ABMI Wetland Inventory

Technical documentation
ABMI Geospatial Centre
March, 2021

ABMI ALBERTA BIODIVERSITY
MONITORING INSTITUTE

Table of contents

1. About

The Alberta Biodiversity Monitoring Institute (ABMI) province-wide wetland inventory is divided into three project areas representing ecoregions of Alberta with distinct wetlands. These three regions are: (i) the boreal and foothills region, which is characterized by large peatland complexes; (ii) the prairie region with wetlands usually occurring in small depressional potholes, and; (iii) the Rocky Mountain region where wetlands are constrained to narrow valleys. These three separate data sets all used open access Sentinel-1 and -2 data with some form of machine learning to classify four classes of the Canadian Wetland Classification System (bog, fen, marsh, swamp (Warner et al., 1997)) plus upland and open water. Each data set used specifically designed methodologies to capture unique spatial and temporal wetland patterns distinct for each area. These are described in detail in the three sections below.

2. Boreal and Foothills Inventory

2.1 Introduction

The ABMI boreal/foothills wetland inventory data is constrained by the Boreal, Shield, and Foothills Natural Region of Alberta (Figure 2-1). This area is characterized by extensive peatland (bog and fen) complexes. This region is the largest project area and has by far the most wetlands, area wise. This methodology uses a U-Net convolutional neural network machine learning approach to distinguish the four wetland classes. The detailed methods and results can be seen in (DeLancey et al., 2020) but a brief methods summary is below.
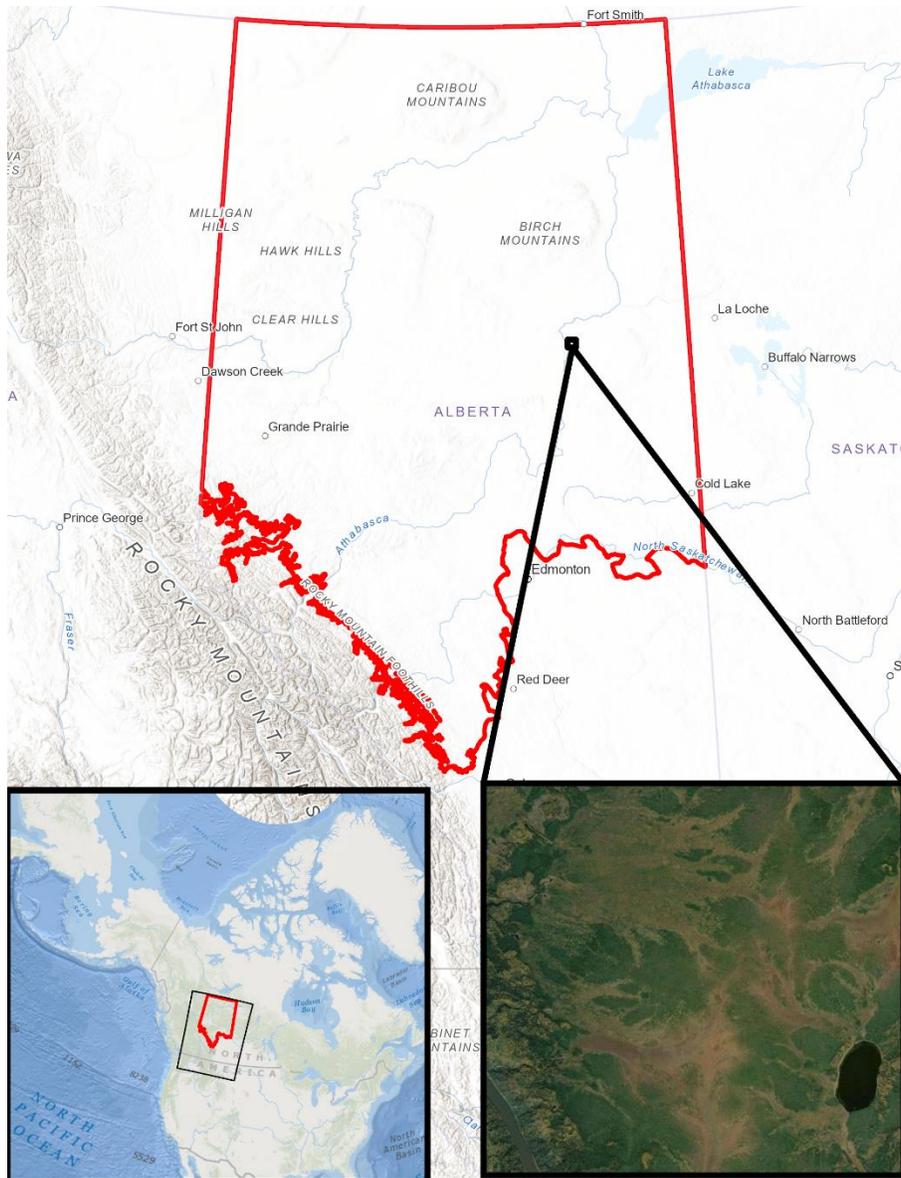
Figure 2-1: The Boreal/Foothills project region. Insets show the project area in the context of North America (left) and a typical peatland complex in north eastern Alberta (right).

2.2 Methods

2.2.1 Data

Data for the landcover classification comes from three sources: Sentinel-1 (S1) Synthetic Aperture Radar (SAR) data, Sentinel-2 (S2) optical data, and Advanced Land Observing Satellite (ALOS) Digital Surface Model (DSM). All input variables can be seen in Table 2-1. All data sets used in this study were acquired, processed, and downloaded through the Google Earth Engine (GEE) JavaScript API (Gorelick et al., 2017). Each Sentinel-1 image in GEE was pre-processed with the Sentinel-1 toolbox using the following steps: thermal noise removal, radiometric calibration, and terrain correction using the SRTM 30 m DEM. All Sentinel-1 dual

pol (VV VH) images over Alberta during the spring/summer time period (May 15th – Aug 15th) for the years 2017 and 2018 were used. This yielded 1,123 Sentinel-1 images. All these images were further processed with: an angle correction (Gauthier et al., 1998), edge mask for dark strips on the edges of images, and a multi-temporal filtering using a two month window (Bruniquel et al., 1997). To get the static backscatter inputs the mean pixel value of the image stack was calculated.

Sentinel-2 top of atmosphere (TOA) data was acquired over all of Alberta for the same period as the Sentinel-1 data. Note that Sentinel-2 surface reflectance products were not available in GEE at the start of product generation. All images with cloudy pixel percentages of less than 50% were used. This yielded a total of 4,479 Sentinel-2 images. All cloud and shadow pixels were masked out using an adapted Google Landsat cloud score algorithm and a Temporal Dark Outlier Mask (TDOM) method. To get the static Sentinel-2 inputs, the median pixel of each band in the pixel stack was chosen. This was done to eliminate any outlier bright or dark pixels. All vegetation indices seen in Table 2-1 were calculated with these Sentinel-2 median bands.

The ALOS 30 m Digital Surface Model (DSM) was acquired over all of Alberta. To match the resolution of Sentinel-1 and Sentinel-2 data, the DSM was resampled to 10 m and turned into floating point data type. Additionally a 5 x 5 pixel spatial mean filter was applied to the DSM for the purpose of creating more realistic hydrological indices (DeLancey et al., 2019). With the 10 m ALOS DSM, topographic indices were then calculated in SAGA version 5.0.0 (Conrad et al., 2015) across the province of Alberta.

The training and validation data for all models are derived from photo-interpreted polygons, which come from the Alberta Biodiversity Monitoring Institute's Landcover Photoplots data (ABMI, 2016). The ABMI photoplots are attributed and spatially explicit polygons, derived from high resolution 3D image interpretation. They include information on wetland type, wetland structure, forest type/structure, and more. The ABMI plots have undergone ground-truthing and are typically highly accurate (high 90% range) when compared to field data. For this study we extracted the following classes from the LC3 field: open water – 0, fen – 1, bog, – 2, marsh – 3, swamp – 4, upland – 5, wetland general – 6. It should be noted that we did not train models with the shallow open water class because the ABMI photoplots data does not have accurate representations of this class.

Table 2-1: List of input variables in the CNN model. Each variable lists its respective data source, description, equation, and, if needed, citation.

Variable	Data source	Equation	Description
ARI	Sentinel-2	$\left(\frac{B8}{B2}\right) - \left(\frac{B8}{B3}\right)$	Anthocyanin Reflectance Index. An index sensitive to anthocyanin pigments in plant foliage which is often associated with plant stress or senescence (Gitelson et al., 2001).
Band 2	Sentinel-2	-	The blue band of Sentinel-2. Central wavelength at 492 nm.
Band 3	Sentinel-2	-	The green band of Sentinel-2. Central wavelength at 559 nm.

Band 4	Sentinel-2	-	The blue band of Sentinel-2. Central wavelength at 664 nm.
DSM	ALOS	-	The raw elevation values from the ALOS DSM.
NDVI	Sentinel-2	$\dfrac{(B8 - B4)}{(B8 + B4)}$	Normalized Difference Vegetation Index. Index for estimating photosynthetic activity, and leaf area (Rouse Jr et al., 1974).
NDWI	Sentinel-2	$\dfrac{(B3 - B8)}{(B3 + B8)}$	Normalized difference Water Index (McFeeters, 1996).
PSRI	Sentinel-2	$\dfrac{(B4 - B2)}{(B5)}$	Plant Senescence Reflectance Index. A ratio used to estimate the ratio of bulk carotenoids to chlorophyll (Hatfield et al., 2010).
REIP	Sentinel-2	$702 + 40\left(\dfrac{\left(\frac{B4 + B7}{2}\right) - B5}{(B6 - B5)}\right)$	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., 2010).
TPI	ALOS	-	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 750m moving window for this purpose (Weiss, 2001).
TRI	ALOS		Topographic Roughness Index generated in SAGA.
TWI	ALOS	-	Saga Wetness Index. A SAGA version of the Topographic Wetness Index. Potential wetness of the ground based on topography (Böhner J, 2002).
VBF	ALOS	-	Multi Resolution Index of Valley Bottom Flatness (Gallant et al., 2003). This index measures the degree of valley bottom flatness at multiple scales. Large flat valleys are typical landscapes for wetland formation.
VH	Sentinel-1	-	Vertical polarization sending horizontal polarization receiving SAR backscatter in decibels.

2.2.2 Machine Learning Algorithms

The segmentation Convolutional Neural Net (CNN) was implemented in the Python programming language using the Keras (Atienza, 2018) deep learning library. The inputs used by our CNN model were: ARI, Band 2, Band 3, Band 4, DSM, NDVI, NDWI, PSRI, REIP, TPI, TRI, TWI, VBF, VH (Table 2-1). Every layer except DSM was clipped high and low based on 95th and 5th percentiles, and then standardized with mean subtraction and divided by the standard deviation. The training patch size was 224 by 244 (by 14 depth) and the label patch was 49 by 49 (by 6 depth). The output activations for the CNN were sigmoid units. The model was trained using the Keras Nadam optimizer (Nesterov Adam optimizer (Dozat, 2016)) with a combination of binary crossentropy and dice coefficient loss for the objective loss function. Candidate training patch indexes were created using a simple moving window with a stride of 10 and simple label counts were generated. During training, patches were randomly selected from the patch list and randomly rotated left or right by 90 degrees, flipped horizontally or vertically, or left as is. Since the wetland classes marsh and swamp were somewhat rarer than the other classes, during batch creation (using a batch size of 24) we ensured that there were at least six patches containing each of those labels. Using a geometrically decaying learning rate, the model was trained for 110 epochs where each epoch was composed of 4,800 training samples. Model training took approximately 3-4 hours and prediction over all of Alberta at 10 m resolution took a similar amount of time. Training and prediction were completed on a desktop with 64Gb of RAM and one Titan X (Maxwell) GPU.

2.2.3 Quality control

Areas of known upland classes from the ABMI's Human Footprint Inventory (ABMI, 2018) were automatically classified as upland habitat. These known areas include: cultivation, harvest areas, roads, mines, and urban areas. To smooth the prediction, a 5 x 5 pixel modal filter was applied to the final output. Lastly a cursory quality check was done at the scale of 15 x 15 km tiles. This QC process only fixed obvious errors such as missing large lakes or presence of large seams in the data.

2.3 Results

Overall, the Boreal/Foothills wetland data achieved 85% accuracy (0.58 kappa statistic) compared to the ABMI photoplot data. Table 2-2 shows the confusion matrix for six classes.

Table 2-2: Confusion matrix of the Boreal/Foothills section of the ABMI Wetland Inventory. This is generated from 300,000 random points placed inside the ABMI photoplots data set aside for validation. Producer and user accuracy are in italics while overall accuracy is in italic and bold.

	Open water	Fen	Bog	Marsh	Swamp	Upland	User
Open water	3898	85	1	217	100	313	*84.48*
Fen	38	17773	3007	323	5632	2440	*60.84*
Bog	0	3795	2374	0	534	365	*33.59*
Marsh	199	437	0	1136	643	445	*39.72*
Swamp	50	3352	678	200	4473	2198	*40.85*
Upland	626	5888	270	1634	10769	225756	*92.17*
Producer	*81.02*	*56.73*	*37.50*	*32.36*	*20.19*	*97.51*	***85.24***

3. Prairie inventory

3.1 Introduction

The ABMI prairie wetland inventory data covers the Grassland and Parkland Natural Region of Alberta. This area is distinct due to the prevalence of agriculture and pothole wetlands which are mainly marshes and swamps. This area is very seasonal with distinct wetland flooding cycles (typically flooded in spring and dry in the fall). The methodology for this area tries to fully capture these seasonal wetland cycles for 2017-2020 and achieve the best minimum mapping unit possible as many wetlands in this area are very small (below 800 m^2).

3.2 Methods

3.2.1 Study area

The study area encompasses the Grassland and Parkland regions of Alberta plus portions of some other Natural Regions to form a contiguous area (Figure 3-1). Wetlands in this area are typically marshes, swamps, and open water. This typical landscape can be seen in the top inset of Figure 3-1.
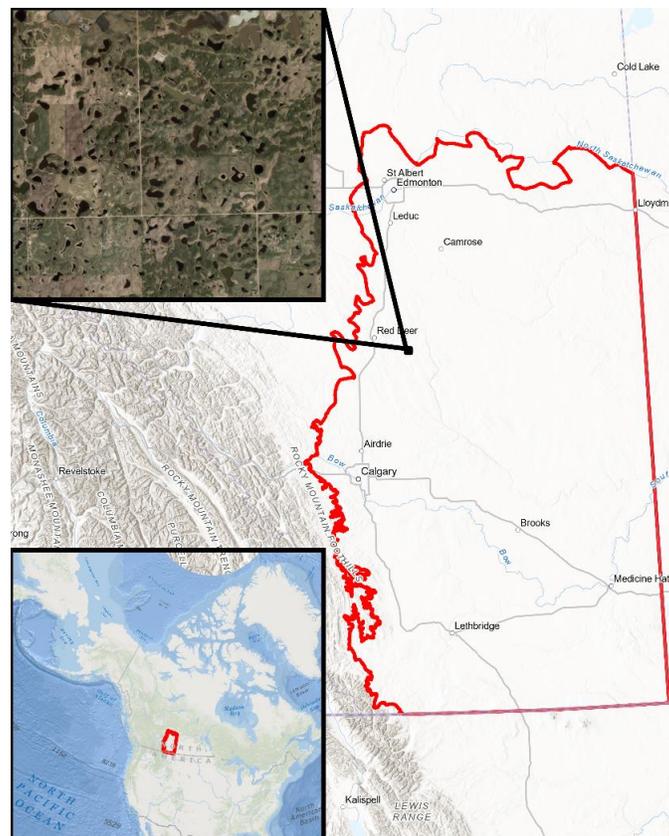

Figure 3-1: Location of the southern wetland mapping project (red). Top inset shows a typical landscape in this area dotted with marshes, open water, and swamps.

3.2.2 Data

The data for this project are from S2, S1, and the ALOS DSM. All these data were processed and acquired with GEE JavaScript API (Gorelick et al, 2017). Data for spring 2020 is S2 surface reflectance from April 24th - May 1st (Figure 3-2). Data for summer 2019 was from S2 surface reflectance during the time period July 13th - Aug 2nd (Figure 3-2). Data for fall 2018 were from S2 top of atmosphere reflectance from October 16th - October 23rd (Figure 3-2). S1 and the ALOS DEM were used to classify wetland segments into a wetland class. S1 ground range detected multi-temporal data were used in the VV and VH polarizations. The ALOS DEM was used to calculate topographic indices such as topographic wetness and topographic position (Hird et al., 2017; DeLancey et al., 2020).
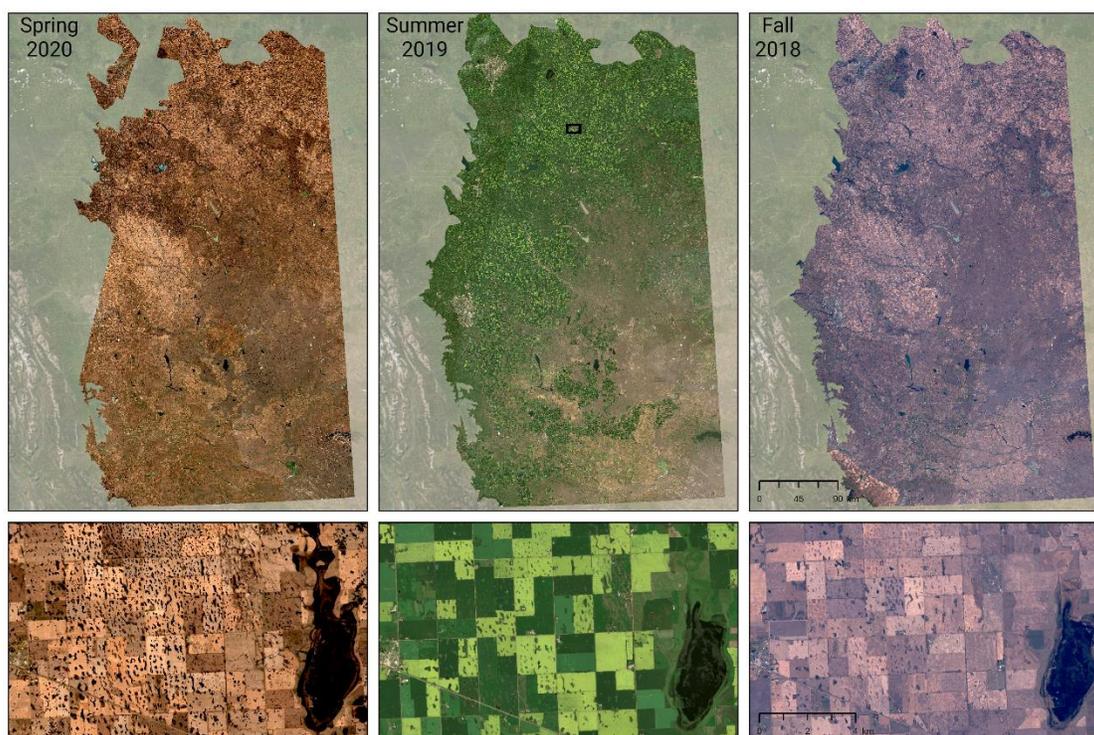
Figure 3-2: Sentinel-2 RGB imagery for spring 2020, summer 2019, and fall 2018. Note the different stages of flooding of the wetlands in each season showcased by bottom panels.

3.2.3 Wetland classification

3.2.3.1 Spring segmentation

The first step in the wetland classification was to segment the spring 2020 image into flooded and non-flooded areas. Nearly all low-lying wetland/open water areas were flooded in late April 2020 after the snow melt. Segmentation was completed with four bands (blue, green, red, NIR) from the spring 2020 image. The Simple Non-Iterative Clustering algorithm (SNIC) segmentation algorithm was used in GEE (Mahdianpari et al, 2020) at a 2 m scale to capture

small pothole wetlands. This segmentation was then exported to generate a flooded/non-flooded label data set.

These segments were labeled, by an interpreter, as flooded or not flooded based on the spring 2020 S2 image. A total of 3,000 segments was labeled and fed back into the segmentation algorithm as training data. The segments were then classified into wetlands and non-wetland areas using the random forest machine learning model. The algorithm was trained on 2,000 segments using 50 trees. Wetland segments were then classified across the whole study region. The original RGB imagery, segmentation, and resulting classification can be seen in Figure 3-3.
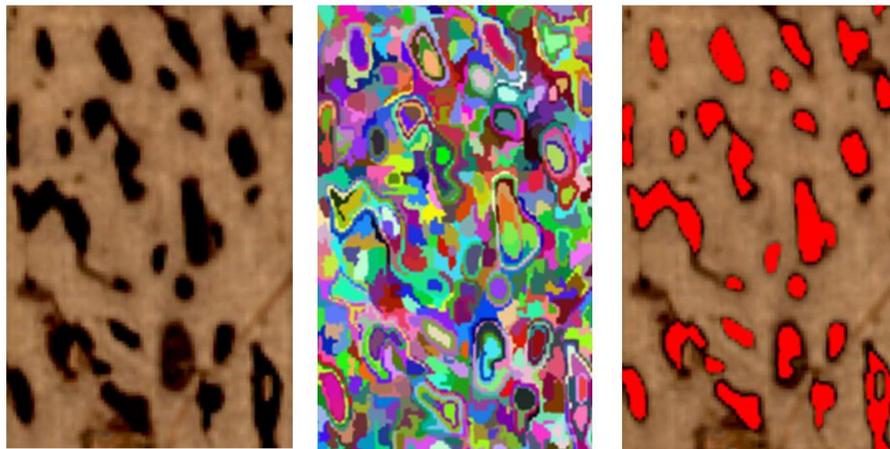
Figure 3-3: Segmentation of flooded areas based on a spring 2020 image. Left panel - spring RGB imagery; middle - segmentation of that image at 2 m, and; right - classification wetland areas.

3.2.3.2 Summer and fall segmentation

The summer 2019 and fall 2018 images were also classified into flooded and non-flooded areas using the random forest algorithm generated in the spring segmentation phase. This segmentation should generally show us permanent open water areas since fall 2018 was very dry across southern Alberta and open water in all three seasons should be a good estimation of permenant open water. The fall segments were also overlaid onto the spring segments using the union function. This was done to differentiate the open water from the marsh/swamp areas of the flooded potholes (see Figure 3-4).
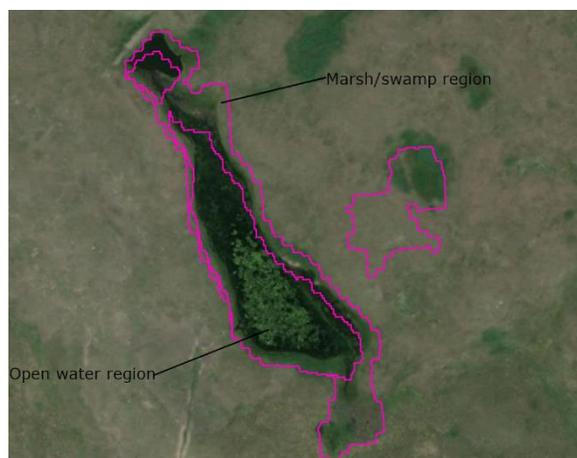
Figure 3-4: Example wetland area with open water region (fall 2018 segment) and marsh/swamp region (spring 2020 segment).

3.2.3.3 Assigning a wetland class

After combining the spring and fall flooded segments, 1,084,312 possible wetland segments were identified in southern Alberta with a minimum size of 400 m^2. We then randomly sampled 2,000 of these polygons and selected an additional 900 (class balancing) to generate a training data set. A trained photo interpreter used spring and fall S2 imagery along with ESRI high resolution satellite imagery to produce wetland class labels for these polygons (i.e., marsh, swamp). The photo interpreter would also assign "upland" class if it was found that the polygon was not actually in a wetland or open water area.

These label data were then combined with all remote sensing data that could be extracted from these wetland polygons. A list of all remote sensing data extracted for the wetland polygons can be seen in Table 3-1. These data include static S2 data, multi-temporal S1 and S2 data, and topographic data from ALOS. The spatial mean and standard deviation of these data were included as possible variables, as well as the area and perimeter ratio of the polygons. This totaled 92 possible input variables. All these variables, plus labels, were then put into an exploratory analysis script/Rmarkdown script which includes the following: violin plots, 3D plots, and machine learning variable importance.

Table 3-1: List of remote sensing variables extracted for wetland polygons.

Variable	Data source	Equation	Season	Spatial reducer
ARI	Sentinel-2	1/B3-1/B5	Spring, summer and fall	Mean and sd
Area	Geographic	-	-	-
AreaRatio	Geographic	Area/Perimeter	-	-
dARI	Sentinel-2	$ARI_{spring} - ARI_{fall}$	-	Mean and sd
B2	Sentinel-2	-	Spring, summer and fall	Mean and sd
B3	Sentinel-2	-	Spring, summer and fall	Mean and sd
B4	Sentinel-2	-	Spring, summer and fall	Mean and sd
B8	Sentinel-2	-	Spring, summer and fall	Mean and sd
B11	Sentinel-2	-	Spring, summer and fall	Mean and sd
dB11	Sentinel-2	$B11_{spring} - B11_{fall}$	-	Mean and sd
B12	Sentinel-2	-	Spring, summer and fall	Mean and sd

NDVI	Sentinel-2	(B8 - B4)(B8 + B4)	Spring, summer and fall	Mean and sd
dNDVI	Sentinel-2	$NDVI_{spring} - NDVI_{fall}$	-	Mean and sd
dSummerNDVI	Sentinel-2	$NDVI_{spring} - NDVI_{summer}$	-	Mean and sd
NDWI	Sentinel-2	(B3 - B8)(B3 + B8)	Spring, summer and fall	Mean and sd
dNDWI	Sentinel-2	$NDWI_{spring} - NDWI_{fall}$	-	Mean and sd
TPI250	ALOS	-	-	Mean and sd
TWI	ALOS	-	-	Mean and sd
VH	Sentinel-1	-	Spring and fall	Mean and sd
dVH	Sentinel-1	$VH_{spring} - VH_{fall}$	-	Mean and sd
DPOL	Sentinel-1	VV/VH	Spring and fall	Mean and sd
dDPOL	Sentinel-1	$DPOL_{spring} - DPOL_{fall}$	-	Mean and sd
VV	Sentinel-1		Spring and fall	Mean and sd
dVV	Sentinel-1	$VH_{spring} - VH_{fall}$	-	Mean and sd

Based on the results of our data exploration (which can be seen here: https://rpubs.com/edelance/SWM-explore), we input the following variables into a RandomForest algorithm in GEE: B11fall_mean, B12fall_mean, B4fall_mean, B8fall_mean, B8spring_mean, NDWIfall_mean, SWI_mean, TPI_mean, VVfall_mean, dB11_mean, dB12_mean, dNDWI_mean, dVH_mean.

3.2.4 Hydroperiod attribution

Hydroperiod (the frequency of time a wetland is covered with water) was calculated for a smaller area of southern Alberta. This hydroperiod data was calculated using multiple S2 images from 2017-2020 and classifying each image into flooded and non-flooded segments much like section 2.3.1. These binary rasters were then added together and divided by the total number of images in the final image stack to get percent of time flooded. Each polygon with hydroperiod information in the wetland inventory was then assigned a hydroperiod value by taking the mean hydroperiod inside the wetland or open water polygon.

3.2.5 Google Earth Engine code base

About 90% of the prairie wetland inventory workflow was completed in Google Earth Engine. The code links and code descriptions can be seen below.

1. Acquire spring 2020 S2 imagery - https://code.earthengine.google.com/8e975a57ddfdd7ef3785b1e7d72d2762
2. Acquire summer 2019 S2 imagery - https://code.earthengine.google.com/ce7e624ef8899fd4abecee785e977289
3. Acquire fall 2018 S2 imagery - https://code.earthengine.google.com/da5d1019589e1ca14ff6dbed63c60845
4. Spring flooded segmentation - https://code.earthengine.google.com/3b476d420aaf0e1b050d9b0cea1c5151
5. Summer flooded segmentation - https://code.earthengine.google.com/04d66310ce53caa74a57e680be0884bf
6. Extract multi sensor wetland segment stats - https://code.earthengine.google.com/a3706f3c07cd632ba3af116404c3898e
7. Classify wetland polygons into wetland class - https://code.earthengine.google.com/3d31493d82938a41fec13da27563fa7f

3.2.6 Quality control

The wetland inventory polygons were inspected at a scale of 1:33,367. Wetland polygons were removed if they were deemed to be false wetlands and some wetland polygons were manually digitized if they were missed by the classification algorithm. Additionally, areas with steeper slopes were considered not to be wetlands and thus these areas were masked out. Very limited alteration of the wetland boundaries or changing of the wetland class was done in the QC process.

3.2.7 Accuracy assessment

To assess the accuracy of the wetland classification results, independent validation data was acquired using ESRI base layer imagery, and the multiyear/season S2 data. Seven 1 x 1 km plots were chosen and all wetlands over 400 m^2 were digitized and assigned a wetland class under the Alberta Wetland Classification System. As suggested by the Alberta Wetland Inventory Standards (Government of Alberta, 2020), wetland perimeters were not assessed and therefore central points were generated for each wetland polygon (multiple points for large wetlands and a single point for smaller wetlands). Additionally, random points were generated in non-wetland areas to serve as upland reference data. Each reference point was then compared to the modeled results and the accuracy assessment was reported in a confusion matrix along with a kappa statistic and per-class F1-score.

3.3 Results

In total, 1,050,794 wetland polygons were identified in the study area with a minimum mapping unit of 400 m^2. The results show ecologically relevant shapes which capture the patterns seen in prairie pothole landscapes. The accuracy assessment shows a 93% accuracy when distinguishing wetland, open water, and upland areas (Table 3-2) and shows a 90% accuracy when classifying at the wetland class level with a kappa statistic of 0.80 (Table 3-3). User accuracies are shown to be ≥82% for each class. The per-class F1-scores are as follows: open water = 0.85, marsh = 0.80, swamp = 0.69, and upland = 0.96. Note that fen was present in the predicted data, but it was not abundant enough to be available in the validation data.

Table 3-2: Confusion matrix (at the wetland, open water, upland level) for the Prairie Wetland Inventory product. Producer and user accuracies are in italics and overall accuracy is in bold.

	Open water	Wetland	Upland	User
Open water	20	4	0	*83.33*
Wetland	3	119	5	*93.70*
Upland	0	25	341	*93.17*
Producer	*86.96*	*80.41*	*98.55*	**92.84**

Table 3-3: Confusion matrix (at the wetland class level) for the Prairie Wetland Inventory product. Producer and user accuracies are in italics and overall accuracy is in bold.

	Open water	Marsh	Swamp	Upland	User
Open water	20	3	1	0	*83.33*
Marsh	3	75	9	4	*82.42*
Swamp	0	5	30	1	*83.33*
Upland	0	14	11	341	*93.17*
Producer	*86.96*	*77.32*	*58.82*	*98.55*	***90.14***

4. Rocky Mountain Inventory

4.1 Introduction

The final study area of the ABMI wetland inventory consists of the Rocky Mountain natural region of Alberta. Wetlands are less common in the Rocky Mountain ecoregion than other natural regions of Alberta. Bogs, in particular, are rarely found in this ecoregion and were not represented in the training data, thus, only the wetland classes of marsh, swamp, fen, and open water were classified for the Rocky Mountain ecoregion. Topography plays a large role in constraining the distribution of wetlands in this region. This area was classified with an object-oriented supervised Random Forest machine learning algorithm. The classification strategy focused on topographic variables and spectral variation across time to make the most of relatively sparse training sites and cloud cover constraints.

Figure 4-1: The Rocky Mountain ecoregion. Insets shows the Vermillion Lakes wetland complex near Banff.

4.2 Methods

4.2.1 Data

The training and validation data for the classification is derived from photo-interpreted polygons from the photoplots data sets generated by the Alberta Biodiversity Monitoring Institute (ABMI 2016). A few additional training sites were added with an inventory of labeled Google Photosphere sites.

The classification imagery in the Rocky Mountain ecoregion wetland classification comes from two sources: Sentinel-2 optical data for the 2020 snow-free season, and topographic data derived from the Advanced Land Observing Satellite (ALOS) Digital Surface Model (DSM). Unlike the rest of Alberta, Sentinel-1 SAR data was not used in this classification because of issues of radar layover, foreshortening, and shadowing in mountainous areas, and because its inclusion did not improve the accuracy of the final product.

Sentinel-2 Level 2 imagery was a median composite of all cloud-free dates between mid-June and mid-September acquired through the GEE JavaScript API (Gorelick et al. 2017) (early June data was too cloudy for use, and spring imagery was avoided due to lingering snow cover at high elevations). All images with a cloudy pixel percentages of less than 30% were used, and remaining cloud pixels were masked out using Sentinel-2's built-in quality band (QA60) and its cirrus cloud detecting band (Band 1). Cloud shadows were avoided by using the median for the composite. The median composite image was visually inspected to ensure no clouds or cloud shadows remained.

The Rocky Mountain ecoregion wetland classification used topographic indices TPI (Topographic Position Index), TWI (Topographic Wetness Index) and VBF (Valley Bottom Flatness) calculated in SAGA GIS from ALOS 30 m (resampled to 10m) DSM, the same topographic data as used in the Boreal/Foothills and Prairie Alberta wetland classifications (Hird et al, 2017). Two versions of the TPI were used in the classification, one calculated with a 250 m moving window and one with a 750 m moving window.

4.2.2 Classification Approach

An object-oriented supervised classification was implemented with the Random Forest machine learning algorithm. The imagery was segmented into clusters (objects) with Google Earth Engine's SNIC algorithm. In Google Earth Engine, the size of the clusters is determined by the map scale at which clustering is done. For this region, a scale of 5 m per pixel was chosen to best represent the size and shape of natural features. This is not to be confused with the input imagery's native resolution, which was 10-20 m.

4.2.2.1 Derivation of Variables

The sixty variables ultimately used in the Rocky Mountain ecosystem landcover classification were derived from:

1. Topographic variables as described above, either averaged (mean) or standard deviation taken over the extent of each cluster. The standard deviations represented the variability of the topographic characteristics over space within a contiguous sample of a class (the cluster). According to the Random Forest algorithm, the most important variable in this category was the mean of Valley Bottom Flatness, although all the topographic variables and their standard deviations rated quite highly in this region.

2. Bands from median composite, either averaged (mean) or standard deviation taken over the extent of each cluster. The standard deviations represented the variability of the band over space within a contiguous sample of a class (the cluster). Examples: B12_mean (the second shortwave infrared band), B8_sd (standard deviation of the near infrared band).

3. Spectral indices from median composite, either averaged (mean) or standard deviation taken over the extent of each cluster. The standard deviations represented the variability of the index over space within a contiguous sample of a class (the cluster). The spectral indices calculated for this classification were:

EVI = Enhanced Vegetation Index (Huete et al., 1997)
NARI = Normalized Anthocyanin Reflectance Index (Bayle et al., 2019)
NDWI1 = Normalized Difference Wetness Index for detecting water bodies (McFeeters 1996)

NDW12 = Normalized Difference Wetness Index for water content of leaves (Gao, 1995)
Examples of variables: NDWI1_mean, NDWI1_sd

4. The standard deviation of a spectral band or index over the entire image collection (June to September). This represents the spectral variability of the class over time. This value can then be either averaged or the standard deviation taken per cluster. Examples of variables representing the standard deviation over time by cluster means are NDWI2_sd_mean and B2_sd_mean. The standard deviation of this value represents variability across both time and space. Examples are EVI_sd_sd and B12_sd_sd.

5. For selected variables, the difference between the 90th percentile (high values) and the 10th percentile (low values) over the date range was calculated. Percentiles were used rather than min and max to avoid outliers. This variable represents the range of a band or index for the class, and also could be either averaged per cluster or the standard deviation taken. Examples are diffNDWI2_mean and diffB7_sd.

Table 4-1 shows the final list of 60 variables used in the classification in order of importance within categories. In general, the topographic variables were the most important, followed by spectral indices and band means. The 10 least important or most redundant variables were removed after several trials of Random Forest. Variable removal stopped when class accuracies as shown in the validation error matrix started to fall.

Table 4-1: Variables used in Rocky Mountain Wetland classification

Topographic	Spectral Index from median composite	Spectral Band from median composite	Spectral Index – SD over snow-free season	Spectral Band - SD over snow-free season	Difference between 90th and 10th percentile over season
VBF_mean	NDWI1_mean	B12_mean	NDWI2_sd_mean	B12_sd_sd	diffNDWI2_mean
TPI750_mean	EVI_mean	B11_ mean	EVI_sd_mean	B2_sd_mean	diffB7_sd
TPI250_mean	NARI_mean	B6_ mean	NARI_sd_mean	B2_sd_sd	diffNARI_mean
SWI_mean	NDWI2_mean	B7_ mean	EVI_sd_sd	B8_sd_sd	diffNDWI2_sd
TPI250_sd	NDWI1-sd	B5_ mean	NDWI2_sd_sd	B7_sd_sd	diffB12_mean
VBF_sd	EVI_sd	B8_ mean	NARI_sd_sd	B8_sd_mean	diffNARI_sd
TPI750_sd	NARI_sd	B3_mean		B7_sd_mean	diffB12_sd
SWI_sd	NDWI2_sd	B8_sd		B5_sd_mean	diffB7_mean
		B7_sd		B11_sd_sd	
		B3_sd		B11_sd_mean	
		B2_mean		B5_sd_sd	
		B2_sd		B12_sd_mean	
		B4_mean			
		B4_sd			
		B12_sd			
		B11_sd			
		B6_sd			
		B5_sd			

4.2.3 Google Earth Engine code base

The whole classification workflow, with the exception of training data preparation and quality control post-processing, was completed in Google Earth Engine and can be accessed by the following link:

https://code.earthengine.google.com/df5c265a3dd0354849dd51d7a27df5d6

4.2.4 Quality Control

In the post processing phase, areas of known upland classes from the ABMI's Human Footprint Inventory (ABMI 2018) were automatically classified as upland habitat. The classification result was scanned through visually, manually editing any obvious errors - these mainly took the form of small mountain lakes which were confused with shadows. To smooth the result ArcPro's Majority Filter which uses a 3 x 3 moving window was applied to the final output three times, and PCI Geomatica's SIEVE function was applied to remove isolated pixels. Finally, the raster dataset was converted to polygons and upland polygons and polygons less than 400 m^2 were removed.

4.3 Results

The Rocky Mountain Ecoregion classification identified 25,761 wetland and water polygons (minimum mapping unit 400 m^2) covering 3.2 % of the ecoregion area. Wetlands are mainly found in valleys and other low-lying areas. The most common wetland class is swamp with 0.9% of the ecoregion area, followed by fen at 0.7% and marsh at 0.5%. Open water (both deep and shallow water bodies) takes up 1.0% leaving 96.8% of the region as upland.

Thirty percent of the ABMI photo interpreted data prepared for training the classification were held back as validation points. According to this validation data, the overall accuracy of the classification is 84.5% (Table 4-2) with a kappa value of 0.69. The classification can distinguish between upland and lowland (wetland and water) with a 90% overall accuracy. Individual wetland class accuracies range from 58% to 69% with the greatest incidence of confusion between swamp and fen.

Table 4-2: Confusion Matrix comparing the results of the Rocky Mountain Wetland Classification against 1292 validation points. Producer and user accuracies are in italics and overall accuracy is in bold.

	Open water	Fen	Marsh	Swamp	Upland	User
Open water	53	0	0	0	7	*88.3*
Fen	1	85	6	21	28	*60.3*
Marsh	1	9	38	4	10	*61.3*
Swamp	4	16	8	100	44	*58.1*
Upland	0	15	7	20	816	*95.1*
Producer	*89.8*	*68.0*	*64.4*	*69.0*	*90.2*	**84.5**

5. ABMI Wetland Inventory

The final ABMI Wetland Inventory, with all three wetland regions merged together, contains 3,093,620 wetland polygons. Figure 5-1 shows swamp and marsh potholes dominating the prairie regions of Alberta and peatlands (fen and bog) dominating the Boreal. According to this data, 78% of the province is upland, 18% wetland, and 4% open water. By wetland class, fen makes up 12% of Alberta by area followed by swamp 3%, marsh 2% and bog 2%. The transition between the Prairie and Boreal region shows a large difference in predicted marsh habitat. This is due to seasonal and temporary marshes being captured in the Prairie Wetland Inventory while the Boreal inventory was developed with a static classification approach. The next version of the ABMI Wetland Inventory aims to add in the classification of seasonal and temporary wetlands in the Boreal transition zone to better reflect seasonal wetlands across all of Alberta.
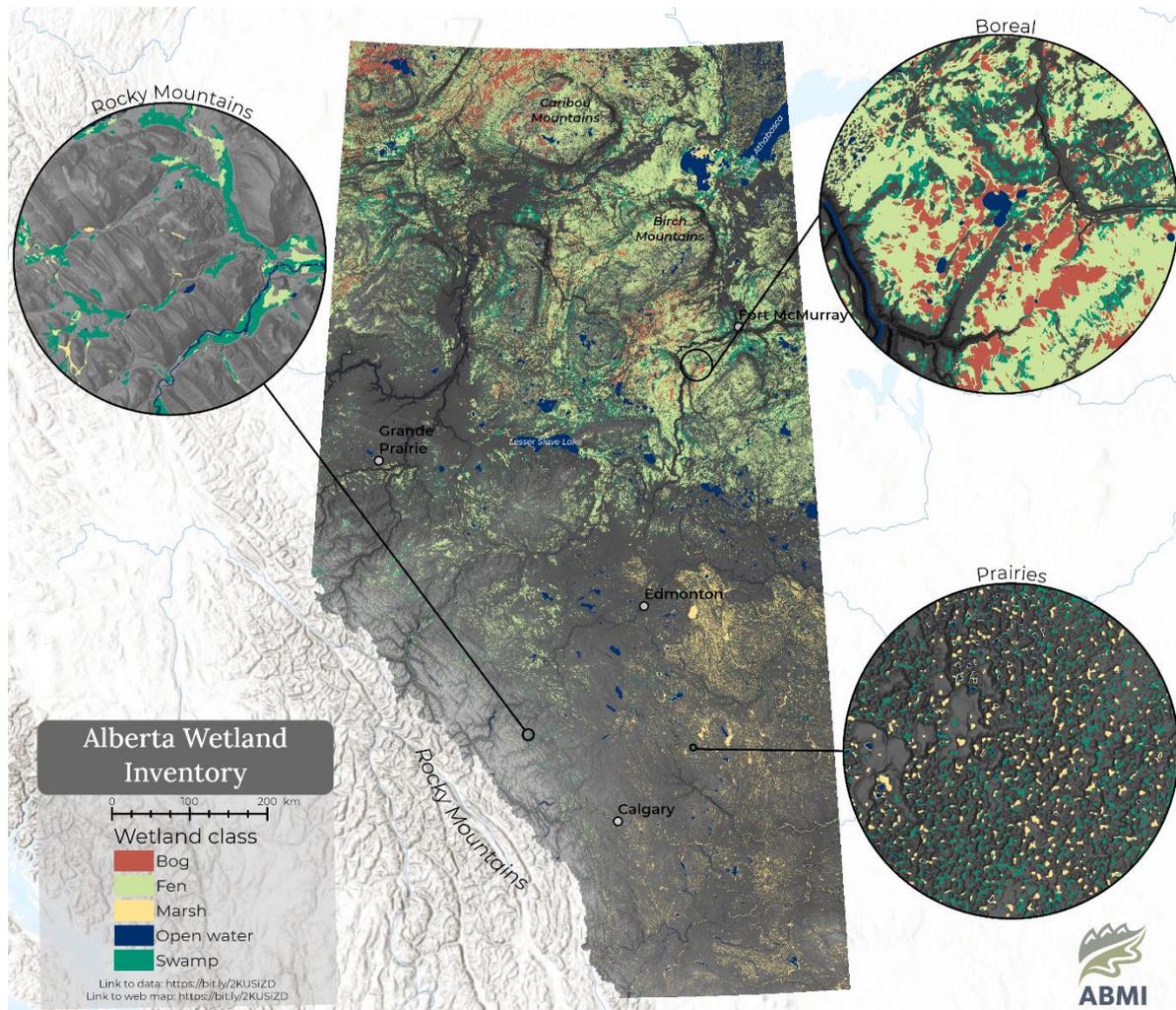

Figure 5-1: The ABMI Wetland Inventory with the three region merged together. The ABMI wetland class colour palette can be seen the ABMI Wetland Inventory metadata.

6. References

ABMI. (2016). *ABMI 3x7 Photoplot Land Cover Dataset Data Model*.

ABMI. (2018). *Human Footprint Inventory 2016*.

Atienza, R. (2018). *Advanced Deep Learning with Keras: Apply deep learning techniques, autoencoders, GANs, variational autoencoders, deep reinforcement learning, policy gradients, and more*: Packt Publishing Ltd.

Bayle, A., Carlson, B. Z., Thierion, V., Isenmann, M., & Choler, P. (2019). Improved mapping of mountain shrublands using the sentinel-2 red-edge band. *Remote sensing, 11*(23), 2807.

Böhner J, K. R., Conrad O, Gross J, Ringeler A, Selige T. (2002). Soil regionalisation by means of terrain analysis and process parameterisation. *EUROPEAN SOIL BUREAU �!RESEARCH REPORT NO. 7*.

Bruniquel, J., & Lopes, A. (1997). Multi-variate optimal speckle reduction in SAR imagery. *International Journal of Remote Sensing, 18*(3), 603-627.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific Model Development, 8*(7), 1991-2007.

DeLancey, E. R., Kariyeva, J., Bried, J., & Hird, J. (2019). Large-scale probabilistic identification of boreal peatlands using Google Earth Engine, open-access satellite data, and machine learning. *PLOS ONE, 16*(6).

DeLancey, E. R., Simms, J. F., Mahdianpari, M., Brisco, B., Mahoney, C., & Kariyeva, J. (2020). Comparing deep learning and shallow learning for large-scale wetland classification in Alberta, Canada. *Remote sensing, 12*(1), 2.

Dozat, T. (2016). *Incorporating nesterov momentum into adam*.

Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water resources research, 39*(12).

Gao, B.-C. (1995). *Normalized difference water index for remote sensing of vegetation liquid water from space*. Paper presented at the Imaging Spectrometry.

Gauthier, Y., Bernier, M., & Fortin, J.-P. (1998). Aspect and incidence angle sensitivity in ERS-1 SAR data. *International Journal of Remote Sensing, 19*(10), 2001-2006.

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and photobiology, 74*(1), 38-45.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*(1), 18-27.

Hatfield, J. L., & Prueger, J. H. (2010). Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. *Remote sensing, 2*(2), 562-578.

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., & Bonfil, D. (2010). *Assessment of leaf area index by the red-edge inflection point derived from VENµS bands*. Paper presented at the Proceedings of the ESA hyperspectral workshop, Frascati, Italy.

Huete, A., Liu, H., Batchily, K., & Van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote sensing of Environment, 59*(3), 440-451.

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425-1432.

Rouse Jr, J. W., Haas, R., Schell, J., & Deering, D. (1974). *Monitoring vegetation systems in the Great Plains with ERTS*.

Warner, B., & Rubec, C. (1997). The Canadian wetland classification system. *Wetlands Research Centre, University of Waterloo, Waterloo, Ontario.*

Weiss, A. (2001). *Topographic position and landforms analysis.* Paper presented at the Poster presentation, ESRI user conference, San Diego, CA.