2022/23 Wetland Inventory Program

Annual Report

Prepared for the Oil Sands Monitoring Program (OSM), June 2023

ABMI ALBERTA BIODIVERSITY MONITORING INSTITUTE

Table of Contents

Executive Summary

The accurate and timely detection of wetlands in Alberta's Oil Sands Region is crucial for effective and efficient monitoring, particularly given the ongoing natural and anthropogenic pressures exerted on wetlands in this portion of the province. Up-to-date knowledge of current conditions provide an important baseline for evaluating future changes to wetland extent, location, and condition or status over time.

For 2022/23, we had the following objectives:

1. Develop and test a ground-based field protocol that will support wetland mapping
2. Create and test a machine-learning approach to wetland classification
3. Create geospatial wetland data products for a pilot area

To achieve these objectives, we worked closely with Alberta Environment and Protected Areas and Ducks Unlimited Canada to develop, pilot, and evaluate a workflow for accurate wetland mapping.

Key outcomes of this project include:

- We completed the first version of a field protocol through extensive technical review with our collaborators and on the ground field-testing.
- We developed and tested two different empirical, data-driven approaches to wetland mapping – both achieved over 80% accuracy at the wetland class level. One approach was also tested at the wetland form level and achieved 64% overall accuracy, with variability in accuracy across wetland forms.
- Mapping to wetland form requires further development and testing. The overall accuracy of this product will benefit greatly from several additional sources of data:
- Large field reference dataset from across the oil sands region
- Complete coverage of LiDAR for the oil sands region and incorporation of LiDAR into the wetland classification workflow
- Additional training and validation datasets from updated 3 x 7 km air photo-interpreted sample areas

Below, we provide an in-depth description of the 2022/23 wetland inventory project work.

1.0 Background

In 2022/23, the Oil Sands Monitoring (OSM) program initiated the development of a wetland inventory to support accurate and timely detection of wetlands in Alberta's Oil Sands Region. This information is important, given the ongoing natural and anthropogenic pressures exerted on wetlands in this portion of the province. The ability to report on future changes requires knowledge of current conditions (i.e., a reliable baseline), enabling assessments of shifts in wetland area, size, extent, and status over time. Contemporary spatial inventories help to support legislation and policy development that are key to managing these sensitive ecosystems. Existing inventories (e.g., Alberta's Merged Wetland Inventory) can not only contain high rates of error, but are inconsistent in their vintages and methods, and are presently out-of-date. A reliable and updated baseline is crucial for future wetland monitoring in the Oil Sands Region.

Further to the need for a spatially consistent and timely wetland inventory to support the OSM program, a current provincial initiative, approved by Environment and Protected Areas, to develop methods and deliver updated wetland inventory products that follow Alberta's wetland inventory mapping standards is underway at present. Work undertaken during 2022/23 is an important contribution to this existing initiative and harmonizes with this work through the use of consistent approaches, datasets, and methods.

Objectives for the wetland inventory work in 2022/23 included:

1. Develop and test a field protocol to collect wetland inventory data important for the calibration and validation of wetland classification models based on remote sensing data;
2. Support the development and validation of a remote sensing data driven machine learning approach for wetland classification and mapping in the oil sands region;
3. Provide geospatial data products (wetland classification map), and a technical report (indicating summary of methods, results, data quality measures, and recommendations for future wetland classification efforts);
4. Provide documentation of final field validation protocols; and
5. Provide wetland data (Version 1) for further assessment and alignment with Environment and Protected Areas provincial-scale data.

The following document summarizes contributions for the 2022/23 year, and progress against the above objectives.

2.0 Summary of Contributions

This work was delivered collaboratively by the Alberta Biodiversity Monitoring Institute (ABMI) and Ducks Unlimited Canada (DUC), bringing together and leveraging expertise from both organizations. Appendix A lists the staff contributions from each organization. The authors acknowledge the financial support of the Oil Sands Monitoring Program.

3.0 Methods Summary

3.1 Study Area

The study area for this project was in the Alberta Oil Sands Region (*Figure 1*). The study area was chosen to align with ongoing wetland monitoring efforts being undertaken by the OSM Wetlands Surveillance Monitoring Program (i.e., its boundaries cover most wetland field sites and a diversity of wetland types that have been visited as part of this program). In this way, geographical overlap between the two types of OSM wetland efforts is aligned.

This work was completed in parallel with two other boreal forest pilot projects collaboratively by the ABMI and DUC, whereby the collective results will be used to inform future wetland mapping over Alberta's forested region.

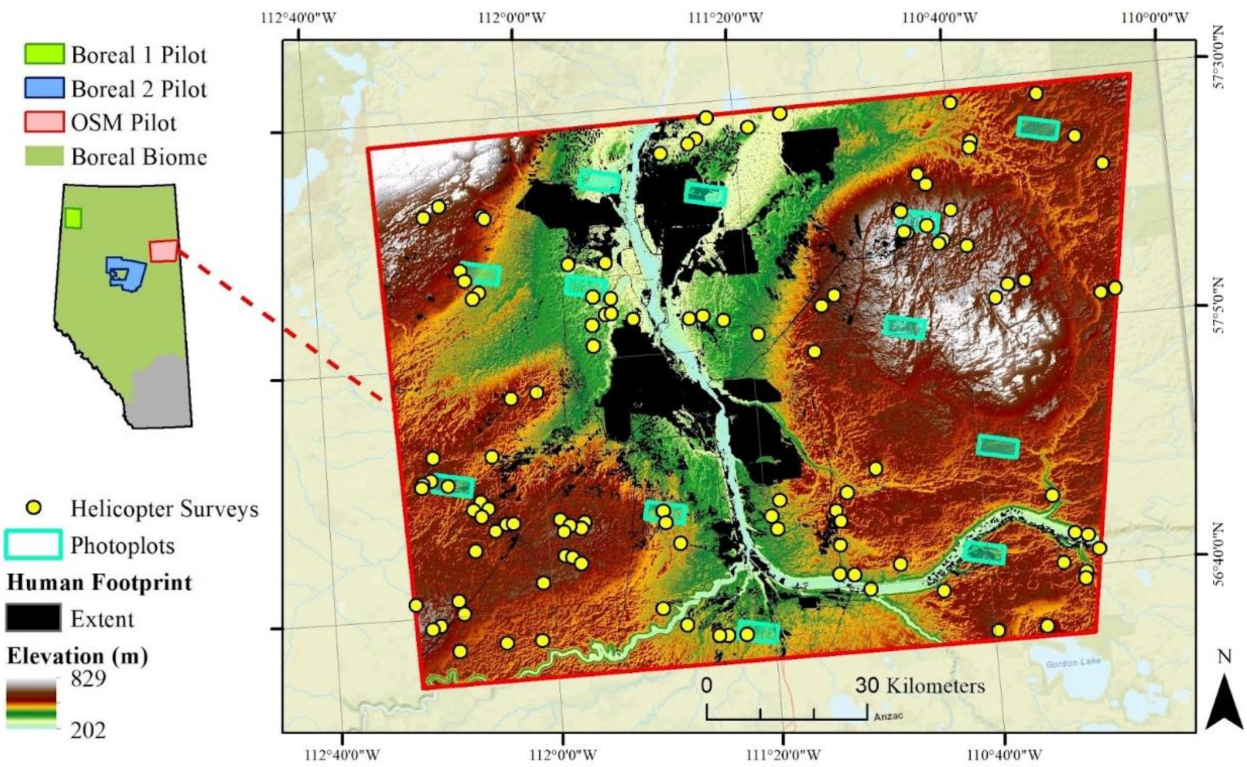

Figure 1: *Project area location. Reference datasets used for wetland modeling, including helicopter field surveys and photoplots, are displayed along with the extent of anthropogenic activity.*

3.2 Wetland Labels and Definitions

The primary goal of this project was to develop a wetland inventory that identifies and delineates both the wetland *class* and *form* of a given wetland, according to the Alberta Wetland Classification System (AWCS; ESRD, 2015). This includes five wetland classes (which are complementary to the Canadian Wetland Classification System; CWCS; NWWG, 1997) and 13 wetland forms (*Figure 2*). The AWCS was developed specifically for wetlands in Alberta and includes a suite of key indicators used to classify wetlands. Classification to the form detail is a requirement of the Government of Alberta's (GOA's) Wetland Mapping Standards and Guidelines (GOA-AEP, 2020) for the boreal/foothills zone.

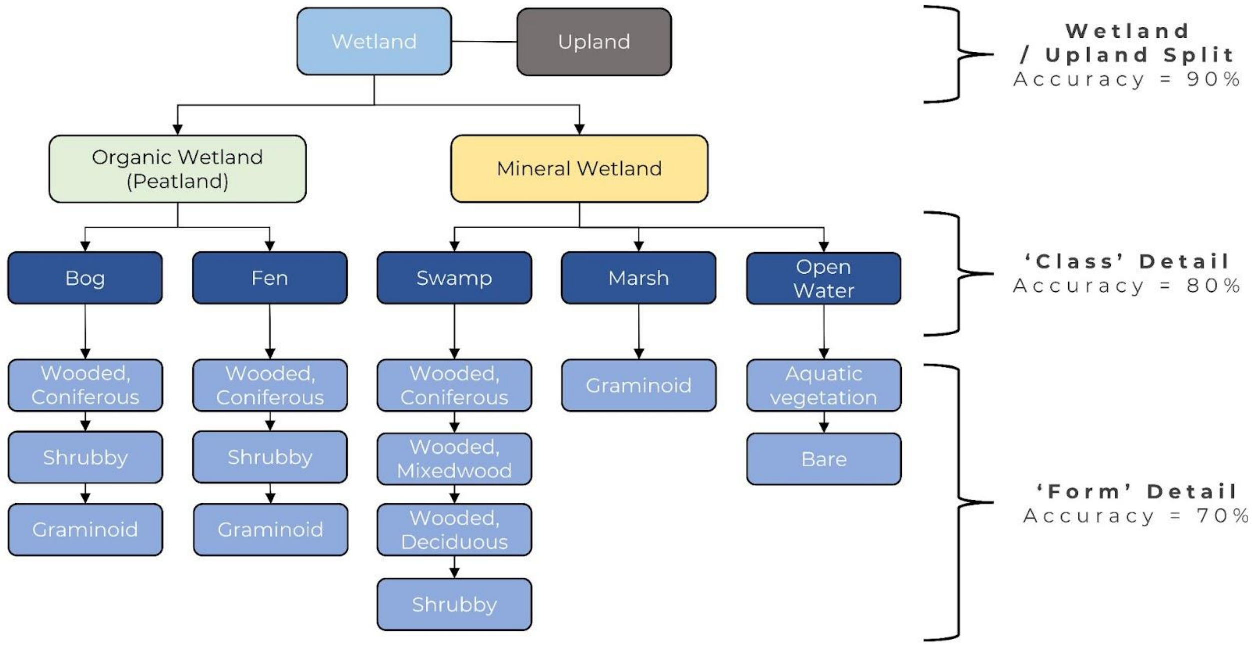

Figure 2: *Wetland classification schema adopted from the AWCS. The minimum overall accuracy requirements for each detail, as identified in the GOA's Wetland Mapping Standards and Guidelines, are indicated on the right.*

3.3 Training and Testing Reference Data Preparation

Model training data was acquired from the ABMI 3×7 km photoplot repository overlapping the project area (ABMI, 2016). The ABMI photoplots are detailed and comprehensive inventories characterizing moisture, management status, vegetation features, wetlands, land use, infrastructure, and land cover and cover approximately 5% of Alberta. However, these vector-based photoplots were created over a range of dates (e.g., some are now many years old), by various analysts, and thus required a comprehensive review and editing process before their use in deep and machine learning model training. DUC led this process, whereby photoplots and their land cover calls were assessed in reference to the AWCS wetland class and form definitions. Vegetation attribute cutoffs and hydrological cues (e.g., hydrodynamic regimes, moisture conditions, etc.) were used to correct any known errors in the photoplots. The photoplot attribute information and available high-resolution optical imagery were also used to digitize new polygon boundaries for small wetlands not originally captured by the photoplots as the GOA wetland mapping standards indicate a minimum mapping unit of 0.9 ha for the boreal/foothills zone. Any new polygons were digitized at a consistent scale range of 1:2000 to 1:5000. A total of 12 photoplots were strategically selected (i.e., with high wetland coverage, diversity, and general land cover uniqueness) across the OSM project area and corrected according to the AWCS wetland definitions.

Model testing data was obtained from DUC's historical collection of helicopter-based field sites (*Table 1*). These field sites were collected as polygons in vector format. All DUC field sites were reviewed for quality and consistency with a focus on thematic land cover class, disturbance (e.g., human impact), and boundary extent. Any edits to field site boundaries were done by digitizing at a scale of 1:2000 to 1:5000. A total of 115 field sites were available within the OSM project boundary and were used solely for model assessments. Evidently, this sample repository was limited in size; a dedicated field campaign would have greatly benefited the accuracy assessment process.

Table 1: *Summary of DUC historical helicopter survey sites.*

Class	Count	Form	Count
Shallow Open Water [W]	7	Submersed and/or floating aquatic vegetation [A]	3
		Bare [B]	4
Marsh [M]	3	Graminoid [G]	3
Fen [F]	36	Wooded, coniferous [Wc]	21
		Shrubby [S]	4
		Graminoid [G]	11
Bog [B]	16	Wooded, coniferous [Wc]	11
		Shrubby [S]	5
		Graminoid [G]	0
Swamp [S]	13	Wooded, coniferous [Wc]	3
		Wooded, mixedwood [Wm]	0
		Wooded, deciduous [Wd]	4
		Shrubby [S]	6
Upland	40	-	-

3.3.1 Field Protocol Development to Inform Future Validation Efforts

Field protocols were developed that describe collection procedures to produce validation data supporting the calibration and validation of Earth Observation (EO) satellite-based wetland mapping to the AWCS class and form level at a minimum. The primary purpose of these protocols is to collect critical physical and biological data that efficiently and effectively enable the accurate ground-based classification of a given wetland to class, form, and type levels. The protocols follow the AWCS document (GOA:ESRD 2015), the Alberta Wetland Classification Field Guide (DUC 2021) and the Alberta Wetland Mapping Standards and Guidelines (GOA: AEP 2020a). The protocol was developed by ABMI with four rounds of technical review and input from DUC and Alberta Environment and Protected Areas (EPA). *Table 2* summarizes key milestones in protocol development.

Table 2: *Important milestones reached during the development of a field protocol for ground validation data collection in support of wetland inventory mapping.*

Date	Draft	Details
July 28, 2022	1	Initial draft completed and sent to EPA for review
August 29, 2022	2	Revisions completed based on review and discussions with EPA and DUC. Revisions included restructuring to focus on three main data elements: vegetation, soil, and water. Draft completed and sent to EPA for review.
October 7, 2022	3	Revisions completed based on review and discussions with EPA and DUC. Revisions included updates on site selection process, tree and shrub definitions, boundary delineation approach, site layout, equipment list and disturbance observations. Draft completed and used for field testing.
January 20,2023	4	Revisions completed based on changes identified in field testing. Draft completed and sent to EPA for review.
January 30, 2023	5	Revisions completed based on EPA review. Draft completed.

Field testing was completed on October 11-12, 2022, by ABMI near Donalda, Alberta. During testing, the protocol was completed at 24 wetland sites in two quarter sections (NE27-41-19-W4 and NW18-40-19-W4). The objective of the field testing was to assess the amount of time the protocol took to complete in the field and identify any challenges implementing the protocol. A short description of select challenges identified and subsequent updates to the protocol document are described below.

- The vegetation indicator community descriptions on the field sheet were not adequate to describe observed wetlands. The protocol and field sheet were revised so that each characteristic species encountered was documented and indicator species from cropland wetlands were added from Stewart and Kantrud (1971).
- The field testing identified the need to be able to capture mucky mineral soil types. The mucky mineral soil category was added to the field sheet and detail on how to distinguish this soil type in the field was added to the SOP based on the USDA hydric soil field guide (USDA 2018).

The final field protocol involves the collection of information on the following:

- Vegetation (e.g., height, cover, indicator species presence);
- Soil peat presence, depth, and decomposition, along with observations of other wetland soil indicators (e.g., gleying);
- Water physiochemistry and hydrological indicators (e.g., pH, electrical conductivity);
- Locational information on wetland class, form, or type boundaries;
- An AWCS form or type-level call;
- A set of photos; and
- Observations of human footprint and/or fire disturbance.

The field protocol document (i.e., standard operating procedures) is provided in *Appendix B*.

3.4 LiDAR Data Procurement and Processing

LiDAR datasets were acquired through the Forest Resource Improvement Association of Alberta from Alberta-Pacific Forest Industries and covered 74% of the OSM study area (*Table 3*). ABMI acquired additional LiDAR data covering another 26.7% of the study area (*Table 3*) and is anticipated to acquire the remaining coverage during 2023 (*Table 2*). The latter gap is located largely on the mineable region of the study area. Members of Canada's Oil Sands Innovation Alliance were contacted by way of the OSM Program's Industry Data Sharing Request Form for LiDAR datasets covering the mineable portion of the OSM study area (11.4 %), but while two meetings were held, no data was supplied.

Table 3: *Summary of LiDAR dataset coverage of the OSM study area.*

Source	Area (km^2)	Percent of Study Area
Alberta-Pacific Forest Industries	9,445	74.7%
ABMI 2020 acquisition	3,337	26.7%
ABMI 2023 acquisition (planned)	1,445	11.4%

To utilize LiDAR in our model pipeline it first required processing. Briefly, the steps included in processing LiDAR data include:

1. Assembly of all LiDAR sources and confirmation that full coverage of a study area exists (note: as described, full coverage of the OSM study area is not yet available).

2. For each LiDAR source (e.g., ABMI 2022 acquisition):
    a. generation of a Digital Surface Model (DSM) at 10m resolution;
    b. generation of a bare earth Digital Elevation Model (DEM) at 10m resolution; and
    c. creation of a Canopy Height Model (CHM) by subtracting the DEM from the DSM.
3. Reprojection and mosaicking of the DEM and CHM outputs into common a projection and extent such that continuous coverage over the study area is produced.

With a new LiDAR-based DEM, it is possible to generate a series of topographic layers to replace the satellite-based topographic layers. In addition, the CHM provides a new layer that is not currently used in the modeling approaches described here. Furthermore, it is possible that other LiDAR-based derivatives could also be created (such as point densities by height), and which follow very similar creation procedures as those used in generating the DEM.

As full coverage of the OSM study area by LiDAR datasets was not available at the time of the described work, LiDAR inputs were not included in wetland class or form model training and prediction across the study area.

3.5 Satellite EO Data Collection and Processing using Cloud-Computing Resources

Multi-temporal and multi-source Earth Observation (EO) data was collected from the Google Earth Engine (GEE) cloud-computing platform following that of Delancey et al. (2020). Although, in this project, we also utilized multi-seasonal image compositing methods (e.g., spring and summer). This is because Merchant et al. (2020) established that wetlands are mapped more accurately using multi-temporal data, which captures the dynamic ecohydrological characteristics of wetlands. Image sources collected from the GEE included Sentinel-2 optical imagery, Sentinel-1 synthetic aperture Radar (SAR) imagery, and Advanced Land Observing Satellite (ALOS) Digital Surface Model (DSM) topographic data (Table 4). All GEE data collection was performed using JavaScript coding.

Table 4: *List of EO data sources used for wetland mapping.*

Image Source	Image Type	Season	Processed Image Count	Native Spatial Resolution
Sentinel-1	SAR	Spring	90	10 m
		Summer	93	
Sentinel-2	Optical	Spring	176	10 m
		Summer	248	
ALOS DSM	Topographic	-	1	30 m

For Sentinel-2, Level-2A surface reflectance (i.e., bottom of atmosphere) data were used, which were atmospherically corrected using the European Space Agencies (ESA) Sen2Cor algorithm. Images were queried for two seasonal periods, June to July (i.e., early season composite) and August to September (i.e., late season composite), and for the years 2020-2022. By using the quality assessment (QA60) band, we selected only images with a cloudy pixel percentage of less than 20%, and then subsequently performed cloud masking. Sentinel-2 reflectance composites were then created by calculating the index median from the time-series stack, although only for bands collected at 10 or 20-m resolution (i.e., 60-m bands were not used). Temporal aggregation through median metrics is an advantageous and popular method for gap filling cloud-masked areas (Carrasco et al., 2019).

Sentinel-1 is a two-satellite constellation mission (Sentinel-1A and 1B) that collects C-band (5.6 cm wavelength, 5.405 GHz) SAR data from the microwave portion of the electromagnetic spectrum. Sentinel-1 can operate in several acquisition modes; for this project we selected the Level-1 Interferometric Wide Swath (IW) Ground Range Detected (GRD) mode, which leverages a dual-polarization SAR sensor with 12 or 6-day repeat imaging (Potin et al., 2012). Polarization channels included: (1) vertically transmitted and vertically received (VV), and (2) vertically transmitted and horizontally received (VH). The IW mode collects data with spatial resolution (range × azimuth) at 20 × 22 and pixel spacing (range × azimuth) at 10 × 10. Sentinel-1 images were preprocessed within GEE using the Sentinel-1 toolbox, and included thermal noise removal, radiometric calibration, and terrain correction. Sentinel-1 images were collected for the same time period and seasons as the Sentinel-2 data, and then similarly underwent a median pixel function.

Height above sea level elevation data was acquired from the ALOS DSM. For this, elevation measurements were converted from the ellipsoidal height based on ITRF97 and GRS80 in GEE, using the EGM96†1 geoid model. The topographic data was resampled to 10 m spatial resolution to match the Sentinel-1 and –2 image sources.

3.5.1 EO Features

Numerous bands and mathematical indices were extracted from the multi-source and multi-season satellite data (45 in total) to support our artificial intelligence- (AI-)driven modeling. These EO features, which largely follow Delancey et al. (2020) and Merchant et al. (2023) are listed in *Table 5*.

Table 5: *List of EO features (i.e., variables) processed and considered for wetland classification modeling.*

Feature Source	Features	Season
SAR	VV, VH, VV/VH, Radar Vegetation Index (RVI), Span	Spring and summer
Optical	Blue, Green, Red, Red Edge (RE1, RE2, RE3), Near Infrared (NIR), Narrow Near Infrared (NNIR), Short wave Infrared (SWIR1, SWIR2), Enhanced Vegetation Index (EVI), Normalized Burn Ratio (NBR), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Soil Adjusted Vegetation Index (SAVI)	Spring and summer
Topographic	Elevation, Topographic Position Index (TPI), Topographic Wetness Index (TWI), Multi Resolution Index of Valley Bottom Flatness (VBF), Topographic Roughness Index (TRI)	-

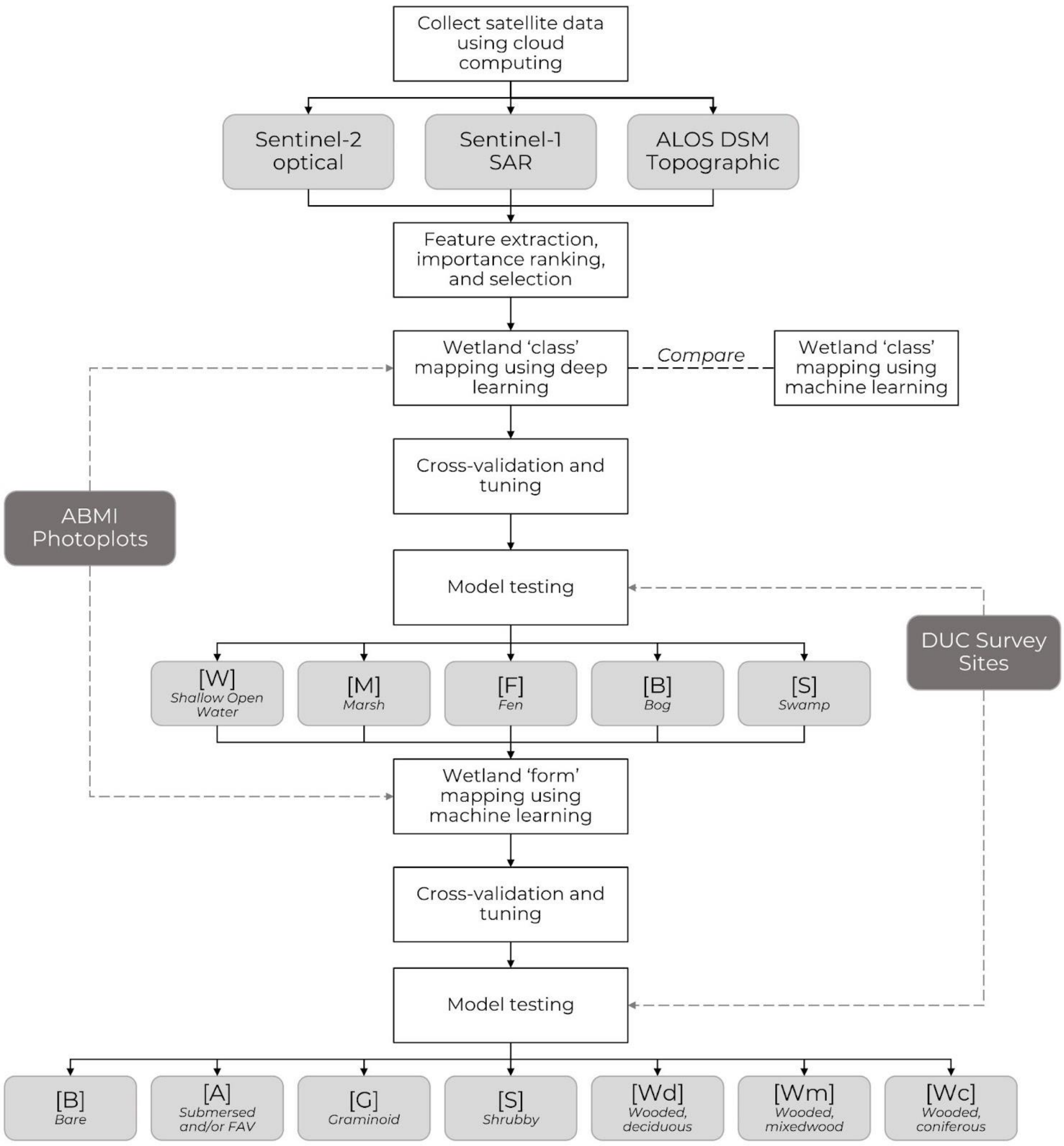

Figure 3: *Wetland mapping workflow applied in this project.*

3.6 *Wetland Classification Modeling using Artificial Intelligence*

EO data classification was completed using both deep (DL) and machine learning (ML) methods. The overall mapping workflow proposed in this project can be found in *Figure 3*. DL was first used for wetland class mapping because these AI models typically outperform traditional ML models on unstructured data, however they require much more training data to do so. It is for this reason that they generally perform well with fewer prediction labels, but also why ML was then used for wetland form mapping (i.e., comprising more labels and thus less training data per label). Nevertheless, we still tested both AI methods (DL and ML) for class detail mapping. The hybridized approach used in this study, which utilized both forms of AI by integrating the classified products post-modeling, is described in the following sections.

3.6.1 Deep Learning Modeling for Wetland Class Detail

Wetland class definitions, as identified by the AWCS (i.e., open water, marsh, fen, bog, and swamp), were first mapped using the DL Convolutional Neural Network (CNN) methods developed by Delancey et al. (2020). A CNN approach was chosen for AWCS class detail mapping because CNNs perform well at exploiting spatial context due to filtering and making accurate generalizations (Kattenborn et al., 2021). CNN modeling was implemented in the Python programming language using the Keras deep learning library. The CNN framework was based on a U-Net architecture originally developed by Ronneberger et al. (2015). U-Net is an encoder-decoder that performs image segmentation, by producing an output with similar spatial dimensions as the input. The U-Net models developed in this project used the same general architecture as in Delancey et al. (2020).

With our DL modeling, two CNNs were trained using slightly different input layers and training regimes (Table 6), with the resultant class likelihood maps then averaged to produce a final result. Initially, we performed an experiment that involved training a CNN for each available input layer with the idea of selecting the top-N layers based on model performance. The performance for the models trained using single layers was low and noisy, making feature selection in this manner somewhat problematic. Using the layers that were discovered in the ML feature selection process comprised a starting point. Since the feature selection process evolved, we ended up with two slightly different sets of input layers from different times during the process. Some layers were subsequently eliminated. It was noticed that the models benefitted a small amount by removing the “late” version when both the “early” and “late” were initially used. Both CNNs used a square input window (i.e., patch) of size 192 pixels, were trained for 90 epochs with batch size of 16, had an initial learning rate of 0.003, employed an artificial over-sampling of tiles containing ‘bog’, and used a loss function being the sum of cross entropy loss and dice loss.

The differences between the two models were the input layers (outlined below) and the learning rate schedule. Model 1 had its learning rate halved every 20 epochs and Model 2 had its learning rate halved every 40 epochs. Training multiple models in different ways followed by results averaging is a common approach in AI modeling. Selecting which models to use was based on two main factors: the first being cross-validation scores that were competitive and the second being the stability of the model. The cross-validation metrics were calculated by applying a newly trained model to two ABMI 3x7 photoplots that were not included in the data used to train or monitor the model (during training). The same holdout test plots were used throughout the model building process.

The stability of the model was gauged by comparing the performance of the fully trained model (at epoch 90) to a version of the model trained at 60 epochs. It was observed that for many models trained there was some fluctuation in the training and validation performance throughout the epochs, which indicates that either the model is having difficulty converging on a solution or that it is still exploring its weight-space. Selecting models with a greater amount of stability throughout its later epochs gives us more confidence in its reliability. This process was performed 'by inspection', and further experimentation would be required to fully assess the relationship between stability and overall performance. Both CNN models were trained using data gathered from the 'Boreal 1', 'Boreal 2', and 'OSM' pilot regions.

Table 6: *Input layers used in the two CNN models developed for class mapping. Note that S2 indicates Sentinel-2 optical derivatives, S1 indicates Sentinel-1 radar derivatives, and ALOS indicates topographic derivatives. See Table 3 for index acronym definitions.*

Model 1	**Model 2**
S2 Late BLUE S2 Early NDVI S2 Late NDVI S2 Early NDWI S2 Early NIR S2 Early RED S2 Late RED S2 Early SWIR1 S2 Late SWIR2 S2 Early RE1	S2 Late BLUE S2 Early GREEN S2 Early NDVI S2 Early NDWI S2 Early NIR S2 Late RED S2 Early SWIR1 S2 Early RE1
DEM ALOS SWI ALOS TPI750 ALOS TRI ALOS VBF ALOS	DEM ALOS SWI ALOS TPI750 ALOS TRI ALOS VBF ALOS
S1 Early Span S1 Late Span S1 Early VH S1 Late VH S1 Late VV	S1 Late RVI S1 Early Span S1 Early VH S1 Late VV

3.6.2 Machine Learning Modeling for Wetland Form Detail

Following class modeling, the AWCS wetland form definitions were then mapped using ML. An Extreme Gradient Boosting (XGBoost) algorithm was chosen for this (Chen & Guestrin, 2016). XGBoost has attracted much attention in data science, computer vision, and remote sensing due to its state-of-the-art results compared to other benchmark ML algorithms such as Random Forests (RF) and Support Vector Machines (Jafarzadeh et al., 2021). XGBoost is part of a group of ensemble learning methods which implements a Gradient Boosting algorithm, whereby an ensemble of decision trees, which are considered "weak learners", are then grouped together to create a "strong learner" (*Figure 4*). This process operates sequentially (i.e., rather than in parallel, like a bagging algorithm), whereby the algorithm continuously corrects the previous weak learning trees until a stopping condition is met and a final strong learner is achieved.

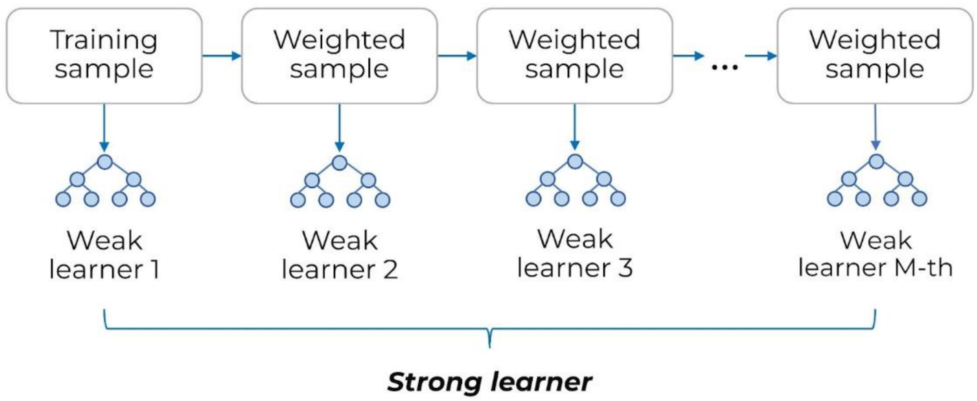

Figure 4: *Schematic diagram of a boosting, sequential XGBoost algorithm.*

We employed XGBoost modeling using the Classification and Regression Training (caret) package available in R Statistical software (Kuhn, 2008). To achieve form mapping, we randomly sampled the updated ABMI 3×7 photoplot polygons, with samples labeled based on their plant community/vegetation structure or basin characterization captured at the AWCS form detail (e.g., bare, SAAV, graminoid, shrubby, wooded deciduous, wooded coniferous, or wooded mixedwood). Samples were created by randomly generating 1000 points per form, resulting in a total of 7000. A balanced sampling design was chosen as this often produces more precise estimates than simple random sampling (Stevens & Jensen, 2007). This random sampling procedure was then repeated four times, resulting in four unique subsets of form samples. This allowed us to develop four separate XGBoost models. The predictions from each XGBoost model were summarized into a final classification by taking the modal (i.e., majority) form value. Once this modal layer was obtained, the final wetland form label for each pixel was determined by intersecting (i.e., integrating) the form with the class predictions. In other words, the wetland class outputs were used as image objects for spatially constraining and then determining the final wetland form. For example, if the predicted wetland class was bog and the form was shrubby, then the final label would be shrubby bog.

All XGBoost models were independently optimized by tuning the important hyperparameters through an exhaustive grid search method. The hyperparameters considered for tuning can be found in Table 7. The final values for each hyperparameter were selected based on a statistical *k*-fold cross-validation assessment, which is described in more detail later.

Table 7: *Overview of XGBoost hyperparameters tuned.*

Hyperparameter	Description	Values	Tuning Method
max_depth	Maximum tree depth	10, 20, 30	
nrounds	Number of trees	100, 300, 500	
eta	Learning rate	0.1, 0.3, 0.5	
subsample	Row sampling	0.5, 0.75, 1.0	Cross-validation grid search
gamma	Used for tuning of Regularization	0, 0.5, 1	
min_child_weight	Minimum leaf weight	1, 2, 3	
colsample_bytree	Column sampling	0.6, 0.8, 1	

We also evaluated different post-processing techniques for the effect on form mapping accuracies. This included the following: 1) no post-processing; 2) smoothing using a conventional 5×5 pixel majority filter; and 3) smoothing by overlaying an object-based segmentation produced externally in GEE with the Simple Non-Iterative Clustering (SNIC; Tassi & Vizzari, 2020) algorithm. Based on our preliminary experiments, we found that no post-processing left too many spurious and noisy pixels, while the GEE SNIC algorithm produced unsatisfactory image objects and even hindered classification accuracy. As such, the majority filter function was chosen for post processing, as this produced a wetland map closely resembling the ecological patterns of wetlands.

3.6.3 Feature Selection

Implementing a practical feature selection (FS) methodology is critical for effectively removing redundant and/or irrelevant variables, improving computational efficiencies, and in many instances boosting classification accuracies. Moreover, the importance of FS is emphasized when using multi-dimensional, high-resolution EO datasets, and in particular when working with computationally demanding AI algorithms (Maxwell et al., 2018). To address this, we computed relative variable importance (VI) for both class and then form wetland modeling based on the Mean Decrease in Accuracy (MDA) algorithm generated from XGBoost models. Thus, two sets of rankings were generated from this procedure. VI is determined by whether a variable was selected to split during the tree construction process, and how much the error changed the result. The higher the value of MDA, the more important the variable is to modeling. Once VI was computed for each classification detail, the variables were ranked and the top 20 were selected as model inputs.

3.6.4 Model Cross-Validation and Accuracy Assessment

The partitioning of reference data used for AI model development (both DL and ML), in particular for training and validation, is incredibly important as it influences model architecture, the selection of hyperparameters, and model performance (Lyons et al., 2018). Cross-validation methods, which are commonly applied in applied AI, are used to estimate model performance on unseen data.

As outlined in section 3.6.1, two AMBI 3×7 photoplots were held out to evaluate each model once it was trained. These plots contained all wetland classes and since they remained constant, comparisons between models were possible. To monitor the performance of a model *during* training, another small set of four ABMI 3×7 photoplots were predicted after each training epoch in order to evaluate the progress of the model's training. It is well-known that a model will likely perform better on samples taken in close proximity to the samples used to train the model, a concept often referred to as 'spatial auto-correlation'. This makes the decision to use separate ABMI 3×7 photoplots to monitor training and to evaluate model performance more robust.

For XGBoost ML models, we used the *k*-fold statistical validation method for estimating generalization error and tuning hyperparameters (Kohavi, 1995). With *k*-fold cross-validation, the training dataset is randomly divided into *k* smaller groups and the ML model is trained using *k*-1 of the folds; the ensuing ML model is validated with the remaining data. Accuracy from the *k*-fold cross-validation run is measured by taking the mean of the model scores in the loop. All XGBoost models were tuned and validated using 10-fold cross-validation.

After finalizing DL and ML model tuning, all output classifications were tested with several accuracy metrics. These metrics were calculated using the unseen DUC survey sites (i.e., the reference sites not used in model training or validation). Accuracy metrics for testing evaluations included overall accuracy and per-class recall, precision, F1, and intersection over union (IOU):

$$Overall\ accuracy = \frac{Correctly\ classified\ pixels}{Total\ pixels}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$IOU = \frac{Area\ of\ overlap}{Area\ of\ union}$$

where TP is True Positive, FP is false positive, and IOU represents the ratio between the intersection of all positive predictions and all ground truth predictions (Maxwell et al., 2021). F1, which represents the harmonic mean of precision and recall, and IOU were the primary per-class metrics reported.

4.0 Summary of Key Results

4.1 Feature Importance

Figure 5 shows the normalized (0 to 1) Mean Decrease in Accuracy (MDA) values and therefore relative variable importance (VI) for the top 20 variables, for both class and form modeling. These rankings suggest that topographic information is very important for boreal wetland classification, as many of the ALOS DSM derivatives appear high on each list (i.e., further to the left on the x-axis). Elevation, for example, was by far the most important predictor for both classification details.

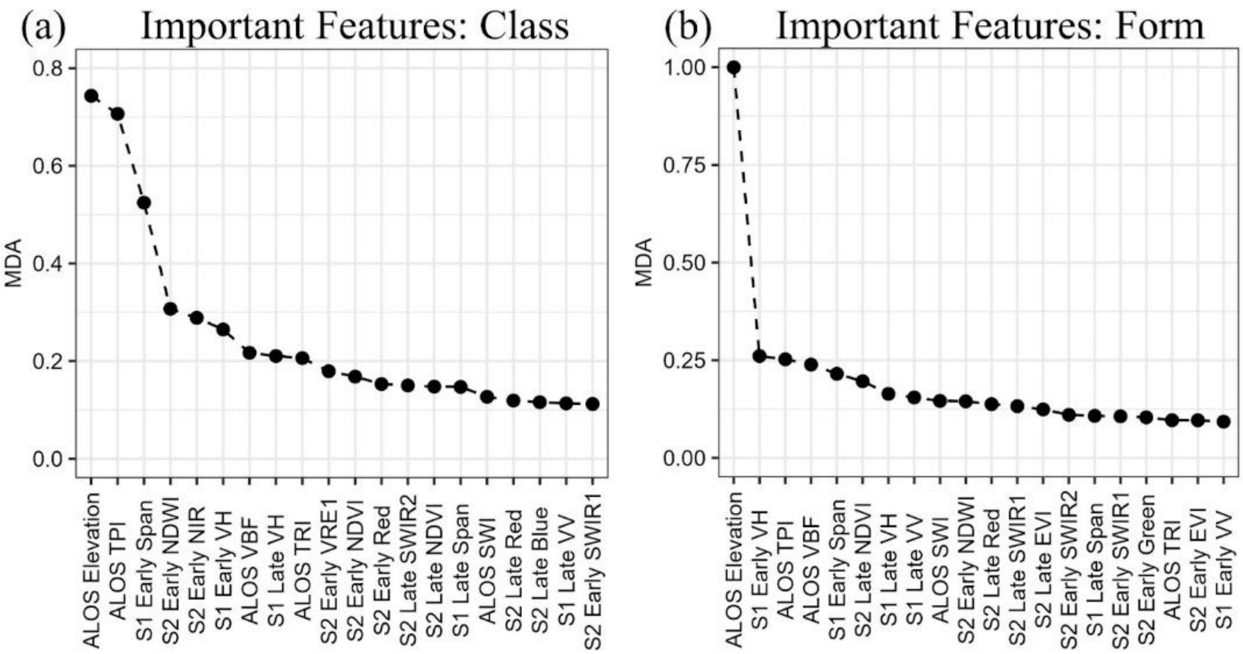

Figure 5: *Variable importance (VI) for the top 20 variables for (a) class and (b) form modeling. VI was computed using the Mean Decrease in Accuracy (MDA) algorithm. Values were scaled from 0 - 1. Variables are arranged in descending order of importance (left to right).*

4.2 Wetland Modeling Accuracies

Table 8 shows the overall map accuracies achieved from both ML and DL. Both methods achieved the minimum overall accuracy target (i.e., 80%) at the class detail Government of Alberta's (GOA's) Wetland Mapping Standards and Guidelines (GOA-AEP, 2020). However, the XGBoost model outperformed the U-Net model by 6%. Therefore, the XGBoost class classification was advanced for spatial integration with the form classification. Due to limited time, only a ML approach was tested for form mapping, as this has proven to work well in previous boreal mapping projects. After integrating the class detail map with structural predictions from the XGBoost modeling, the resulting form classification was evaluated to have 64% overall accuracy. This result, which does not meet the minimum standard of 70% for the boreal/foothills zone according to the Government of Alberta's (GOA's) Wetland Mapping Standards and Guidelines (GOA-AEP, 2020), is discussed further in the recommended next steps.

Table 8: *Overall map accuracies achieved by each AI model.*

Model	Detail	Overall Accuracy
U-Net	Class	83%
	Form	-
XGBoost	Class	89%
	Form	64%

Table 9: *Per-class accuracy metrics achieved by each AI model.*

	U-Net		XGBoost			XGBoost	
Class	**F1**	**IOU**	**F1**	**IOU**	**Form**	**F1**	**IOU**
Shallow Open Water	0.86	0.75	0.98	0.97	Bare	0.04	0.02
					Submersed and/or floating aquatic vegetation	0.69	0.53
Marsh	0.13	0.06	0.21	0.12	Graminoid	0.21	0.12
Fen	0.73	0.57	0.87	0.76	Graminoid	0.39	0.24
					Shrubby	0.02	0.01
					Wooded, coniferous	0.49	0.32
Bog	0.49	0.32	0.85	0.79	Graminoid	-	-
					Shrubby	0	0
					Wooded, coniferous	0.82	0.69
Swamp	0.44	0.29	0.77	0.23	Shrubby	0.31	0.18
					Wooded, coniferous	0.03	0.02

	U-Net		XGBoost			XGBoost	
					Wooded, deciduous	0.12	0.06
					Wooded, mixedwood	-	-
Upland	0.97	0.94	0.87	0.86	Upland	0.92	0.86

Table 9 provides a breakdown of per-class and per-form accuracy metric results. It is evident that accuracies varied considerably both between, and within classes. In addition, while similar trends in relative accuracies between classes are consistent between the U-Net and XGBoost models, the latter produced higher class accuracies. All classes produced by this model, except marsh, produced an F1 score above 0.75. At the form level, however, only open water with submersed and/or floating aquatic vegetation, wooded coniferous bog, and upland produced comparable F1 scores (*Table 9*).

4.3 Wetland Mapping Outputs

The final wetland maps of the OSM project area are presented in *Figure 6* (class) and *Figure 7* (form). Fen was found to be the most prevalent wetland class, occupying 21% of the project area. This was followed by swamp (19%), bog (10%), marsh (4%), and shallow open water (2%).

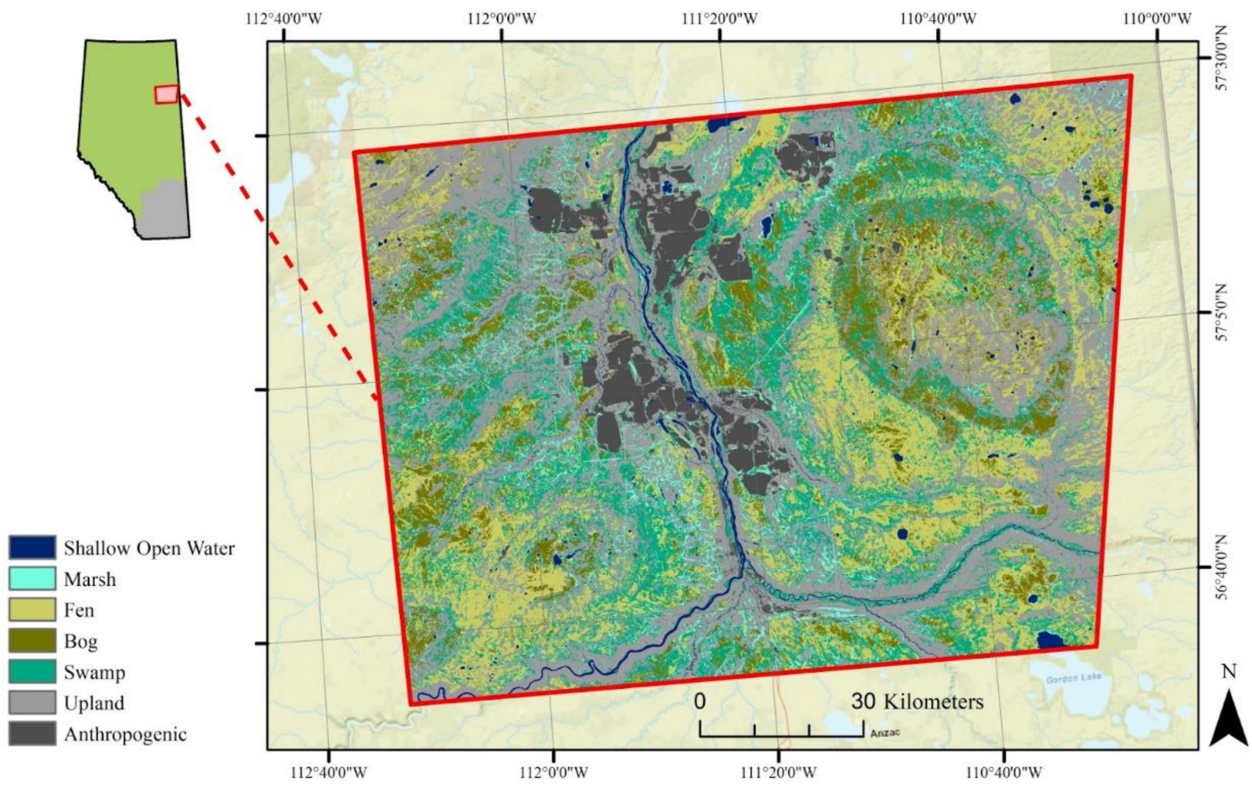

Figure 6: *Wetland class map of the OSM project area. Overall accuracy = 89%.*

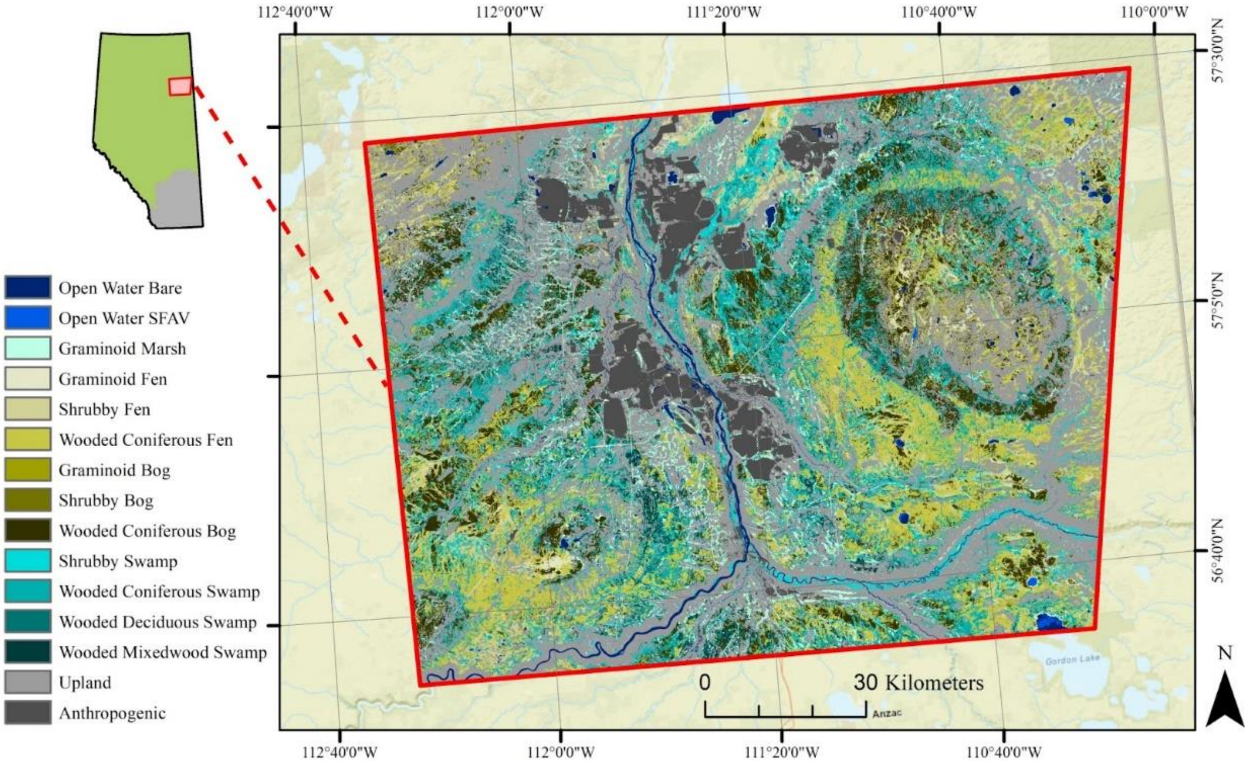

Figure 7: *Wetland form map of the OSM project area. Overall accuracy = 64%.*

5.0 Recommended Next Steps

In this project, we demonstrated an AI-driven approach to modeling the spatial extent of boreal wetlands. The implemented methodology builds on earlier collaborative work by DUC and ABMI. The preliminary findings from this project were largely promising, yet unexpected, warranting further experimentation and analysis. However, limitations in time, remote sensing data, and field validated samples presented challenges in exploring important questions related to these results.

Field-validated data of sufficient size is a critical component in AI modeling, especially when making predictions about complex natural systems like wetlands. We highly recommend that any future wetland mapping updates across Alberta's boreal biome dedicates appropriate resources to rigorously collecting a large, detailed, and representative field reference dataset. Such a dataset would permit a more statistically sound and defendable evaluation of any wetland classification generated. In the current project, we only had available a select number of historical survey sites that sporadically intersected the OSM region. Moreover, these were collected for other mapping projects which used different remote sensing methods, technologies, and sampling designs. Due to its sporadic extent and limited size, this sample repository lacks the ability to provide a thorough and full statistical assessment of our wetland maps. This was especially apparent for wetland classes containing few survey sites, such as marsh (3 sites total), or many of the wetland forms (e.g., wooded coniferous swamp). Hence, a dedicated and well-planned sampling design and accuracy assessment program is needed for each wetland inventory map produced over Alberta. Nevertheless, we can assume a higher confidence in the reported class detail overall accuracy (89%) than the form detail overall accuracy (64%) is due to the number of samples per thematic label. For example, fens, bogs, and uplands were represented by the greatest numbers of samples, and their reported F1-scores were all > 0.85. A user of this map can have higher confidence in those reported accuracies in comparison to other classes or forms evaluated with fewer samples (e.g., marsh).

It is also imperative to discuss the absence of LiDAR imagery in our modeling. Initially, this project sought to incorporate LiDAR imagery into the wetland mapping workflow by substituting the moderate resolution ALOS DSM data with it. We hypothesized that LiDAR would elevate our classification accuracies as a result of its higher spatial resolution and ability to preserve important wetland features that manifest from topographic variations. Despite sufficient LiDAR data not becoming available during this project, we nevertheless still achieved encouraging accuracies at the class detail using both ML (89%) and DL (83%). However, it was at the form detail where the LiDAR's omission was most impactful. Without LiDAR, we were unable to process and analyze valuable wetland predictors, such as vegetation canopy height and density metrics, and bare earth topographic layers. These terrain variables have shown to be foundational in achieving wetland classification targets in several of our organizations' previous mapping initiatives. Therefore, we strongly recommend that future wetland mapping projects incorporate and evaluate the efficacy of LiDAR model inputs.

There are several remaining questions and experimental findings from this project that require further exploration. For example, an unexpected outcome from this work was that the XGBoost methodology produced both a quantitatively and qualitatively better class detail map than the U-Net approach. It is well known that DL algorithms are complex and sensitive to their parameterization. As such, additional time would have permitted further fine-tuning and experimentation with the U-Net model, which holds immense potential for class detail mapping. We would have also liked to assess the U-Net's form modeling capabilities. Moreover, restrictions in time and budget meant that only a select number of ABMI 3×7 photoplots were edited and used for model training and validation stages. Enhancement of additional photoplots would have increased the learning capacity of both the DL and ML algorithms, increased the sampling of more rare wetland classes and forms, and improved the generalization capabilities of our models. Another question left rather unexplored from this project is the impact of sampling design on mapping accuracies. For instance, our ML modeling employed a balanced sampling design for both classification details. Experimenting with different sampling approaches, such as simple random or proportionate allocation, may have yielded improved results, particularly at the form detail where the spatial extent of wetland forms varies immensely. Moreover, adjusting the number of samples collected per label, and the number of AI models included in the ensemble for final predictions (e.g., our U-Net used the modal prediction from two models, and XGBoost used four models), may have also altered classification accuracies.

With regard to the field protocol developed as part of this work, further refinement in collaboration with EPA and DUC is anticipated. An area of possible refinement is further alignment with the Alberta wetland identification and delineation directive (GOA 2015). It is recommended that ongoing reviews of the protocol document focus on continuing to evaluate the soil information and detail required for the purposes of satellite-based wetland inventory validation. It could also be beneficial to address whether adjustments might be needed to better accommodate the protocol's application within the boreal transition zone of the province, where wetlands may not easily fall into a boreal-type or prairie pothole-type categorization. It is also recommended that future work involves the use of the developed protocol to rigorously collect critical ground validation data to support improved wetland inventory mapping efforts.

Abbreviations

AI	Artificial Intelligence
ALOS	Advanced Land Observing Satellite
AWCS	Alberta Wetland Classification System
CHM	Canopy Height Model
CNN	Convolutional Neural Network
CWCS	Canadian Wetland Classification System
DL	Deep Learning
DEM	Digital Elevation Model
DSM	Digital Surface Model
EO	Earth Observation
FP	False Positive
FS	Feature Selection
GEE	Google Earth Engine
GOA	Government of Alberta
MDA	Mean Decrease in Accuracy
ML	Machine Learning
RF	Random Forest
SAR	Synthetic Aperture Radar
SNIC	Simple Non-Iterative Clustering
TP	True Positive
VI	Variable Importance
XGBoost	Extreme Gradient Boosting

References

ABMI. (2016). ABMI 3x7 Photoplot Land Cover Dataset Data Model. Edmonton, AB, Canada.

Alberta Environment and Sustainable Resource Development (ESRD). (2015). Alberta Wetland Classification System. Water Policy Branch, Policy and Planning Division, Edmonton, AB.

Carrasco, L., O’Neil, A. W., Morton, R. D., & Rowland, C. S. (2019). Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. Remote Sensing, 11(3), 288.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

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.

Ducks Unlimited Canada (DUC). (2021). Alberta Wetland Classification System Field Guide. Edmonton, Alberta.

Government of Alberta. (2015). Alberta Wetland Identification and Delineation Directive. Water Policy Branch, Alberta Environment and Parks. Edmonton, Alberta.

Government of Alberta (GOA) – Alberta Environment and Parks (AEP). (2020). Alberta Wetland Mapping Standards and Guidelines: Mapping Wetlands at an Inventory Scale v1.0. Edmonton, Alberta.

Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., & Homayouni, S. (2021). Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. Remote Sensing, 13(21), 4405.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS journal of photogrammetry and remote sensing, 173, 24-49.

Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

Kuhn, M. (2008). Building predictive models in R using the caret package. Journal of statistical software, 28, 1-26.

Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J., & Elith, J. (2018). A comparison of resampling methods for remote sensing classification and accuracy assessment. Remote Sensing of Environment, 208, 145-153.

Maxwell, A. E., Warner, T. A., & Guillén, L. A. (2021). Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 1: Literature review. Remote Sensing, 13(13), 2450.

Merchant, M., Haas, C., Schroder, J., Warren, R., & Edwards, R. (2020). High-latitude wetland mapping using multidate and multisensor Earth observation data: a case study in the Northwest Territories. Journal of Applied Remote Sensing, 14(3), 034511-034511.

Merchant, M. et al. (2023). Leveraging Google Earth Engine Cloud Computing for Large-scale Arctic Wetland Mapping. In Review.

National Wetlands Working Group, The Canadian Wetland Classification System, 2nd ed., University of Waterloo, Waterloo, Ontario (1997).

Stevens, D. L., & Jensen, S. F. (2007). Sample design, execution, and analysis for wetland assessment. Wetlands, 27(3), 515-523.

Stewart, R.E. and H.A. Kantrud. (1971). Classification of Natural Ponds and Lakes in the Glaciated Prairie Region. Bureau of Sport Fisheries and Wildlife, U.S. Fish and Wildlife Service, Washington, D.C., USA. Resource Publication 92. 57 pp.

Tassi, A., & Vizzari, M. (2020). Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms. Remote Sensing, 12(22), 3776.

United States Department of Agriculture, Natural Resources Conservation Service. (2018). Field Indicators of Hydric Soils in the United States, Version 8.2. L.M. Vasilas, G.W. Hurt, and J.F. Berkowitz (eds.). USDA, NRCS, in cooperation with the National Technical Committee for Hydric Soils.

Appendix A

Table A-1: *Summary of the staff and subcontractors involved in wetland inventory work.*

Person	Title	Organization	Role
Thorsten Hebben	Alberta Manager of Provincial Operations	DUC	High level facilitation, coordination and oversight, client liaison
Kevin Smith	National Manager Boreal Program	DUC	High level facilitation, coordination and oversight
Alain Richard	Head Conservation Partnerships, Boreal	DUC	Project Coordinator (DUC components) and overall project coordination with ABMI Project Coordinator
Michael Merchant	Remote Sensing Scientist	DUC	Geospatial data analysis, machine learning, report writing, training/testing data development
Rebecca Edwards	Remote Sensing Specialist	DUC	Data and image interpretation, modeling and geospatial analysis and field protocol support
Jim Herbers	Executive Director	ABMI	High level facilitation, coordination and oversight
Dr. Cynthia McClain	Geospatial Centre Director	ABMI	Project management, staff supervision, overall coordination with DUC, reporting
John Simms	Sr. Machine Learning Engineer	ABMI	Field protocol development, training and validation data collection design
Dr. Jenet Dooley	Lead Wetland Scientist	ABMI	Field protocol development, training and validation data collection design
Jennifer Hird	Lead Scientist, Earth Observation Insights	ABMI	Field protocol development and monitoring design, reporting, geospatial and project management support
Dr. Branko Hricko	Lead, Geospatial Foundations	ABMI	LiDAR data management and QA/QC, post-processing, technical advising on aerial data collection
Cris Gray	GIS Analyst	ABMI	Geospatial data management; acquiring and compiling existing LiDAR data and aerial imagery; cartography
Fiona Gregory	Geospatial Data Analyst	ABMI	Geospatial data analysis and image processing

Appendix B

Alberta Biodiversity Monitoring Institute, Ducks Unlimited Canada, Alberta Environment and Protected Areas, 2023, Field Inventory Protocol to Support Satellite Remote Sensing-Based Wetland Mapping in Alberta, 2023-06-15. Alberta Biodiversity Monitoring Institute, Alberta, Canada

See following pages for a copy of the above document.

Field Inventory Protocol to Support Satellite Remote Sensing-Based Wetland Mapping in Alberta

2023 Field Season

Alberta Biodiversity Monitoring Institute
Ducks Unlimited Canada
Alberta Environment and Protected Areas

About the Document

Acknowledgements

This protocol document was developed by Jennifer Hird and Jenet Dooley with technical review and input from Rebecca Edwards, Lyle Boychuk, and Michael Merchant from Ducks Unlimited Canada and Craig Mahony, Josh Montgomery and Danielle Cobbaert from Alberta Environment and Protected Areas. In addition, the Alberta Environment and Protected Areas Wetland Policy team provided technical review.

Disclaimer

Use of this Material

Document Status & History

List of Updates

Update Date	Version	Section	Description of Changes
2022-07-28	1.0	All	Initial Draft
2022-08-29	2.0	All	Revisions based on collaborative discussions between ABMI, EPA and DUC; some restructuring to focus more

Update Date	Version	Section	Description of Changes
			on three main data collection elements: vegetation, soil, and water.
2022-10-07	3.0	2. Wetland Site Selection 3. Field Parameters and Protocols	Clarification and flowchart of site selection processes added, based on EPA/ABMI/DUC collaborative discussions. Revisions made to tree and shrub strata definitions, boundary delineation requirements/approach and site layout, equipment list, and recording observations of disturbance; based on EPA/ABMI/DUC collaborative discussions
2023-01-20	4.0	All	Minor editorial changes for consistency and updates to reflect needed changes identified during field trial
2023-01-27	5.0	All	Editorial changes and updates to incorporate EPA edits
2023-06-15	5.1		Updates to incorporate elements of the EPA wetland policy team review. Changes to reflect 2023 field implementation. Protocol finalized for 2023 field efforts.

Suggested Citation

Alberta Biodiversity Monitoring Institute, Ducks Unlimited Canada, Alberta Environment and Protected Areas, 2023, Field Inventory Protocol to Support Satellite Remote Sensing-Based Wetland Mapping in Alberta, 2023-06-15. Alberta Biodiversity Monitoring Institute, Alberta, Canada. Report available at: abmi.ca [Date Cited].

Contact Information

If you have questions or concerns about this publication, you can contact:

ABMI Information Centre
1-107 CCIS
University of Alberta
Edmonton AB Canada
T6G 2E9
780-492-5531
E-mail: abmiinfo@ualberta.ca

Table of Contents

Glossary of Acronyms and Specific Terms

TERM	DEFINITION
ABMI	Alberta Biodiversity Monitoring Institute
AEPA	Alberta Environment and Protected Areas, Government of Alberta
AWCS	Alberta Wetland Classification System (AEP: GOA 2020b)
BTZ	Boreal Transition Zone
DUC	Ducks Unlimited Canada
EO	Earth observation (reference to satellite sensors)
GPS	Global positioning system
Ground cover	Non-woody (herbaceous) vegetation, bryophytes, bare/exposed ground, litter, or water
Shrub	Woody vegetation species that typically have multiple, branched stems and are < 3 meters tall. Examples include willows, red-osier dogwood, prostrate shrub species, alders, bog birch.
Tree	Woody vegetation species that typically have a single, well defined trunk and are > 3 meters tall. Examples include larch, spruce, birch and balsam poplar.
Wetland complex	A set of contiguous wetland sites (each representing a homogeneous wetland class, form, or water permanence type) that are selected at the same time for sampling during a field campaign. As wetland ecosystems commonly consist of more than one adjacent wetland class, form, or type, these are often sampled together.
Wetland site	An area characterized by a single, relatively homogeneous wetland class or form, or water permanence type; the polygonal unit of analysis used in this document for the purposes of satellite-based wetland mapping

1. Background

This document describes the wetland field data collection protocols used for collecting field wetland inventory data to support the calibration and validation of Earth observation (EO) satellite-based wetland mapping in Alberta. The protocols follow the Alberta Wetland Classification System (AWCS; GOA:ESRD 2015), the Alberta Wetland Classification Field Guide (DUC 2021) and the Alberta Wetland Mapping Standards and Guidelines (GOA: AEP 2020a).

The primary purpose of these protocols is to collect critical physical and biological data that efficiently describes classification of Alberta's wetlands to class, form and type levels enabling accurate classification at these levels.

Figure 1 shows the wetland classes, forms, and water permanence types outlined in the AWCS for describing Albertan wetlands. Note that this protocol will collect information on ephemeral permanence types to inform their distinction from temporary classes. Salinity and alkalinity types are not shown.

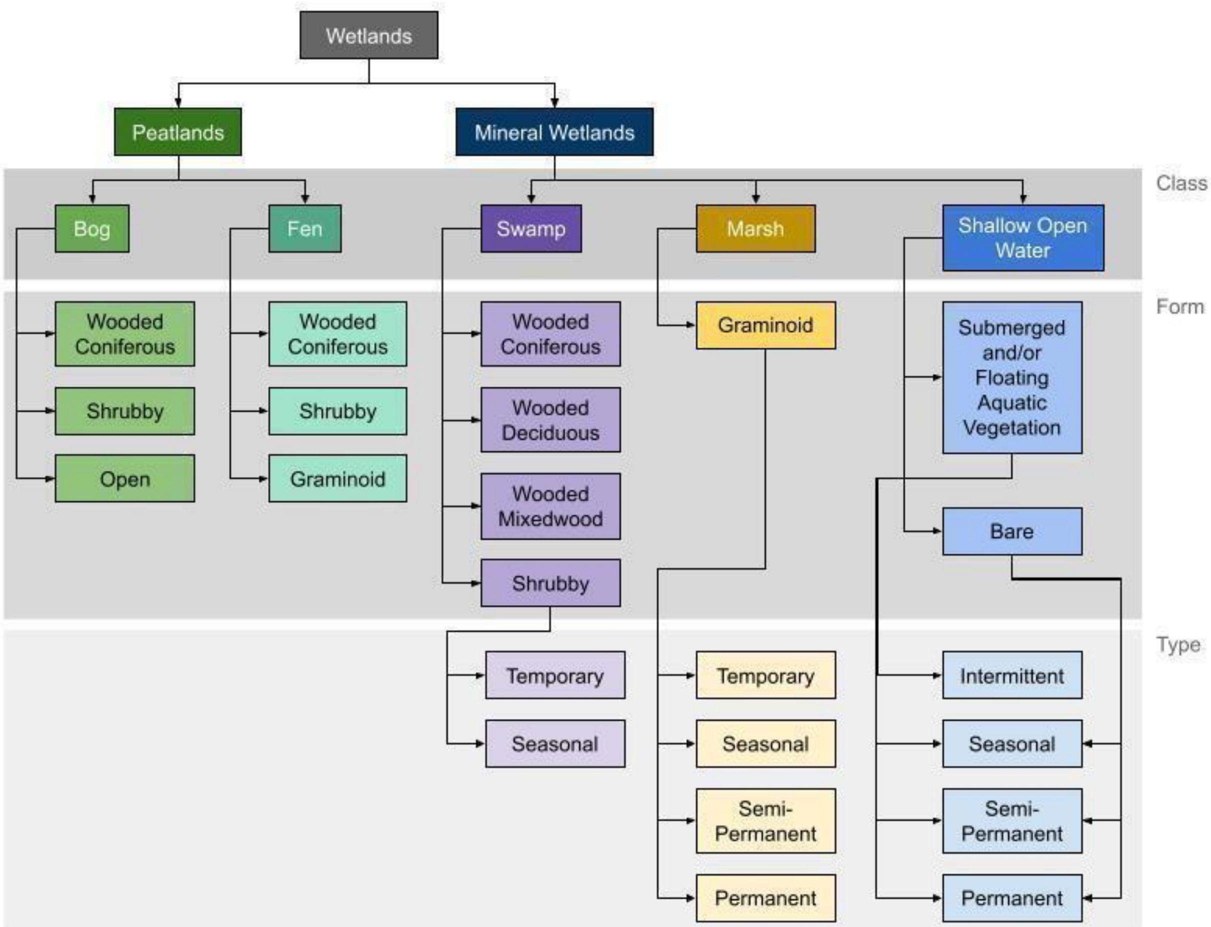

Figure 1. Flow chart outlining the wetland classes, forms, and water permanence types as found in the Alberta Wetland Classification System. Adapted from (GOA:ESRD 2015).

Wetland Mapping Zones

As outlined in the Alberta Wetland Mapping Standards and Guidelines (GOA: AEP 2020a), the Prairie/Parkland zone comprises the Grassland and Parkland Natural Regions of Alberta (Nat. Reg. Comm. 2006), where prairie pothole, mineral wetlands such as seasonally- and annually-dynamic marshes, and in more northern areas swamps, dominate (GOA: AEP 2020a). The Wetland Mapping Standards combine the Boreal and Foothills Natural Regions into the Boreal/Foothills zone (GOA: AEP 2020a), but for the purposes of the present documentation, we also include the Rocky Mountain and Canadian Shield Natural Regions in this zone as well. The latter zone is characterized by vast peatland systems of fens and bogs and less dynamic water table fluctuations. Figure 2 shows the distribution of Natural Regions across Alberta and the boundaries of these two zones.

The Boreal Transition Zone (BTZ) marks the transition between the Boreal/Foothills and Prairie/Parkland zones (GOA: AEP 2020a), and as such, its wetlands display characteristics from both. No defined boundary of the BTZ exists, but it is identifiable by the mixture of northern and southern Alberta wetland characteristics within its landscapes.

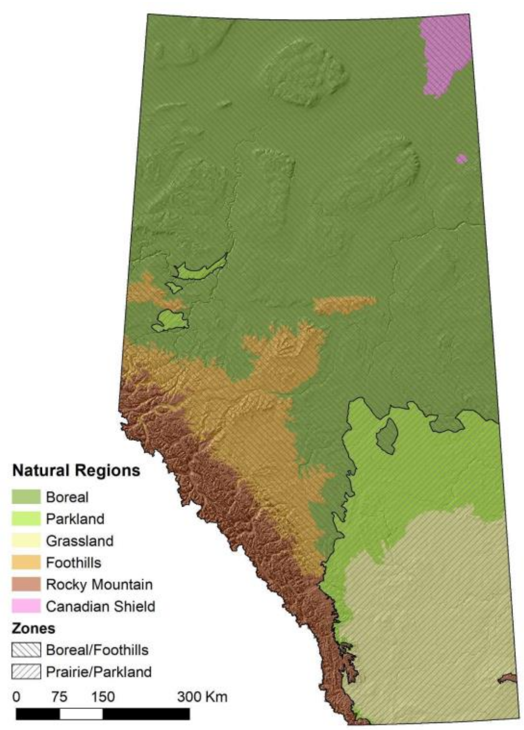

Figure 2. Map of the Natural Regions of Alberta, showing the Boreal/Foothills and Prairie/Parkland zones (data source: Natural Regions Committee, 2005).

The current EO-based mapping of Alberta's wetlands is done using object-based image analysis, wherein relatively homogeneous groups of pixels are grouped together into 'objects' or 'segments', treated as a single unit or polygon, and categorized as a single wetland form, or where appropriate, type (e.g., water permanence type in prairie areas). The resulting inventory is therefore polygon-based and is to be calibrated and validated at this unit of analysis, which we refer to as wetland sites. This contrasts with pixel-based approaches to wetland mapping, where individual pixels are used in calibration and validation.

2. Wetland Site Selection

The sampling strategy will generally be stratified by geographical location and natural variability and account for accessibility (e.g., with truck vs. helicopter, land access permissions, etc.), as well as the location or coverage of previously existing datasets (e.g., DUC validation data). Site selection will seek a balanced design of the various wetland classes and forms, and consider the classification accuracy of various classes and forms based on previous wetland inventory and classification efforts (i.e. increased sampling of wetland classes and forms with poorer accuracies).

Wetland sites selected for field assessment will conform to and exceed where possible the Minimum Mapping Units (MMUs) established in the ‘Wetland Mapping Standards and Guidelines’ (GOA: AEP 2020a) that are as follows:

- 0.04 hectares (i.e. 400 m^2, or 20 metres [m] x 20 m) in the Prairie/Parkland and Boreal Transition Zones
- 0.9 hectares (i.e. 9000 m^2, or ~95 m x 95 m) in the Boreal/Foothills zone

Adhering to these MMUs is important so as to avoid introducing unreliable, granular noise into classification models. Moreover, these MMUs were selected with both reliability and accuracy in mind, whereby their values put the resolution of satellite imagery in the context of real wetland features. A smaller MMU is necessary in the Prairie/Parkland zone due to the high occurrence of small wetlands, whereas in the Boreal/Foothills zone a larger MMU accounts for the more heterogeneous nature of the vegetated wetlands found there.

In addition to meeting and preferably exceeding the MMU thresholds, wetland sites selected for field assessment should have a minimum width of 40 m in the Boreal/Foothills zone and 5 m in the Prairie/Parkland zone.

2.1 Site Selection Procedure

Wetland site selection will be a desktop-based exercise (Figure 3). This will involve selecting quarter sections within the designated project/mapping area. Quarter sections will be stratified by their Natural Subregion (NSR) and at least one quarter section will be selected from each NSR within the project area. Each quarter section will have their wetland coverage, number, and diversity estimated from existing wetland inventories (e.g., the Alberta Merged Wetland Inventory (AMWI) and/or ABMI’s Alberta Wetland Inventory). Once assessed, the quarter sections with a shortlist of the best representation of wetland classes, forms, types, etc. will be selected from each NSR - e.g., using a combination of the greatest number of wetlands and wetland diversity. A random selection of quarter sections from this shortlist will then be extracted. An analyst will select individual sampling sites within the designated quarter sections using a random stratified sampling approach. For this, wetland class and permanence type (if in the Prairie/Parkland zone) or class and form (if in the Boreal/Foothills zone) will inform the stratification process. In the Boreal/Foothills zone, sites will be selected or stratified within non-impacted and fire-impacted areas and prominent human disturbance types. Wildfires are naturally occurring and frequent in the boreal, and thus we recommend sites be selected in both burned and unburned areas with the proportions being approximately 70/30, respectively, depending on fire history. In the Prairie/Parkland zone, stratification by human disturbance will be considered. The randomly selected quarter sections and sites will be reviewed by an analyst using the most recent satellite imagery available. Sites will be reviewed for their wetland likelihood, and homogeneity. For example, if sites appear to be very mixed (e.g., mixed wetland classes, or mixed vegetation strata) then they will not be selected for in-field sampling. Lastly, site accessibility will be taken into account during quarter section and site selection.

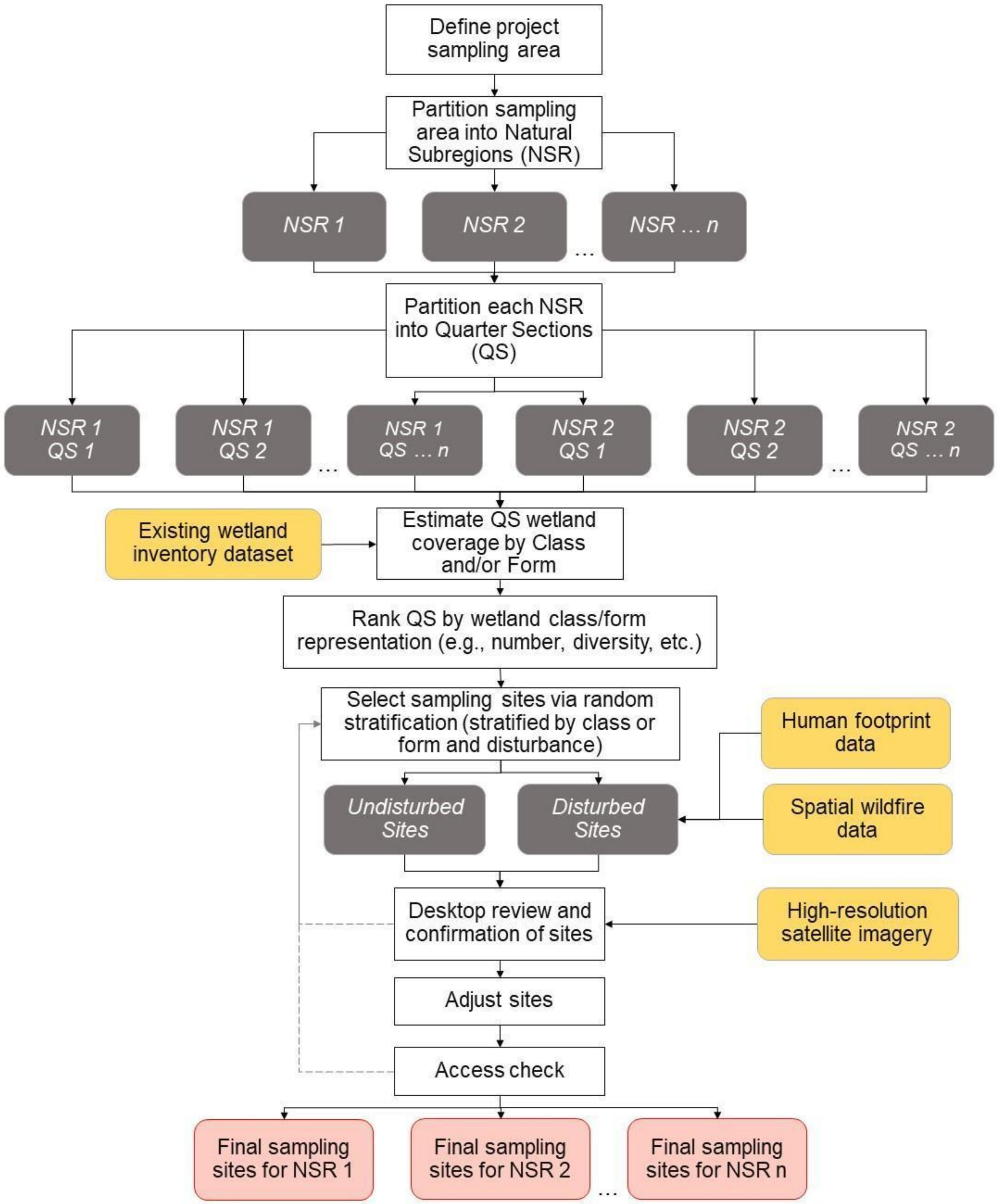

Figure 3. Flowchart depicting the general site selection process. Adapted from Michael Merchant, DUC Boreal Program.

3. Field Inventory Indicators and Protocols

The field inventory indicators and protocols are intended to support accurate, repeatable, and efficient determinations of AWCS class and form that can be used to validate wetland inventory models using a remote sensing and machine learning approach. The wetland field inventory protocols focus on wetland indicators and thresholds used to identify AWCS classes, forms, and water permanence types following the *Alberta Wetland Classification System Field Guide* (DUC 2021).

The following sections provide an overview of the wetland site layout, equipment requirements, and protocols for each set of indicators.

3.1 Overview of Wetland Parameters

Key wetland parameters sampled include:

- Vegetation (e.g., height, cover, indicator species present) (see Section 3.4 for details)
- Soil characteristics (hydric mineral soil indicators, peat depth measurements and level of decomposition (see Section 3.5 for details)
- Water physiochemistry and hydrological indicators (e.g., pH and electrical conductivity, evidence of fluctuating water levels) (see Section 3.6 for details)
- Wetland site boundary delineation e.g., walking with a GPS unit using the route tracking function (see Section 3.7 for details)
- Two representative photographs of each site, 1 landscape photo and 1 top-down photo of the vegetation ground cover (see Section 3.8 for details)
- Wetland class and form determination using the Alberta Wetland Classification Field Guide (DUC 2021) (see Section 3.9 for details)
- Observations of human and fire disturbance at the wetland site (see Section 3.10 for details)

3.2 Wetland Site

Each wetland site must have a minimum width of 40 m in the Boreal/Foothills zone and 5 m in the Prairie/Parkland zone. The wetland indicators will be documented within a homogeneous area representative of the larger wetland site, and along the boundaries of the site.

The homogeneous area will be the mapped feature that remote sensing scientists use to validate and train mapping models. The larger the area and the more accurately the field data represents this area, the more useful it will be. There are minimum sizes required for homogeneous areas in the different zones of the Province as specified in the following two paragraphs.

In the Boreal/Foothills, the standard homogeneous area is a 20 m radius (40 m diameter) circular area within the wetland. The wetland site must be large enough to fit the homogeneous area but can be larger. The vegetation plot is placed in the center of the homogeneous area. Figure 4 illustrates example wetland site layouts for several sites at a sampled wetland complex in the Boreal/Foothills zone.

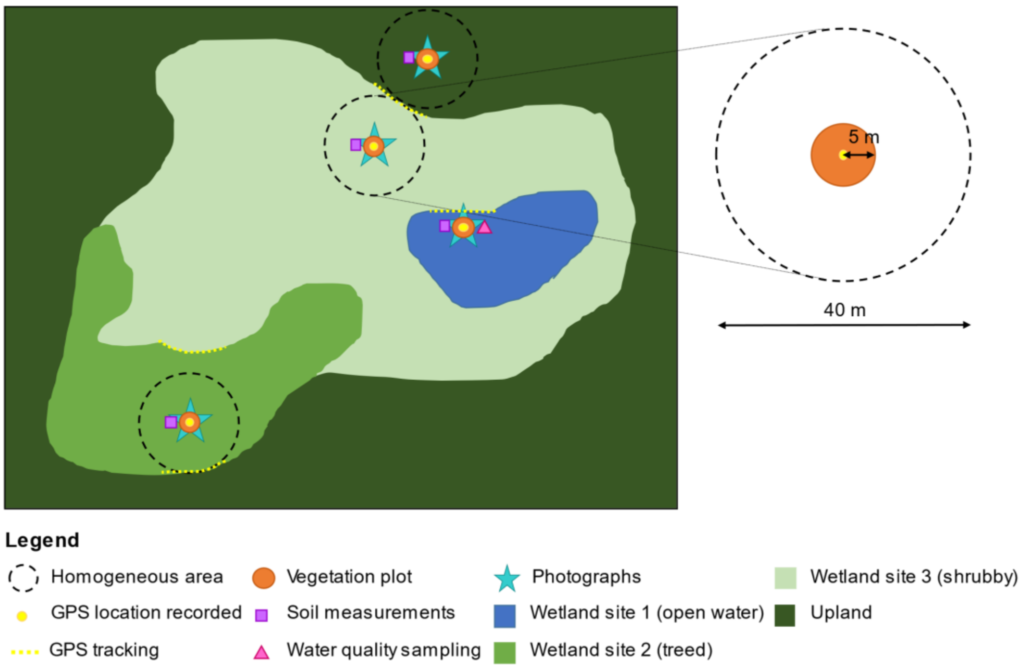

Figure 4. Diagram illustrating the layout of sample collection protocols at a wetland complex comprising three wetland sites and an upland site within the Boreal/Foothills zone.

In the Prairie/Parkland zone, wetlands are typically structured with concentric rings of wetland class, form, and/or type that follow hydrologic gradients of wet areas near the center of depressions and drier areas around the edges. The alignment of vegetation plots should be oriented to bisect each concentric ring. A transect will run perpendicular to these rings, and oriented from the upland to the deepest or wettest portion of the topographic depression (Figure 5). One vegetation plot is placed along the transect within the upland and each distinguishable wetland site it crosses, as indicated by unique wetland class, form, or water permanence type and associated vegetation communities. Each wetland site must be large enough for a 5 m diameter circular homogeneous area to be assessed. The vegetation plot is placed in the center of the homogeneous area. The vegetation plot's location should be equidistant from the closest edges of the class, form, or type it is within.

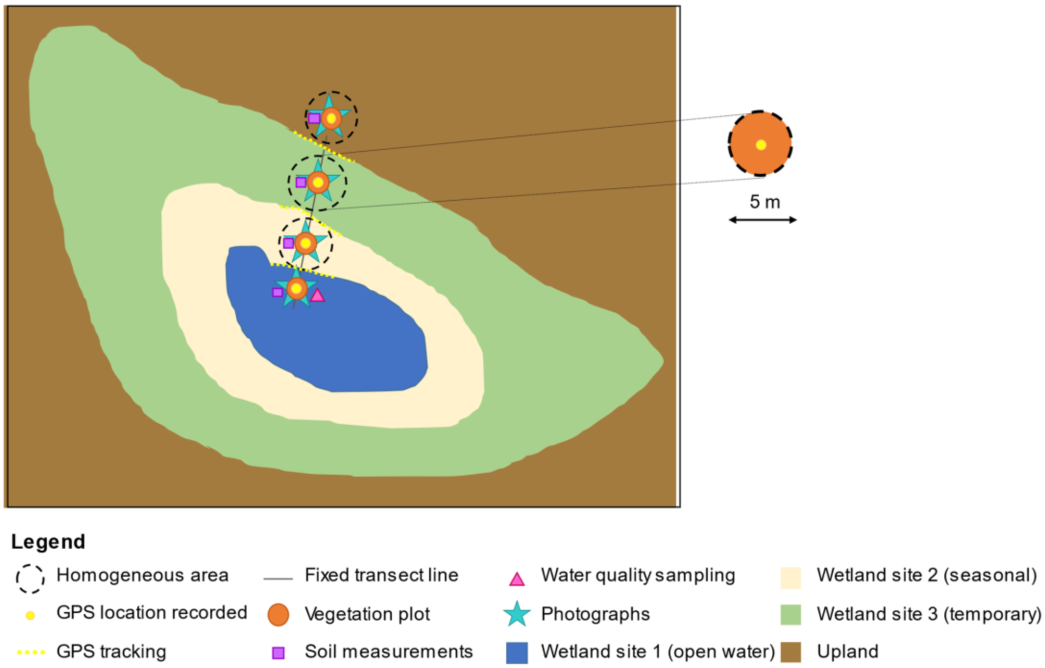

Figure 3. Diagram illustrating the layout of sample collection protocols at three wetland and one upland sites within a wetland complex in the Prairie/Parkland zone.

Wetland Site Layout

For efficient site assessment, the site layout should be established from the direction that you walk into the site. When you enter a wetland site, first establish the homogeneous area. Where possible, homogeneous areas can be larger than the minimum size requirements (20m radius in Boreal/Foothills and 2 .5 m in Prairie/Parkland, as detailed above). Assess as large of an area as possible that is homogeneous. The area should have a uniform, circular shape. Record the diameter of the homogeneous area on the field sheet. If the homogeneous area is very large and/or difficult to navigate, measure the distance reached in five minutes of effort and write greater than this measure on the homogeneous diameter section of the field sheet.

One representative vegetation plot, 5 m radius for Boreal/Foothills and 2.5 m radius for Prairie/Parkland, is placed at the centre of the homogeneous area. Soil samples are observed adjacent to the vegetation plot, and water is sampled within or adjacent to the vegetation plot where water is present (see following sections for further detail). Geolocational position (i.e., GPS point) is recorded at the centre of the vegetation plot. Geolocational information is collected along a maximum of two boundaries (e.g., upland-wetland boundary and a wetland class 1 vs class 2 boundary) using a GPS tracking feature for a distance of 30 m, where the collection of such information can be done in under 10 minutes, per boundary (including time required to walk to the boundary itself).

3.3 Equipment List

General Equipment

- Field sheet
- Aerial photo of site (digital or printed)
    - Include information from pre-existing polygon-based wetland inventory, Human Footprint Inventory, and Alberta Historical Wildfire Perimeter data
- Photo sheet
- Pencils, pens, sharpies
- Tablet with digital field data collection app
- DUC Field guide
- Wetland Classification Decision Key

Vegetation Assessment

- GPS receiver (e.g., hand-held GPS unit, differential GPS unit, GNSS GPS receiver)
- Compass
- Flagging tape and/or pigtail
- 2 meter folding measuring stick
- Inclinometer (Vertex Hypsometer)
- Tape measure (100 m in length)

Soil Assessment

- Organic soil probe
- Measuring tape
- Soil auger (no extensions)
- Water eye dropper for dry soil

Hydrology Assessment

- Water quality multi-probe (E.g, hydrolab)
- Soil auger
- 2 meter folding measuring stick
- Collapsible water bucket and scoop

Boundary Geolocation Protocol

- GPS receiver (e.g., hand-held GPS unit, differential GPS unit, GNSS GPS receiver)

Photography

- Digital camera
- Paper or whiteboard & appropriate marker (Photo sheet)
- Tape measure

3.4 Vegetation Assessment

Vegetation is to be recorded in circular plots with a 5 m radius (10 m diameter) in the Boreal/Foothills zone and 2.5 m radius (5 m diameter) in the Prairie/Parkland zone, centered within the homogeneous area (see section 3.2 for details). Information on observed species will focus on wetland indicator species in particular, though additional dominant species can also be recorded.

3.4.1 Vegetation Plot

Once the location for a vegetation plot has been selected, mark it with a pigtail or flagging tape. Use a GPS-enabled device (e.g., hand-held GPS unit, differential GPS unit, Real-Time Kinematic/GNSS GPS unit) to collect locational coordinates at this marked centre of the vegetation plot. The device used should be appropriate for the field conditions so that it has +/- 3 m accuracy (e.g., a more precise unit may be needed in treed environments). This is to be done using the Alberta 10TM (Forest) coordinate system (datum: NAD83), as designated by the EPSG code 3400. The coordinates (northings and eastings, in metres) are to be recorded on the electronic field sheet, along with any associated locational accuracy estimates.

Use a measuring tape to distinguish the edges of the 5 m or 2.5 m radius circular vegetation plot (i.e., what vegetation is “in” or “out” of the plot).

3.4.3 Vegetation Strata

There are three vegetation strata to observe at each vegetation plot: trees, shrubs, and ground cover. They are defined as follows.

- Trees: woody vegetation species that typically have a single, well defined trunk and are > 3 meters tall. Examples include larch, spruce, birch and balsam poplar.
- Shrubs: woody vegetation species that typically have multiple, branched stems and are usually < 3 meters tall. Examples include willows, red-osier dogwood, prostrate shrub species, alders, bog birch.
- Ground cover: non-woody (herbaceous) vegetation, bryophytes, bare/exposed ground, litter, or water.

At each plot, record the overall cover of the tree and shrub strata. Measure the height of representative tree and shrub individuals for the plot. The selected individuals should be of typical height and species for the plot. The number of individuals measured depends on the overall cover of the strata as seen in Table 1.

Table 1. Number of individual plants to measure for height for each vegetation strata in the plot, based on the overall cover of that strata.

Overall Strata Cover	# of individuals
<=5%	0
6-49%	3
50-75%	4
>76%	5

3.4.3.1 Trees

Record what percent of the tree cover is deciduous and conifer trees. Indicate the cover of tamarack and black spruce. Indicate which species are present in the vegetation plot.

3.4.3.2 Shrubs

Indicate what percent of the shrub cover is ericaceous species (i.e., flowering shrubs that belong to the heath family and are associated with acidic soil. Includes labrador tea, leatherleaf, bog cranberry.) Indicate which species are present in the vegetation plot.

3.4.3.3 Ground Cover

Fill in the % cover for the ground strata categories and select the species from the list that are characteristic of the plot. Characteristic species are those that are most common and together, describe the majority of the species present. You do not need to include species that are relatively rare in the plot. If a characteristic species is missing from the list, write it in.

3.5 Soil Assessment

Soil sampling is done adjacent to each vegetation plot.

Extract a soil core sample using the augur. If the uppermost soil layer is saturated, record this. If it is dry, moisten it using a few drops of water. Determine if the uppermost soil layer is mineral or organic soil by following the USDA field estimation method (USDA 2018). Gently rub a pinch of the wet soil material between your forefinger and thumb. If upon the first or second rub the material feels gritty, it is mineral soil material. If after the second rub the material feels greasy, it is either mucky mineral or organic soil material. Gently rub the material two or three more times. If after these additional rubs it feels gritty or plastic, it is mucky mineral soil material; if it still feels greasy, it is organic soil material. Indicate the soil type on the field sheet. If it is organic soil, determine the Von Post Decomposition rating and the peat depth. See DUC (2021) p.146 for details. If mottling, gleying or oxidized root channels are observed, indicate it on the field sheet and record the depth from the surface that they are first present. If hydrogen sulfide odor is observed, indicate it on the field sheet. See DUC (2021) p. 28 for more information on these indicators. Take a photo of the soil core sample. Include a photo sheet with site name written on it within each photograph.

Site name = Pilot area [BORL1, BORL2, GRSS1, PRKL1, OSM1] – Quarter Section letter – site number

E.g., GRSS1-B-3

3.5.2 Peat Depth

If peat is present, obtain the total peat depth by pushing the organic soil probe into the peat (adding extension rods as necessary) until significant resistance is met. If the depth exceeds 4m, record “>4m”. The augur hole from the sample collection can be utilized to help guide the probe into the peat. Withdraw it from the peat and measure the total peat depth using a measuring tape. Clean the mineral soil from the tip of the rod at each plot to avoid a false reading at the next plot.

3.6 Hydrology Assessment

Water sampling is done at each wetland site, with measurements taken within or adjacent to the vegetation plot located within that site. If there is no water present within the vegetation plot, the homogeneous area around the plot should be searched for water to sample or water collected in the augur hole can be measured. Ensure that the sampled area is in the same wetland class, form, or type as the vegetation plot. Follow the protocols below.

3.6.1 Hydrological Indicators

Record on the field sheet the presence of hummocky microtopography, beaver activity, and water table depth. These should be observed in the vegetation plot and/or within the homogeneous area (20 m radius in the Boreal/Foothills, 2 .5 m radius in the Prairie/Parkland) it represents. See DUC (2021) for more information on hydrological indicators.

Water physiochemistry is sampled using in-situ measurements of pH and electrical conductivity, taken within each wetland site (class, form, or type). If there is standing water present within the wetland site, a minimum of three measurements are to be taken. Use a water quality multi-probe to take measurements just below the surface of the water if using a dip probe or at mid depth if using a cabled probe. The measurements can be taken from water that collects in the auger hole if there is no surface water present.

Estimate the water table depth. Record positive measurements for surface water and negative measurements for water tables below the surface. Record the typical depth for the wetland site by measuring the water depth where you sample water physiochemistry (i.e., 3 surface water locations or auger hole). Estimate deeper standing water depths from shore.

3.7 Delineating 'Site' Boundaries

Delineating the boundary is important for mapping homogeneous wetland sites. The longer and more accurate a delineated boundary, the more valuable it is for validation purposes. Up to two boundaries are delineated at each site, where a boundary can be delineated within 10 minutes or less (including time required for walking to the boundary itself).

Boundaries to be delineated include those between:

- Upland and wetland areas
- Wetland sites (i.e., unique classes, forms, and/or type)

Boundaries should not represent transition areas - they should be the furthest edge that you are confident still represents your homogeneous area. Delineate 'site' boundaries (i.e. boundaries of the target wetland class, form or type) by walking the boundary and using the route tracing function (continuous point collection) for at least 30 m. The device used should be appropriate for the field conditions so that it has +/- 3 m accuracy (e.g., a more precise unit may be needed in treed environments).

In the Prairie/Parkland zone, ensure that the boundary delineations are collected where the vegetation plot transect and boundary cross. In the Boreal/Foothills, delineate the closest boundary to the vegetation plot.

3.8 Wetland Classification

Use the information collected at the wetland site and the flow chart on page 32 of the DUC *Alberta Wetland Classification System Field Guide* (DUC 2021) to classify each wetland site. The resolution required is class and form for all wetland sites and water permanence type for marsh and shallow open water classes.

3.9 Photographs

Two photographs are taken at the vegetation plot located within each wetland site. For data management and recording purposes, include a photo sheet with site name written on it within each photograph.

Site name = Pilot area [BORL1, BORL2, GRSS1, PRKL1, OSM1] – Quarter Section letter – site number

E.g., GRSS1-B-3

Protocols for taking the two photographs area as follows:

1. Take one landscape photograph at eye height from the centre of the vegetation plot toward the wettest/deepest part of the wetland site or complex in the Prairie/Parkland zone, or toward the 'centre' of the wetland site or complex in the Boreal/Foothills zone.

- Stand at the centre of the selected vegetation plot
- Have a crew member stand at the edge of the plot within view of the camera - i.e. 5 metres from the camera's location in Boreal - and holding a photo sheet or whiteboard with the site name written on it.
- Take the photo
- Record the file number of the photo in the camera on the field.

2. Take one ground photograph at approximately 1.5 m from the ground, at the centre of the vegetation plot, oriented directly down toward the ground.
    - Stand at the centre of the selected vegetation plot
    - Point the camera downward at the ground, holding it at eye height
    - Ensure the written site name is present within the camera's field of view
    - Take the photo
    - Record on the field sheet that this photo was taken

3.10 Disturbance

Confirm the presence of any human or natural disturbance indicated on the site map within the homogeneous area. These disturbances may include wildfire and human impacts (e.g., agriculture, roads, seismic lines). Indicate on the field sheet whether the disturbance(s) is(are) within the vegetation plot at the site or adjacent to the vegetation plot but within the homogeneous area being sampled. Record a description of the observed disturbance(s) and type(s) (e.g., road, seismic line, wildfire) on the field sheet.

3.11 Extra Boreal/Foothills Wetland Sites (2 max)

When delineating the site boundary and when measuring the diameter of the homogeneous area in the Boreal/Foothills, check to see if there appears to be another wetland form/class within the polygon shown on the site map. If there is, carry out another assessment on this adjacent site labeling it with the next number available for that quarter section (see section 3 below). Carry out this protocol for up to two extra sites per polygon. If there are more than two extra unique wetland sites note this on the field sheet. If the wetland site is too large or time consuming to carry out the boundary delineation and diameter measurements (as mentioned above) this protocol will be abandoned.

For Prairie/Parkland wetlands this practice is already in place with the assessment plots aligned in a transect perpendicular to and bisecting the concentric ring wetlands (Fig. 5). This protocol is not limited to two extra wetland sites nor is it to be abandoned.

4. Data Quality Management & Integrity

To ensure data quality, the following data management tasks should be completed at the end of each field day:

- Ensure all data sheets are complete and uploaded in Epicollect once in internet service.
    - If using paper datasheets, photograph the sheets (check that it is clear and legible) and upload a copy to your crew folder and correct shift sub folder on FTP. Label the photograph Site name-datasheet.
        - Pilot area [BORL1,BORL2, GRSS1, PRKL1,OSM1] - Quarter Section letter-site number (as noted on map and given GPS point)- datasheet E.g. BORL1-F-3-datasheet
- Download and label all site photos. Put the labelled photos in your crew folder and correct shift sub folder on FTP. Photo names should be changed to the following format: Site name -”landscape” or “ground” or “soil”
    - E.g., BORL2-C-4-ground

Note: If more unique wetland forms/types are found within a polygon, simply label it with the next number available for that quarter section. E.g., if quarter section BORL1-A has three sites but site two has an extra unique wetland within its polygon, name the extra wetland ground photo “BORL1-A-4-ground”

- Download all GPS waypoints either as a single file with the Ozi Explorer program or grouped in a folder with the Gaia program. The file/folder should be named by its quarter section (E.g., BORL1-B). Upload the file/folder to your crew folder and correct shift sub folder on FTP. Individual GPS points should be labeled: Site name -point type. Point type examples are:
    - Vegetation plot: [Sitename]- VP
        - E.g., GRSS1-D-3-VP
    - Site boundary locations: [Sitename] – SB1 or SB2
        - E.g., PRKL1-B-2-SB1 (or PRKL1-B-2-SB2 if a secondary boundary is completed)
- Clean field equipment with disinfectant following legislative requirements to limit spread of invasive species.
- Charge batteries and calibrate equipment as needed for the next field day.

Field crew training will include classroom-based learning (either in-person, online, or through pre-recorded videos) provided by wetland experts familiar with the protocols, such as individuals from the ABMI or DUC. This will be followed by an in-field (in-person) component for hands-on demonstration and practice, similarly provided by wetland experts familiar with the protocols.

5. Resources

Cobbaert, D., Conner, S., Montgomery, J., and Mahoney, C. 2022. Alberta Boreal Wetland Field Protocols. Government of Alberta, Ministry of Environment and Parks. DRAFT.

Ducks Unlimited Canada (DUC). 2021. Alberta Wetland Classification System Field Guide. Edmonton, Alberta. https://borealducks.lpages.co/alberta-wetland-classification-system-field-guide/

Government of Alberta - Alberta Environment and Sustainable Resource Development (GOA:ESRD). 2015. Alberta Wetland Classification System. Water Policy Branch, Policy and Planning Division, Edmonton, Alberta. https://open.alberta.ca/publications/9781460122587

Government of Alberta – Alberta Environment and Parks (GOA: AEP). 2020a. Alberta Wetland Mapping Standards and Guidelines: Mapping Wetlands at an Inventory Scale v1.0. Edmonton, Alberta. https://open.alberta.ca/publications/alberta-wetland-mapping-standards-guidelines-mapping-wetlands-inventory-scale-version1

Government of Alberta - Alberta Environment and Parks (GOA: AEP). 2020b. Alberta Merged Wetland Inventory. https://geospatial.alberta.ca/titan/rest/services/environment/alberta_merged_wetland_inventory/MapServer/

Natural Regions Committee 2006. Natural Regions and Subregions of Alberta. Compiled by D.J. Downing and W.W. Pettapiece. Government of Alberta. Pub. No. T/852.

Pennock, D.J., Watson, K., and Sanborn, P. 2015. Horizon Identification. From: D. Pennock, K. Watson, and P. Sanborn. Field Handbook for the Soils of Western Canada. Canadian Society of Soil Science.

United States Department of Agriculture (USDA), Natural Resources Conservation Service. 2018. Field Indicators of Hydric Soils in the United States, Version 8.2. L.M. Vasilas, G.W. Hurt, and J.F. Berkowitz (eds.). USDA, NRCS, in cooperation with the National Technical Committee for Hydric Soils. https://www.nrcs.usda.gov/sites/default/files/2022-09/Field_Indicators_of_Hydric_Soils.pdf