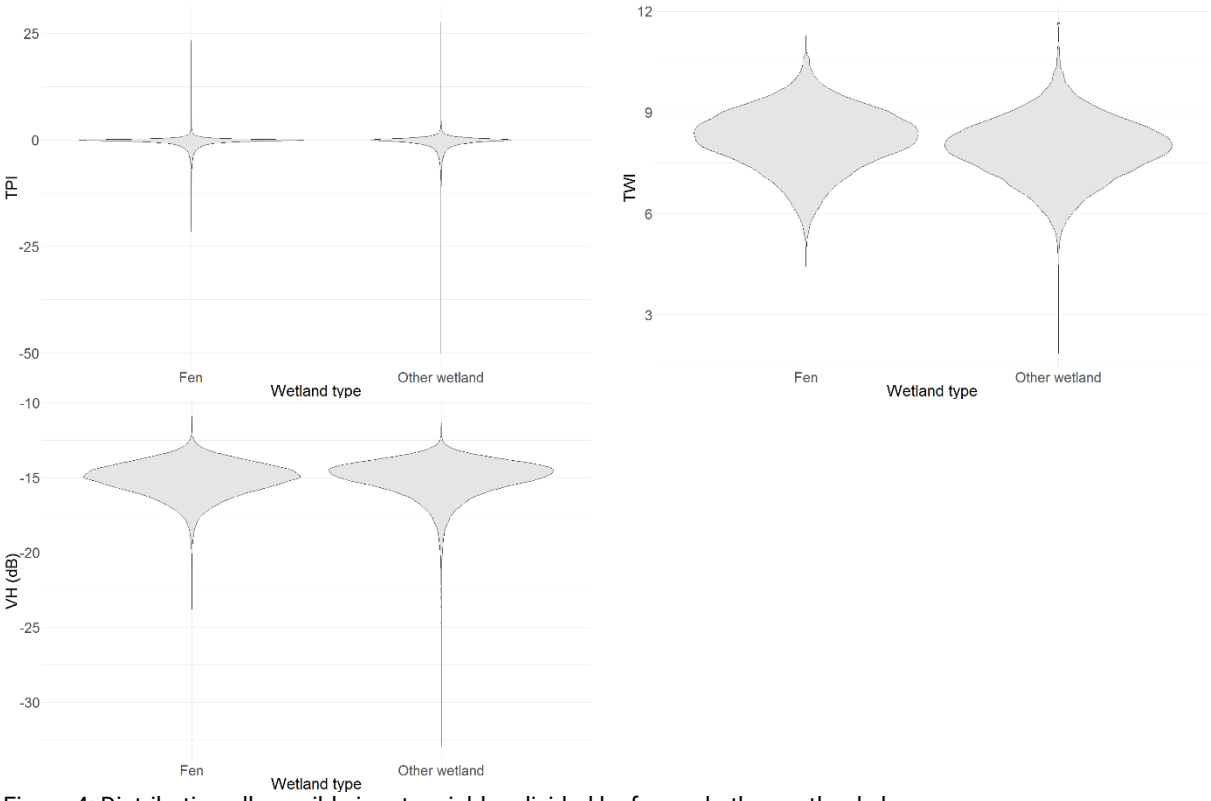


Figure 4: Distribution all possible input variables divided by fen and other wetland classes.

Using information from Figure 4, Table 2, and Table 3 we selected the most important variables for predictive modelling of fen. PC1 and TWI show some separation between classes and prove to be very important in the BRT model and therefore are chosen as model inputs. PSRI and ARI were chosen for model building as they had better correlation to fen presence than other variables (0.12 and 0.16) and were shown to be moderately important in the BRT model. TPI was chosen as it is a good counterpart to the TWI index (Hird *et al.,* 2017) and had high importance in the BRT model. Finally NDPOL was chosen as the most important S1 variable. These six variables provide a good mix of data sources (three from S2, two from DEMs, and one from S1).

Table 2: Cross correlation between the 11 possible input variables and a binary fen/other wetland grid.

	ARI	NDPOL	NDVI	NDWI	PC1	PC2	PSRI	REIP	TPI	TWI	VH	fen
ARI	1.00	-0.02	0.02	-0.24	-0.57	-0.52	0.69	-0.16	0.01	0.07	-0.40	0.12
NDPOL	-0.02	1.00	0.07	-0.04	0.02	0.15	-0.02	-0.24	0.10	0.03	0.04	-0.01
NDVI	0.02	0.07	1.00	-0.88	-0.61	0.51	-0.07	0.01	0.09	0.04	0.45	0.10
NDWI	-0.24	-0.04	-0.88	1.00	0.82	-0.21	-0.29	-0.04	-0.06	-0.08	-0.23	-0.16
PC1	-0.57	0.02	-0.61	0.82	1.00	0.20	-0.61	-0.08	-0.03	-0.13	0.08	-0.18
PC2	-0.52	0.15	0.51	-0.21	0.20	1.00	-0.73	0.02	0.08	-0.02	0.57	-0.06
PSRI	0.69	-0.02	-0.07	-0.29	-0.61	-0.73	1.00	-0.20	-0.02	0.08	-0.50	0.16
REIP	-0.16	-0.24	0.01	-0.04	-0.08	0.02	-0.20	1.00	-0.12	0.08	0.11	-0.05
TPI	0.01	0.10	0.09	-0.06	-0.03	0.08	-0.02	-0.12	1.00	0.25	0.08	0.04
TWI	0.07	0.03	0.04	-0.08	-0.13	-0.02	0.08	0.08	0.25	1.00	-0.05	0.16
VH	-0.40	0.04	0.45	-0.23	0.08	0.57	-0.50	0.11	0.08	-0.05	1.00	-0.02
fen	0.12	-0.01	0.10	-0.16	-0.18	-0.06	0.16	-0.05	0.04	0.16	-0.02	1.00

Table 3: The relative importance of the 11 possible input variables in the boosted regression tree fen predictive model. Variables chosen for predictive modelling are seen in green.

Variable	Relative importance
PC1	23.68
TWI	14.40
REIP	11.23
TPI	10.77
NDPOL	8.36
PSRI	7.88
ARI	6.87
VH	4.81
NDWI	4.70
NDVI	3.80
PC2	3.44

2.4 Wetland classification – machine learning algorithm

To classify the probability of fen occurrence a machine learning algorithm was developed in R Statistical Software (R Core Team, 2013). This algorithm uses a boosted regression tree modelling approach (Elith *et al.,* 2008). To build a model 1,500 random points were placed at a distance of at least 1 kilometer apart in known wetland areas delineated by the 3x7s. Training points were not placed in any locations within known human footprint features, areas with open water, or upland areas. The spatial delineations of these features are taken from the ABMI's Human Footprint Inventory (ABMI, 2017), the ABMI's Boreal surface water inventory (ABMI, 2017), and the ABMI's Boreal Wetland Probability data set (ABMI, 2017). The training fen/other wetland data was taken from the 3x7s. From these 1,500 points a data frame was built describing the values of the six input variables and their corresponding binary fen/other wetland information. This data frame was then put into the boosted regression tree modelling function using a tree complexity of 5, learning rate of 0.005, and bag fraction of 0.5 (Hird *et al.,* 2017). This model output: responses for the four input variables, variables importance, and Area Under the Receiver Operating Characteristic Curve (AUROC) value. The model was then used to predict fen probability given the six input variables. This process was repeated 40 times which generated 40 fen probability grids. This was done to reduce statistical overfitting and spatial auto-correlation (Parisien *et al.,* 2011). The mean value of these 40 grids was used to produce the final fen probability grid. Fen was then classified as any value above of a probability threshold of 0.5 resulting in a binary fen(1)/other wetland(0) raster.

2.5 Cross validation accuracy assessment

An independent cross validation accuracy assessment of the binary fen/other wetland layers was completed by generating 200,000 points in all 3x7 Boreal areas. Each 10m pixel was classified as fen or not fen (water, upland, other wetland). Values from the 3x7 training data and the modeled fen data were then extracted for each point. With this data a traditional accuracy was calculated along with a confusion matrix, and a kappa statistic.

An additional accuracy assessment was done only inside wetland areas to assess the power to differentiate fen from other wetlands. 200,000 points were generated inside 3x7 wetland areas and,

again, a traditional accuracy assessment along with a confusion matrix, and a kappa statistic were calculated.

2.6 Additional processing

Once fen probability was predicted across the boreal, areas with surface water (ABMI Boreal surface water inventory, 2017) were given a fen probability value of 0 while areas with human footprint types seen in Table 4 (ABMI Human Footprint Inventory (HFI) 2014, 2017) were given a fen probability value of 0 and areas deemed as upland areas (ABMI Wetland probability) were assigned a value of 0. The final probability raster was converted into a binary fen/other wetland grid. Another smoothed version of this binary grid was produced by using a 5x5 pixel majority filter.

Table 4: List of HFI 2014 sublayers used to assign a wetland probability of 0.

Human Footprint Inventory 2014 sublayer
Borrow Pits, Sumps, Dugouts, and Lagoons
Roads
Railways
Canals
Verge
Mine sites
Industrial sites
Well sites
Landfill
Other veg surfaces
Wind generation facilities
High density livestock operation
Residential areas
Cultivation

3 Results

The results of the BRT model show that PC1 was the most important input variable (Figure 5). TWI, REIP, and TPI were seen to be moderately important in the BRT model. NDPOL and ARI had the least overall influence on fen prediction. Figure 6 shows the mean partial dependence response curves for all six input variables and the standard deviation around the mean of the 40 models.

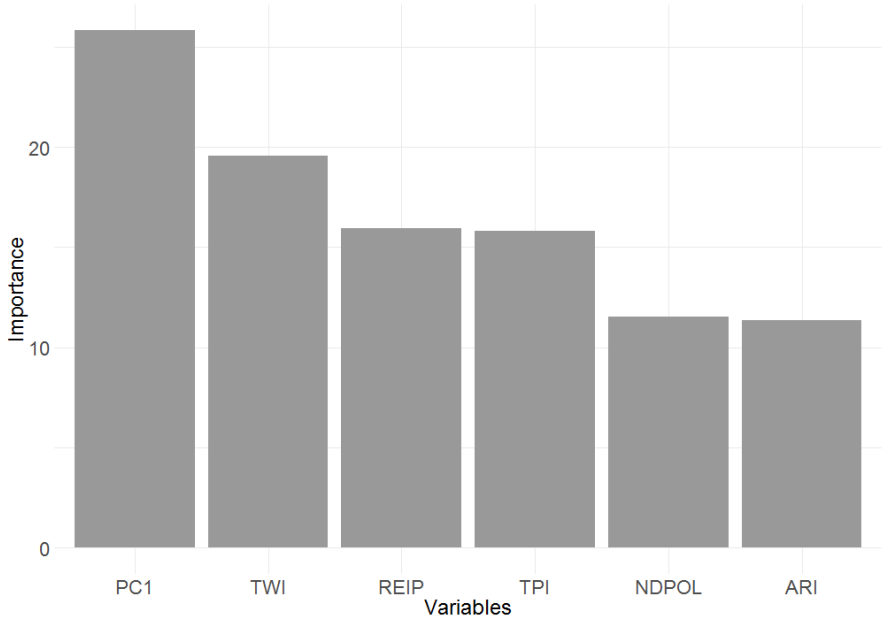


Figure 5: BRT model variable importance of all six input variables.

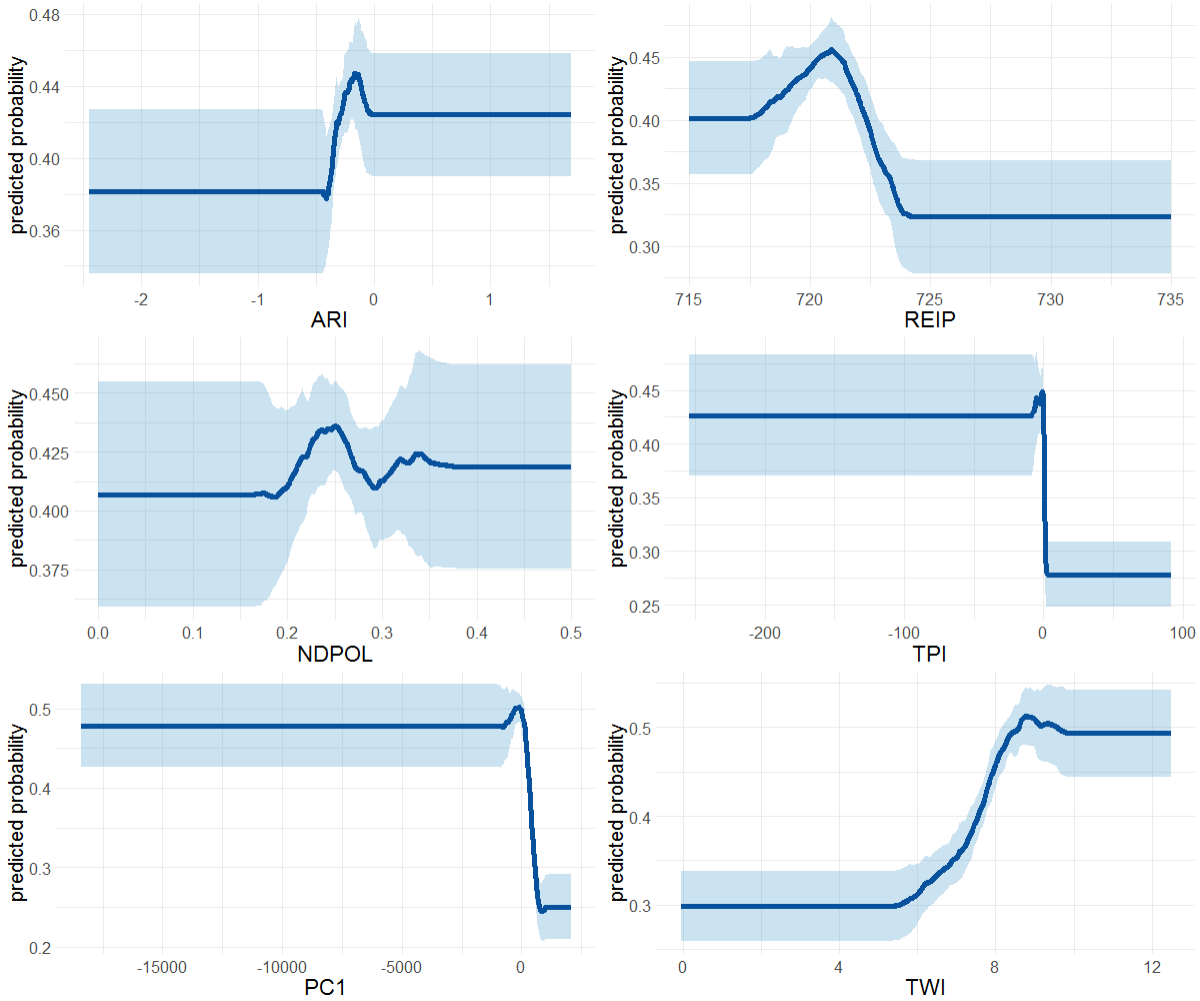


Figure 6: Mean partial dependence response curves of all six input variables for the wetland model. Predicted probability is the estimated probability of wetland occurrence at the given input value with all other variables held at their mean.

Figure 7 shows the model predicted across the Boreal Natural Region with the additional results of the ABMI wetland probability and Boreal Surface water mapped as land cover classes. Fen is mapped as any wetland area with a fen probability of greater than 0.45. Other wetlands (bog, swap, marsh) is thus any area that is not water, upland, or fen. The largest area of fen is the area south west of Fort McMurray around the bend in the Athabasca River. Overall fen makes up 12% of the area of the Boreal Natural Region.

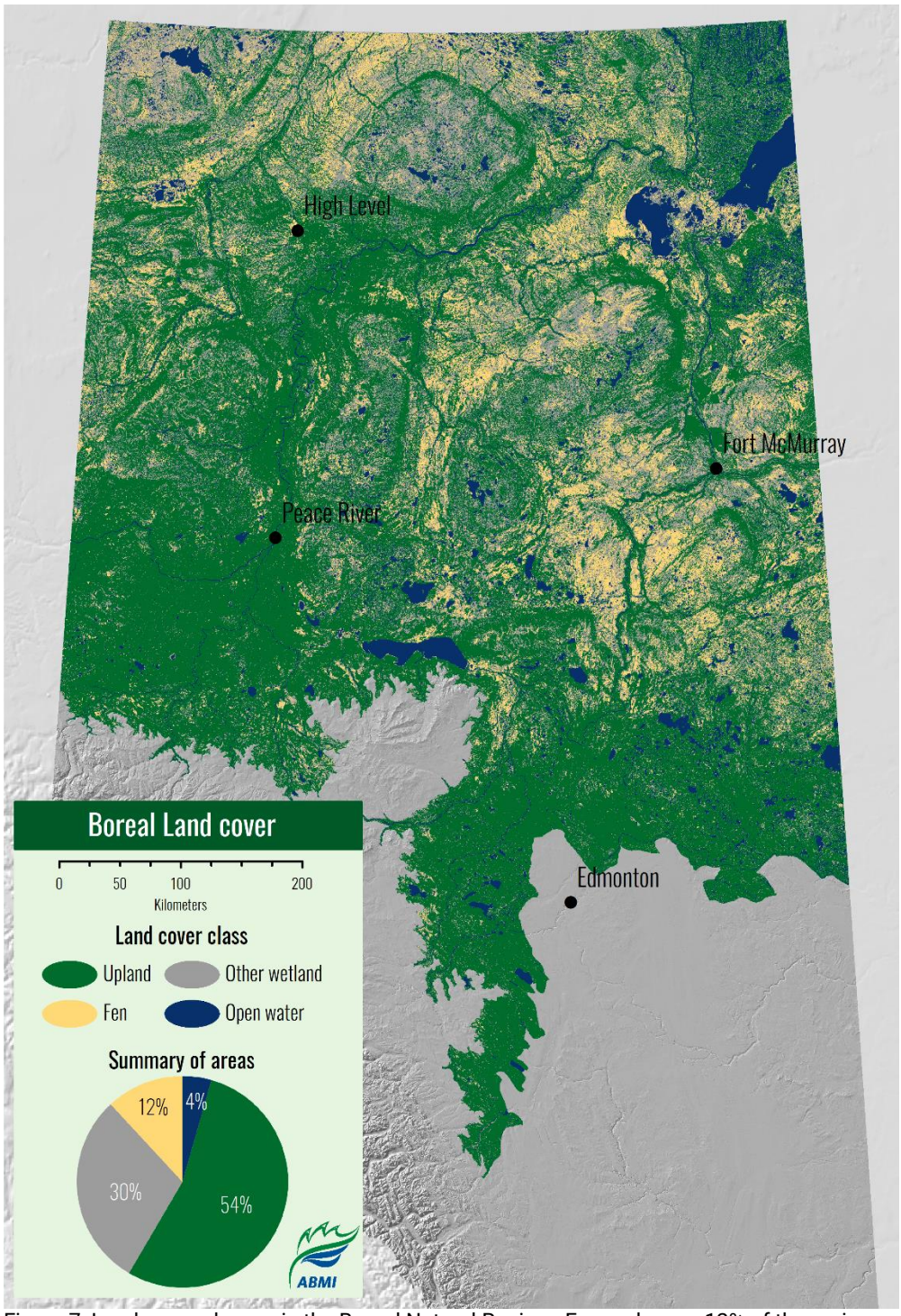


Figure 7: Land cover classes in the Boreal Natural Region. Fen makes up 12% of the region while water, other wetlands, and uplands make up 4, 30, and 54% respectively.

Table 5 shows the results of the fen/not fen cross validation accuracy assessment. The overall accuracy to the 3x7s was seen to be 84.8%, and the kappa statistic was 0.421. Table 6 shows the results of the fen/other wetland cross validation accuracy assessment. The overall accuracy to the 3x7s was seen to be 64.9%, and the kappa statistic was 0.265.

Table 5: Confusion matrix for the fen/not fen cross validation accuracy assessment to the3x7s. Overall accuracy shown in the bottom right in bold.

	Not fen	Fen	User Accuracy
Upland	153,797	20,146	0.884
Fen	10,169	15,620	0.606
Producer accuracy	0.938	0.437	**0.848**

Table 6: Confusion matrix for the fen/other wetland cross validation accuracy assessment to the3x7s. Overall accuracy shown in the bottom right in bold.

	Other wetland	Fen	User Accuracy
Upland	32501	16335	0.666
Fen	9650	15489	0.616
Producer accuracy	0.771	0.487	**0.649**

4 References

Alberta Biodiversity Monitoring Institute Remote Sensing Group. 2016. "ABMI Photo-Plot Quality Control Manual." Edmonton, Alberta.

Alberta Environment and Parks. 2017. "Alberta Merged Wetland Inventory." Edmonton, Alberta.

Belward, A.S., Taylor, J.C., Stuttard, M.J., Bignal, E., Mathews, J., and Curtis, D. 1990. "An unsupervised approach to the classification of semi-natural vegetation from Landsat Thematic Mapper data. A pilot study on Islay." *International Journal of Remote Sensing.* Vol. 11: pp. 429-445.

Böhner, J., Köthe, R., Conrad, O., Gross, J., Ringeler, A., and Selige, T. 2002. "Soil regionalisation by mean of terrain analysis and process parameterization." *European Soil Bureau*, Research Report No.7.

Conrad, O., Bechtel, B. Bock, M., Dietrich, H., Fisher, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. 2015. "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4." *Geoscientific Model Development.* Vol. 8, pp. 1991-2007.

Copernicus Sentinel-1 data [2016, 2017], European Space Agency.

Difebo, A., Richardson, M., and Price, J. 2015. "Fusion of Multispectral Imagery and LiDAR Digital Terrain Derivatives for Ecosystem Mapping and Morphological Characterization of a Northern Peatland Complex." In: *Remote Sensing of Wetlands: Applications and Advances*. Bocca Raton, FL, USA. Taylor and Francis Group.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. "A working guide to boosted regression trees." *Journal of Animal Ecology*, Vol. 77(No.4): pp. 802-813.

Gauthier, Y., Bernier, and Fortin, J-P. 1998. "Aspect and incidence angle sensitivity in ERS-1 SAR data." *International Journal of Remote Sensing*, Vol. 19(No.10): pp. 2001-2006.

Gitelson, A., Merzlyak, M., and Chivkunova, O. 2001. "Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves." *Photochemistry and Photobiology*, 71: 38-45.

Google Earth Engine Team. 2015. "Google Earth Engine: A planetary-scale geospatial analysis platform." https://eathengine.google.com.

Government of Alberta. 2006. Provincial LiDAR dataset. Edmonton, Alberta.

Hatfield, J.L., and Prueger, J.H. 2010. "Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages." *Remote Sensing*, Vol. 2: pp. 562-578.

Herrman, I., Pimstien, A., Karnieli, A., Cohen, Y., Alchanatis, V., Bonfil, D.J. 2011. "LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands." *Remote Sensing of Environment*, Vol. 115 (No.8): pp. 2141-2151.

Hird, J., DeLancey, E.R., McDermid, G.J., and Kariyeva, J. 2017. "Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping." *Remote Sensing*, Vol. 9(No.12): pp. 1315.

Lee, J-S., Wen, J-H., Ainsworth, T.L., Chen, K-S., and Chen, A.J. 2008. Improved Sigma Filter for Speckle Filtering of SAR Imagery. *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 47(No.1): pp.202-213.

McFeeters, S.K. 1996. "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." *Remote Sensing Letters,* Vol. 17(No.7): pp. 1425-1432.

Ozesmi, S.L., and Bauer, M.E. 2002. "Satellite remote sensing of wetlands." *Wetlands Ecology and Management*. Vol. 10, pp. 381-402.

Parisien, M.A., Parks, S.A., Krawchuk, M.A., Flannigan, M.D., Bowman, L.M., Moritz, M.A. 2011. "Scale-dependent controls on the area burned in the boreal forest of Canada, 1980-2005." *Ecological Application*, Vol. 21: pp. 789-805.

R Core Team. 2013. “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W. 1973. “Monitoring vegetation systems in the Great Plains with ERTS.” *In 3rd Symposium,* NASA SP-351 I, pp. 309-317.

Russel, G.D., Hawkins, C.P., and O’Neil, M.P. 1997. “The role of GIS in selecting sites for riparian restoration based on hydrology and land use.” *Restoration Ecology.* Vol. 5, pp. 56-68.

Weis, A. 2001. “Topographic position and landform analysis.” *Poster presentation, ESRI User Conference*, San Diego, California, USA.

Williams, J.A. 1992. “Vegetation Classification Using Landsat TM and SPOT-HRV Imager in Mountainous Terrain, Kananaskis Country, S.W. Alberta.” Alberta Recreation and Parks, Kananaskis Country Operations Branch, Environmental Management, Canmore, Alberta.