Wetlands Ecol Manage (2025) 33:19
https://doi.org/10.1007/s11273-025-10034-7

ORIGINAL PAPER

Wetlands happen: the delineation and classification of opportunistic wetlands in the Athabasca oil sands region of Canada

Virgil C. Hawkes · Jon Hornung · Julio Novoa · Michael T. Miller · Sarah Kohlsmith

Received: 10 October 2024 / Accepted: 17 January 2025 / Published online: 30 January 2025

Abstract Achieving land capability equivalent to that which existed prior to disturbance is the primary goal of reclamation in the Athabasca Oil Sands Region of northern Alberta. To date, most reclamation has focused on the re-creation of upland forest ecosystem analogues. However, a few wetlands have also been constructed. Additionally, wetlands have appeared spontaneously on landforms reclaimed to an upland forest type. Classifying and quantifying these opportunistic wetlands is an important consideration relative to oil sands closure and reclamation planning. Here we describe an approach using topographic and spectral variables to train a machine learning model (random forest) to detect and classify wetlands as an alternative to on-screen visual delineation. The aim was to develop a model that not only predicts where wetlands occur on reclaimed landforms but that is sensitive enough to classify them as to wetland form. Two random forest models were developed that predicted wetland occurrence at two levels: (1) wetland vs. non-wetland (to generate a prediction of all wet areas on reclaimed landforms); and (2) wetland class (with specific emphasis on marsh and shallow open water wetland classes). In addition to successfully predicting wetland occurrence, the resulting models handled the variability in reclamation approach, substrate type, and soil placement depth with high accuracy. This work confirmed that ~ 18% (211 ha) of the upland-reclaimed area at Suncor Energy's Base Plant north of Fort McMurray, Alberta develops not to upland but to unintentional wetland, consistent with earlier studies. The ability to predict wetlands on the landscape could be invaluable when considering metrics of success associated with landscape reclamation in the Athabasca Oil Sands Region and for informing future inquiries around wetland persistence, resilience, and spatial connectivity through time on reclaimed landscapes.

Keywords Opportunistic wetland · Mineral soil wetlands · Athabasca oil sands region · Reclamation · Remote sensing · Machine learning

V. C. Hawkes (✉) · J. Novoa · M. T. Miller
LGL Limited environmental research associates, Sidney, BC, Canada
e-mail: vhawkes@lgl.com

J. Hornung · S. Kohlsmith
Suncor Energy Inc, Calgary, AB, Canada

Introduction

The Athabasca Oil Sands Region (AOSR), in northern Alberta, Canada, is the fourth largest proven oil deposit in the world. It is also an area of diverse upland forests and extensive wetland habitat, with as much as 54% of the area historically covered by wetland vegetation (AEP Alberta Environment and Parks (AEP), 2018). Bitumen extraction in the AOSR, like other natural resource extraction industries, results in large-scale and long-term habitat conversion that alters natural ecosystems, including wetlands (Volik et al. 2020). Environmental regulations require oil sands mines to re-establish, as part of their ongoing operations, functional landscapes of 'equivalent capability' as compared to the pre-existing landscape that integrate with the surrounding landscape; although specific areal requirements are not part of the regulatory guidelines (Government of Alberta 2021; Gosselin et al. 2010). To date ~ 10,000 ha of boreal landscape disturbed for oil sands mining has been permanently reclaimed (AEP Alberta Environment and Parks (AEP), 2023). Most habitat reclamation in the AOSR has focused on the re-creation of upland ecosystems typical of the region with only relatively small areas reclaimed to mineral soil wetland (shallow open water, marsh, swamp and transitional shrub) or peatland (fen, bog) ecosystems (CEMA 2014; Ketcheson et al. 2016; Hawkes and Gerwing 2019; Hawkes et al. 2021; Borkenhagen et al. 2023) even though much of the pre-mined landscape contained peatland wetlands. Constructed wetlands usually are proposed to make up less than 5% of the closure landscape. Nevertheless, the organic matter that was salvaged from wetland bogs and fens prior to disturbance, and commonly referred to as peat-mineral mix, was often laid down during reclamation as an organic rich overlayer on reconstructed upland habitats (Pinno and Hawkes 2015). Following reclamation, researchers noticed wetlands forming on landforms reclaimed to upland forest types. These unplanned, spontaneous developments were referred to as 'opportunistic wetlands' (Little-Devito et al. 2019).

Opportunistic wetlands have been shown to increase the suitability of reclaimed landforms for wildlife (Hawkes et al. 2020) and, importantly, contribute to the development of reclaimed landforms that approximate the mosaic of wetlands, peatlands, and upland forest that dominated the landscape prior to mine development in the region. While the underlying mechanistic factors responsible for opportunistic wetland development have received some attention (Gringas-Hill et al. 2018, Trites and Bayley 2009, Little-Devito et al. 2019, Price et al. 2010, Hawkes et al. 2020) and are of relevance to future wetland reclamation efforts, tools that can reliably predict and quantify opportunistic wetlands are also needed for quantifying the total area of these unplanned developments.

This study builds from earlier work that used photo interpretation and remote sensing analysis to estimate the total area of opportunistic wetlands forming on reclaimed landforms (Hawkes et al. 2020). Documenting the area occupied by opportunistic wetlands, along with their permanence and persistence, is an important consideration of closure and reclamation planning which aims to promote the establishment of functional post-mining landscapes. Here we use topographic and spectral (i.e., optical, radar) variables, applied to previously-collected observations (Novoa and Hawkes 2021) from one reclaimed landform on Suncor's Base Plant north of Fort McMurray, Alberta, to train a machine learning model (random forest) to detect and classify wetlands as an alternative to on-screen visual delineation. The application of machine learning algorithms, including random forest models (Breiman 2001) to ecological datasets has been increasing in recent times because these algorithms are not bound by linearity, can handle a large number of covariates, and can detect complex relationships and interactions (Shoemaker et al. 2018; Bohnett et al. 2020; O'Malley et al. 2024). Furthermore, the application of remote sensing has been applied to modern wildlife studies (e.g., Hawkes et al. 2024) and wetland classification work in Canada (e.g., Mirmazloumi et al. 2021), but to our knowledge, the use of remote sensing attributes and machine learning to detect and classify wetland areas developing on reclaimed landforms is novel and has potentially broad implications with respect to closure and reclamation outcomes in the AOSR. Improvements to on-screen delineation through the use of remote sensing variables in a machine learning model allow for an iterative and optimized approach. By cycling through multiple options for the model inputs, a robust tool that can be applied to other landforms is created. As such, the model could potentially be applied across a broader range of spatial–temporal datasets to identify the location and class of wetlands that have formed on those landforms. The general applicability of these methods to the broader AOSR reclamation landscape is the primary focus of this paper.

Study area

Suncor's Base Plant is located approximately 40 km north of Fort McMurray, Alberta, Canada in the Central Mixedwood subregion of the Boreal Forest Natural Ecoregion of Alberta (NRC 2006) and the Boreal

Plains ecoregion of Canada. The operation has been producing bitumen since 1967 and has a total footprint of 24,426 ha, of which ~ 2,495 ha (10.2% of the total footprint) have been reclaimed. Approximately 1,209 ha of reclaimed land was remotely surveyed by Hawkes and Novoa (2016) and later field-validated by Hawkes et al. (2020). For purposes of wetland delineation, the study area was divided into 11 focal areas (Table 1, Fig. 1). Each of these Areas of Interest (AOI) contained wetlands that formed opportunistically on landforms reclaimed to upland forest ecosites common in the region.

Methods

Field validation

Previous work by Novoa and Hawkes (2021) used various modelling approaches to predict the occurrence and distribution of wetlands and their associated classifications (non-wetland, marsh, shallow open water, and unspecified [transitional] wetlands) on one reclaimed landform on Suncor's Base Plant north of Fort McMurray, Alberta. The predicted outputs required field validation to assess the accuracy of the model to (1) predict the occurrence and extent of opportunistic wetlands, and (2) to accurately classify opportunistic wetlands over all reclaimed landforms, an area of almost 2,500 ha, relative to AESRD (2015) guidelines. Prior work at AOI 7 (Wapisiw Lookout) predicted the occurrence of 9.48 ha of wetlands of three different classes (6.73 ha of marsh, 2.37 ha of shallow open wetland, and 0.38 ha of unspecific wetland; Novoa and Hawkes 2021). A field program was used to validate the results of the desktop exercise and to inform the development of a machine learning model that could be applied to other reclaimed landforms on Suncor's Base Plant. The desktop analysis resulted in thousands of wetland datapoints for Wapisiw Lookout (Novoa and Hawkes 2021). To make these data manageable and permit sample size and location selection, the points were discretized (binned) using a majority filter to create a sample of equal-sized hexagons of ~ 6.5m^2. The boundaries of contiguous hexagons were then dissolved to create homogeneous polygons associated with wetlands and upland habitat. The number of dissolved polygons associated with each of the four classes of interest (marsh, shallow open water, unspecified wetland, and upland) were counted and the percentage of each was used to generate the field sample. A random sample of 104 points was selected with the following distribution: 79 marsh, 11 shallow open water, 9 unspecified, and 5 upland (Fig. 2).

Field-level verifications of predicted wetlands at AOI 7 were completed in July 2023. July was selected for the fieldwork as this was most consistent with the seasonal timing of remote-imagery capture (28 July 2022). Wetland vegetation was also well developed at this time. A high-precision GPS unit (Geode GNS3 Single Band Receiver with sub 30 cm accuracy, Juniper Systems Inc.), pre-loaded with shapefiles of the predicted wetland polygons, was used to navigate to each sample location. Once at the pre-determined location, the field crew surveyed the site to identify the plant communities present, including any obvious wetland communities dominated by hydrophytic vegetation. Based on this survey, the site was classed as either a non-wetland or wetland as per AESRD (2015). If a designation of non-wetland was made (resulting in a 'false positive' finding for that location), the reason for this designation was noted and reference photos were taken. If wetland conditions were noted, the wetland was classified to class, form, type, and permanency type using keys in the *Alberta Wetland Classification System* (AESRD 2015) and

Table 1 Areas of Interest label, name, area (ha), reclamation year, and habitat target for areas assessed for wetlands and wetland classes on Suncor's base plant. Rec Year refers to reclamation year. See Fig. 1 for the distribution of AOIs. *Field validation data not available for these areas

Area of Interest	Name	Area (ha)	Rec Year	Target
AOI-2	MD 2	143.17	2012	Upland
AOI-3	MD 8	201.04	2011	Upland
AOI-4	MD 5	175.42	2009	Upland
AOI-5	SE Dump	42.66	2007	Upland
AOI-6*	Dyke 10	24.27	2015	Upland
AOI-7	Wapisiw Lookout	195.38	2010	Upland
AOI-9	North Dump	25.9	2015	Upland
AOI-10*	Comp Pond	5.69	2010	Upland
AOI-11	North Steepbank	395.95	2015–19	Upland
	Total Area	1,209.48		

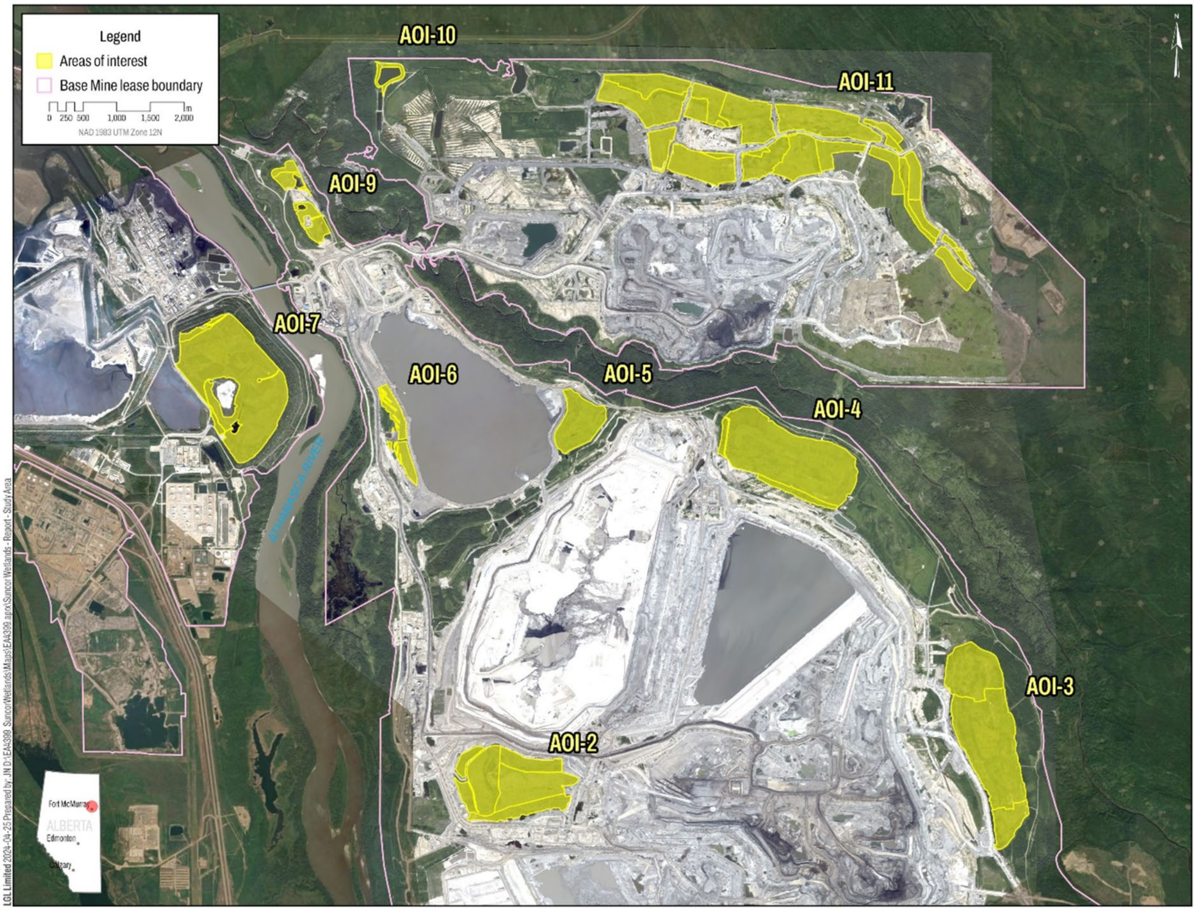

Fig. 1 The distribution of landforms (i.e., the Areas of Interest, AOIs) reclaimed to an upland forest type on Suncor's base plant. See Table 1 for AOI details

Guide for Assessing Permanence of Wetland Basins (Alberta Environment and Parks (AEP) 2014).

The wetland extent was mapped by circumnavigating the wetland perimeter while recording a precise GPS track of the wetland boundary. At representative points along the putative wetland perimeter, vegetation and soil assays were completed as necessary (see below) to identify the point at which the wetland ended and the upland habitat began. During this rapid assessment, the following vegetation indicators (adapted from Government of Alberta [2015]) were used as visual clues in assessing whether the chosen sample point was located within the wetland proper: (1) facultative wetland or obligate species account for over half of the abundant species in the plot; (2) surface encrustations of algae are present; (3) presence of a dominant groundcover of peat mosses (*Sphagnum* spp.), and (4) evidence of morphological adaptations of plants to saturated conditions (e.g., floating leaves, inflated stems, adventitious roots). In some cases, the pre-assigned verification point fell close to, but slightly outside, the field-delineated wetland boundary. At each of these 'false positive' points, a 5.64 m-radius (100 m^2) circular plot was established and a list of dominant plants recorded along with percent cover estimates. By way of contrast, a matching plot was established and sampled just downslope but within the putative wetland boundary. The habitat and vegetation information yielded by these comparative, closely adjacent point observations were later used to inform refinements to the machine learning model.

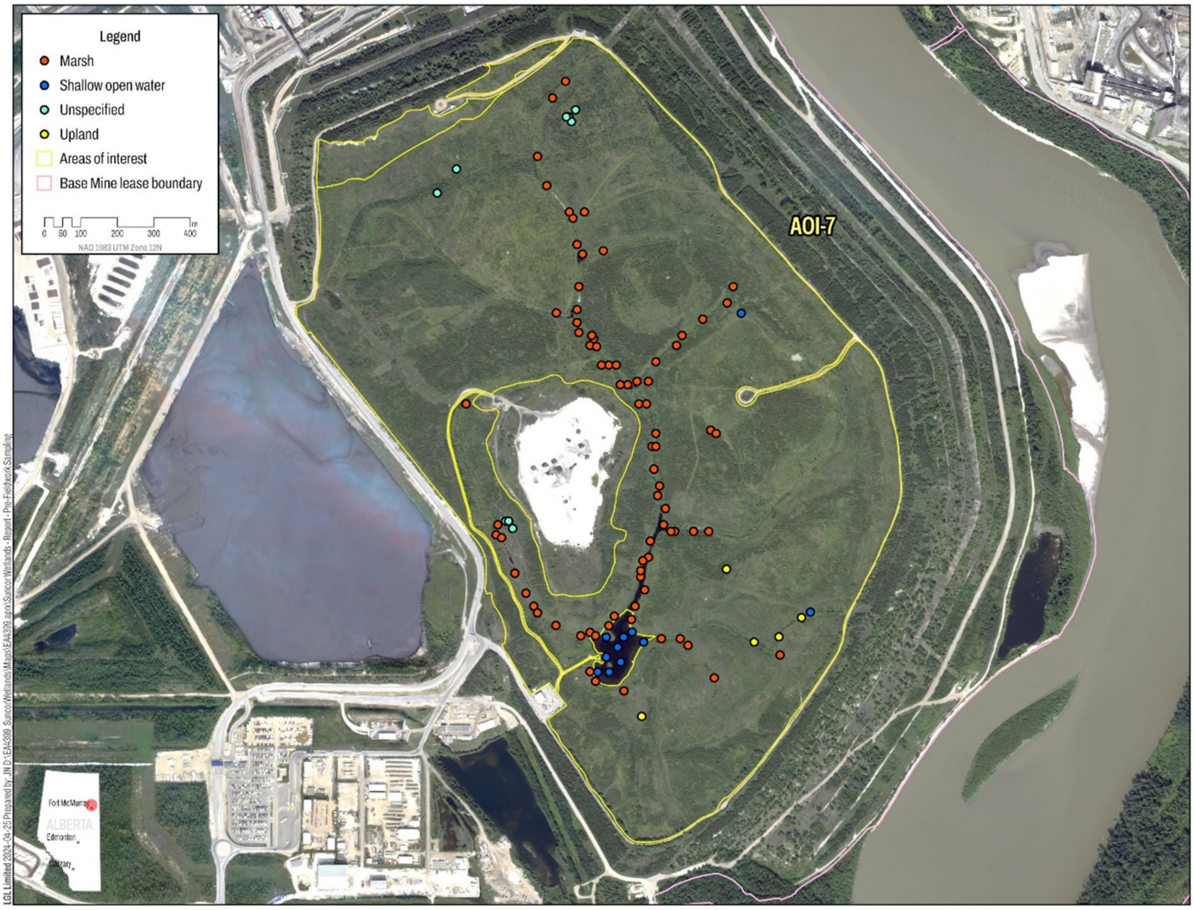

Fig. 2 Distribution of marsh, shallow open water, unspecified, and upland habitats selected for field validation at Wapisiw Lookout (AOI-7) on Suncor's base plant

Modelling

The wetland classification exercise used multiple topographic and satellite-derived variables to develop a random forest supervised machine learning model. The model was then executed for the AOI's on Suncor's base plant. The process used to delineate wetlands in each AOI was broken down into four main steps; manipulation of the raw data inputs, pre-processing the spatial input variables, training and tuning the model, and generating outputs including wetland delineations (i.e., polygons) (Fig. 3).

Data sources

A total of 21 predictor variables were extracted from three main sources, all captured in July 2022: (1) a Worldview-3 multi-spectral satellite image, (2) a LiDAR digital elevation model (DEM), and (3) a Synthetic Aperture Radar (SAR)-derived soil moisture index. The WorldView-3 multispectral image, acquired on 28 July 2022, with 1 panchromatic and 8 multispectral bands at 0.3 m and 1.2 m, respectively, was used to create several water-sensitive and vegetation indices [normalized difference soil

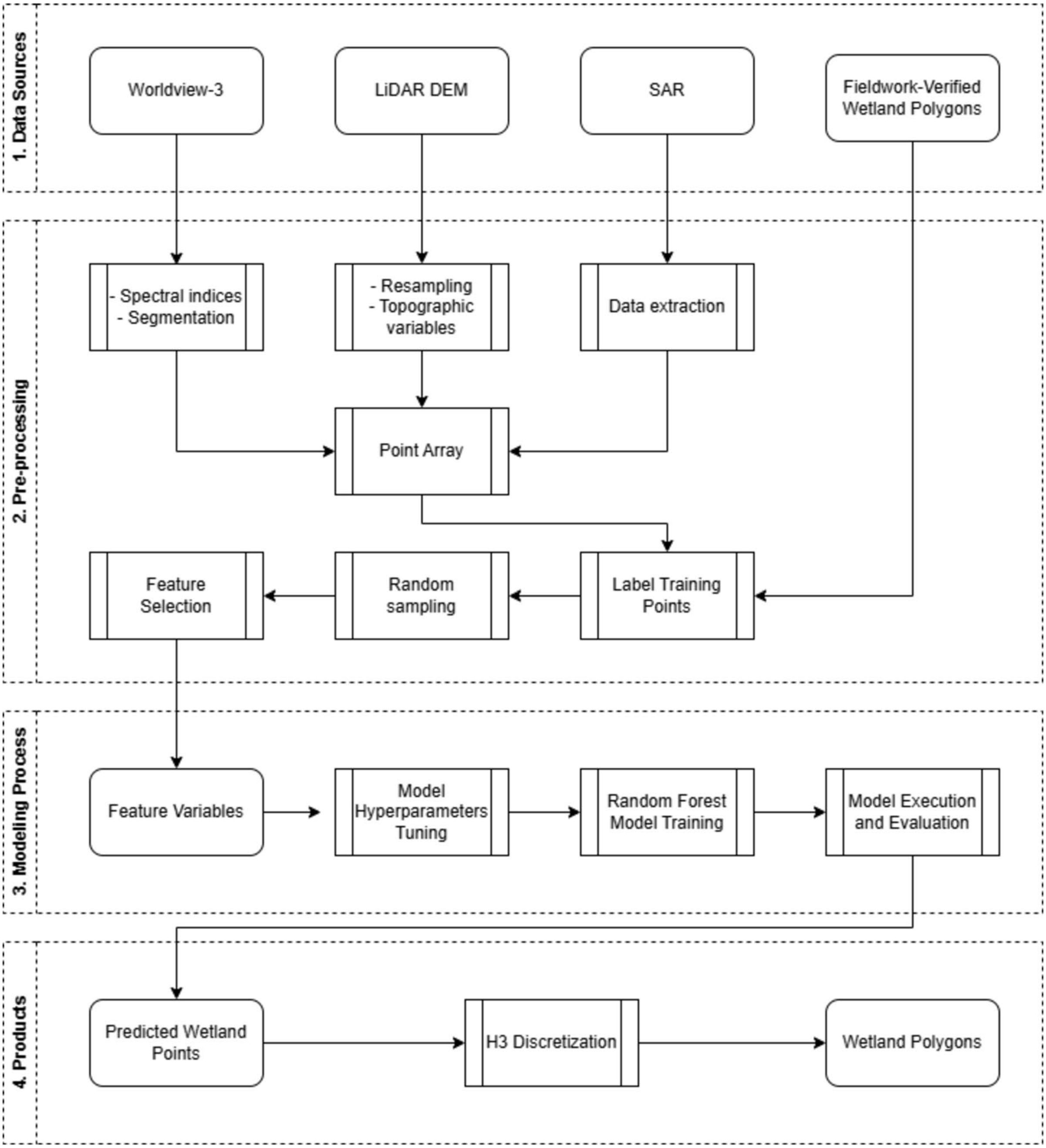

Fig. 3 Workflow associated with the delineation of wetlands on landforms reclaimed to an upland forest type on Suncor's base plant

index (NDSIWV), normalized difference vegetation index (NDVI) and normalized difference water index (NDWI)]. The spatial resolution of 1.2 m was used as the basis for all derived products to ensure all variables spatially matched. In addition to these three indices, an image segmentation variable was also calculated using the algorithm Segment Mean Shift in ArcGIS Pro. This variable is a discretized version of an RGB band combination that represents a more compact, visual representation of the image itself, reducing its inherent spectral variability. The segmentation variable was created using ArcGIS

Pro (Esri 2022) and the Python library scikit-image (van der Walt 2014). This process uses three satellite bands to group adjacent pixels that have similar spatial and spectral characteristics. The bands used for this process were: near-infrared 2 (*NIR2*), red-edge (*RE*), and yellow (*Y*), which are known to improve the discrimination among different vegetated aquatic and terrestrial habitats (Lane et al. 2014). Once the segmentation was calculated, the three resulting bands were transformed into a single grayscale band using the following formula:

Segmentation = 0.2125*R + 0.7154*G + 0.0721*B, where R, G, B are the previously segmented bands.

Once all variables were compiled, a point array was created. This layer consists of an array of locations, spaced at 1.2 m, matching the extent and location of the resampled datasets. Using this array, all variables were transformed from raster (i.e., TIFF files) into tabular form. Next, the training points were labelled using the field-verified wetland polygons obtained from two fieldwork campaigns conducted in 2017 (Miller et al. 2017) and 2023 (this program). Specifically, the points in the array that fell within the boundaries of the field-verified wetlands polygons were assigned a corresponding wetland class value. These labels or classes were then used to train the random forest supervised machine learning model. The points were labeled using the following numeric codes: 0 (non-wetland), 1 (marsh), and 2 (shallow open water).

A digital elevation model (DEM) at 0.5 m of spatial resolution was derived from LiDAR data obtained in June 2022. Using the software SAGA GIS (Conrad et al. 2022), the LiDAR DEM was aligned and resampled from 0.5 m to 1.2 m to match the extent and spatial resolution of the satellite image bands. The downsampled DEM was used to create the following topographic variables: channel network base level, channel network distance, channels distance, closed depressions, convergence index, downslope distance gradient, slope length and steepness factor (*LS factor*), plan curvature, profile curvature, relative slope position, slope (expressed in radians), total catchment area, valley depth, and Topographic Wetness index (Grabs et al. 2009). Nine of these topographic variables are visualized for AOI 7 (Wapisiw Lookout) in Fig. 4.

The synthetic aperture radar (SAR) layers were kept as individual TIFF files where no alignment or resampling operation was performed, due to their coarser spatial resolution of 10 m. The two layers, both of which provide measurements of soil surface moisture used in the model, were: average soil surface moisture chi-B (*AvgSSMchiB*) and average soil surface moisture chi-G (*AvgSSMchiG*). Soil surface moisture refers to the amount of moisture present in the top layer of soil, typically measured in terms of volumetric water content.

Data from previous field validations of wetland area (marsh and shallow open water) in each AOI were used to generate available areas for model training. These areas were transformed into a point grid at 1.2 m intervals from which a random sample could be obtained. The size of the study area, combined with the high spatial resolution of the input variables, creates an imbalanced dataset where the non-wetland class has a significantly larger number of samples than the other classes. This type of problem tends to create biased models where the minority classes suffer from high misclassification rates. When dealing with imbalanced training datasets, a crucial step in generating an accurate model is to apply a random undersampling approach to the majority classes. This technique helps to create a more balanced training dataset, which in turn leads to more accurate models when the minority class is of primary interest (Florath and Keller 2022). The training dataset comprised 1,000 samples from non-wetland areas, with an additional constraint that these samples were at least 50 m away from samples of marsh and shallow open water classes. The dataset also included 2,000 samples each from marsh and shallow open water classes, resulting in a total of 5,000 samples. A total of 5,000 points were selected for model training to strike a balance between generating a model that either overestimated total wetland area (i.e., > 5,000 points) or one that could not detect wetlands in a reliable manner (i.e., < 5,000 points).

Once the array of points was populated with all model attributes, model accuracy and stability were assessed relative to all variables and a reduced set of variables. A mix of a correlation analysis and a visual verification of the distribution of each variable was used to determine which variables could be removed from the model. The Pearson standard correlation coefficient was calculated and used to remove highly correlated variables. Next, boxplots of the distribution of values of each variable relative

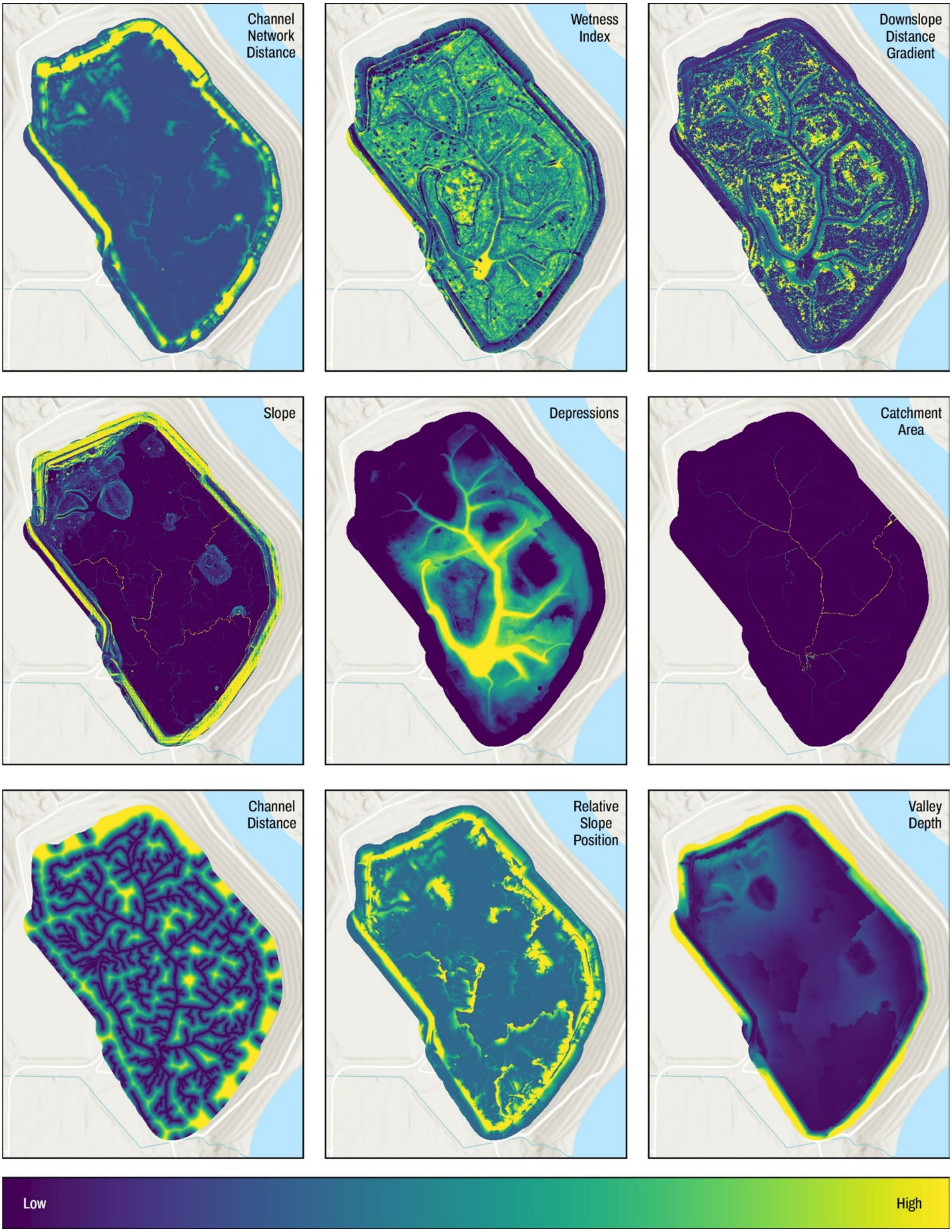

Fig. 4 Visualization of nine topographic variables for AOI 7 (Wapisiw Lookout) extracted from the 0.5 m LiDAR point cloud

to non-wetland, marsh, and shallow open water wetlands were visually evaluated and variables with low discriminatory power were removed (i.e., where heterogeneity of variance was high). Lastly, because the inclusion of correlated predictor variables can lead to unstable estimates and variance inflation (Hastie et al. 2005), Variance Inflation Factors (VIF) of individual variables were examined for potentially strong contributions to multicollinearity (Marcoulides and Raykov 2019) and predictor variables with VIF value > 5 were removed. The retained variables were those that could be used by any model and were not removed as an arbitrary means to reduce potential collinearity between variables (e.g., O'brien 2007). The combination of Pearson correlation coefficients, VIF values, and visual observations of boxplots for heterogeneity among classes supported the data reduction process used to refine the variables considered in the final model used to predict and classify wetlands (Table 2).

Modelling process

Two random forest models were developed to predict the occurrence and distribution wetlands at two levels: (1) wetland vs. non-wetland (to generate a prediction of all wet areas on reclaimed landforms); and (2) wetland class (with specific emphasis on marsh and shallow open water wetland classes). Random forest models are often used in prediction analyses due to their increased accuracy and resistance to multicollinearity compared to linear regression (Hastie et al. 2005). The random forest model – specifically the bootstrapped aggregation of several regression trees – is an ensemble learning method to predict an outcome (Breiman 2001; Brokamp et al. 2017). Outputs of this model include useful accuracy metrics and reports to define the importance of each variable used in the modelling process. Model input is a point feature class where each column is a variable, one column holds the class labels (i.e., wetlands/non-wetland), and optionally one column holds a flag for constraining the sampling area (i.e., no-sample/sample). The model is flexible in terms of the number of variables and their spatial resolution, both of which can be modified. If new variables are available, they can be effortlessly incorporated into the model. All variables in the model must be stored as Float type, while the sampling area constraint and the labels columns must be stored as Integer type. At present the model only supports continuous numerical variables.

Table 2 The satellite bands (visible and Infrared [IR]), their codes, spectral properties, and resolution used to predict the occurrence and distribution of opportunistic wetlands on Suncor's base plant

Data source	Variable name
SAR	AvgSSMchiG
LiDAR DEM	Channels distance
	Closed depressions
	Downslope distance gradient
	Channel network base level
	Valley depth
	Wetness index
Worldview-3	NDWI
	NIR1
	Segmentation

Model optimization was accomplished through a grid search, which systematically evaluated nine distinct hyperparameter combinations across tenfold cross-validation resulting in a total of 90 model fits. This exhaustive search enabled the identification of the best-performing hyperparameter settings for the random forest model. The weighted average F1-score of the wetland vs. no-wetland model was 84%. The training accuracy of 0.83 and validation accuracy of 0.83 demonstrates the model's ability to generalize well to new data.

For the wetland classification model (marsh, shallow open water, and non-wetland) the weighted average F1-score of 0.89 achieved through this process indicates a remarkable balance between precision and recall, demonstrating the effectiveness of the hyperparameter tuning in optimizing the model's performance. 'Precision' quantifies the number of positive class predictions that actually belong to the positive class. 'Recall' quantifies the number of positive class predictions made from all positive samples in the dataset. The F1-score provides a single score that balances both the concerns of precision and recall in one number (Brownlee 2020). Using the optimal hyperparameters identified, the subset of selected variables, and the 5,000 randomly selected samples, a random forest model was trained and evaluated. The dataset was split into training and validation sets using a 90/10 ratio, allowing for a robust assessment of the model's performance. The results of the training were satisfactory, as evidenced by the strong performance across all accuracy metrics. Specifically, the model achieved an average F1-score of 0.88, indicating a good balance between precision and recall. The model's ability to generalize to new data was further evidenced by its training accuracy of 0.94 and validation accuracy of 0.87. These results suggest the random forest model was well-suited for the task at hand and can be relied upon to make accurate predictions.

Following the successful training of the model, a prediction was generated for the entire dataset, comprised of 13,120,508 points. A new point array was created, augmented with a column containing the assigned class labels. While the model achieved an overall accuracy of 0.87, a closer examination of the results reveals a more nuanced picture. Notably, the weighted average F1-score of 0.19 for marsh and shallow open water classes suggests that the model struggled to balance precision and recall, potentially indicating difficulties in distinguishing between classes when applying the model to an unknown dataset.

Products

In the final stage of the wetland prediction process, the predicted points were discretized using the H3 hierarchical hexagonal grid system (Amirpour et al. 2020). This global grid was used at a resolution of Level 14, which corresponds to a spatial granularity of approximately 6.5m^2 per hexagon. By applying spatial analysis operations, the dominant class (i.e., marsh or shallow open water) within each hexagon was identified and subsequently assigned to the respective hexagonal unit. This step enabled the aggregation of predicted points into a cohesive and spatially explicit representation of wetland areas, ultimately facilitating the creation of the predicted wetland polygons.

Results

Field validation

Of the 104 samples selected for field validation, 90 were predicted to be wetlands: 79 consisting of marsh and 11 consisting of shallow open water (SOW). Field validation confirmed a total of 74 wetlands, with 46 classified as marsh and 28 as SOW (Table 3). Some predicted wetlands were reclassified as upland habitat, while some predicted uplands were reclassified as wetlands. Likewise, some marshes were reclassified during fieldwork as SOW, whereas only one SOW was incorrectly labeled as such by the model (Table 3).

For marsh predictions:

- Of the 79 predicted marsh samples, 43 were verified as marsh (54.4%).
- 18 were reclassified as SOW (22.3%).
- Another 18 were reclassified as upland (22.3%).

For SOW predictions:

- Of the 11 predicted SOW samples, 10 were verified as SOW (90.9%).
- One was reclassified as upland.

For other categories:

- All 9 unspecified samples were reclassified as upland.

Table 3 Number and type of wetland features predicted by the 2023 model vs. those identified during field validation work on Suncor's base plant

Category	Predicted (n)	Verified (n)	Reclassified (n)	Accuracy (%)
Marsh	79	43	18 to SOW, 18 to upland	54.4
Shallow Open Water (SOW)	11	10	1 to upland	90.9
Unspecified Wetland	9	0	9 to upland	N/A
Upland	5	2	3 to marsh	40.0
False Positives (Wetland)	29	-	15 within 10 m of wetland	~50.0
Wetlands Predicted Correctly	90	71	28 misclassified as upland	78.9
Uplands Predicted Correctly	5	2	3 misclassified as wetlands	40.0

- Of the 5 predicted upland samples, 2 remained as upland (40%), and 3 were reclassified as marsh (60%).

Overall, 71 of the 90 samples predicted as wetlands (78.9%) were verified as wetlands. However, 28 of the 90 samples (31.1%) predicted to be wetlands were actually uplands. Conversely, 3 of the 5 points predicted as upland (60%) were verified as wetlands (Table 3).

Among the 29 false positives (wetlands incorrectly predicted), 15 (~50%) were within 10 m of an actual wetland, often capturing the riparian shrub zone, which can be ambiguous in terms of wetland classification.

The remaining 14 false positive points (48%) fell clearly within upland habitat (e.g., jack pine stands) having no evident wetland characteristics. When practicable, these habitats were delineated with a GPS track as non-wetland polygons. A few mapped polygons did not capture the full extent of the targeted wetland feature; the boundaries of these polygons were redrawn in the field to indicate their actual size. As well, one previously unmapped minor wetland feature (temporary marsh) was noted by the surveyors as they traversed the Wapisiw catchment. These observations were used as false negative data points to further train the remote sensing model.

Wetland vs. not wetland

The total area of each AOI predicted to be wetland (of any class, i.e., marsh, shallow open water, transitional shrub, swamp, or unspecified) varied from ~ 13 and 22.5% for a total area of 211.1 ha (17.9% of all AOIs assessed), which is consistent with previously reported results (i.e., Hawkes et al 2020; Table 4). The total area of non-wetland habitat assessed in 2023 was 968.42 ha or 82.1% of the AOIs assessed (Table 4). Figure 5 shows the distribution of predicted wetlands (any class) for all AOIs along with examples of predicted wetland distribution at two AOIs considered in this study.

Wetland classification

The 3-class model (marsh, shallow open water, non-wetland) predicted the occurrence and distribution of 56.79 ha of marsh (53.2 ha) and SOW (3.59 ha) habitat across all AOIs (Table 5). Some of the areas used for training were captured in these predicted areas. Adding the additional training areas associated with field data obtained in 2017 and 2023 (i.e., the training data not predicted by the model) increased the total estimated area of marsh to 61.72 ha and SOW to 4.64 ha (Table 5), for a total of 66.36 ha of wetland habitat in two classes (i.e., marsh and SOW). Figure 6 shows the distribution of predicted wetlands (by class: marsh and shallow open water) for all AOIs along with examples of predicted wetland distribution at two AOIs considered in this study.

Table 4 Area of marsh and shallow open water (SOW) habitat predicted by the 2023 model and 2017 field validation work per area of interest (AOI) on Suncor's base plant

AOI	Name	AOI Area (ha)	2023 wetland		2023 upland		Hawkes et al. (2020)	
			ha	% AOI	Ha	% AOI	ha	% AOI
AOI-2	MD 2	143.17	18.8	13.13	124.37	86.87	32.1	22.42
AOI-3	MD 8	201.04	33.8	16.81	167.24	83.19	43.8	21.79
AOI-4	MD 5	175.42	21.7	12.37	153.72	87.63	19.1	10.89
AOI-5	SE Dump	42.66	5.2	12.19	37.46	87.81	6.1	14.30
AOI-6*	Dyke 10	24.27	na	na	na	na	na	na
AOI-7	Wapisiw Lookout	195.38	43.9	22.47	151.48	77.53	12.1	6.19
AOI-9	North Dump	25.9	5.8	22.39	20.10	77.61	1.3	5.02
AOI-10*	Comp Pond	5.69	na	na	na	na	na	na
AOI-11	North Steepbank	395.95	81.9	20.68	314.05	79.32	95.5	24.12
		1,179.52	211.1	17.90	968.42	82.10	210	17.80

*Total area excludes AOIs-6 and 10 due to a lack of training data available with which to predict the occurrence and distribution of wetlands

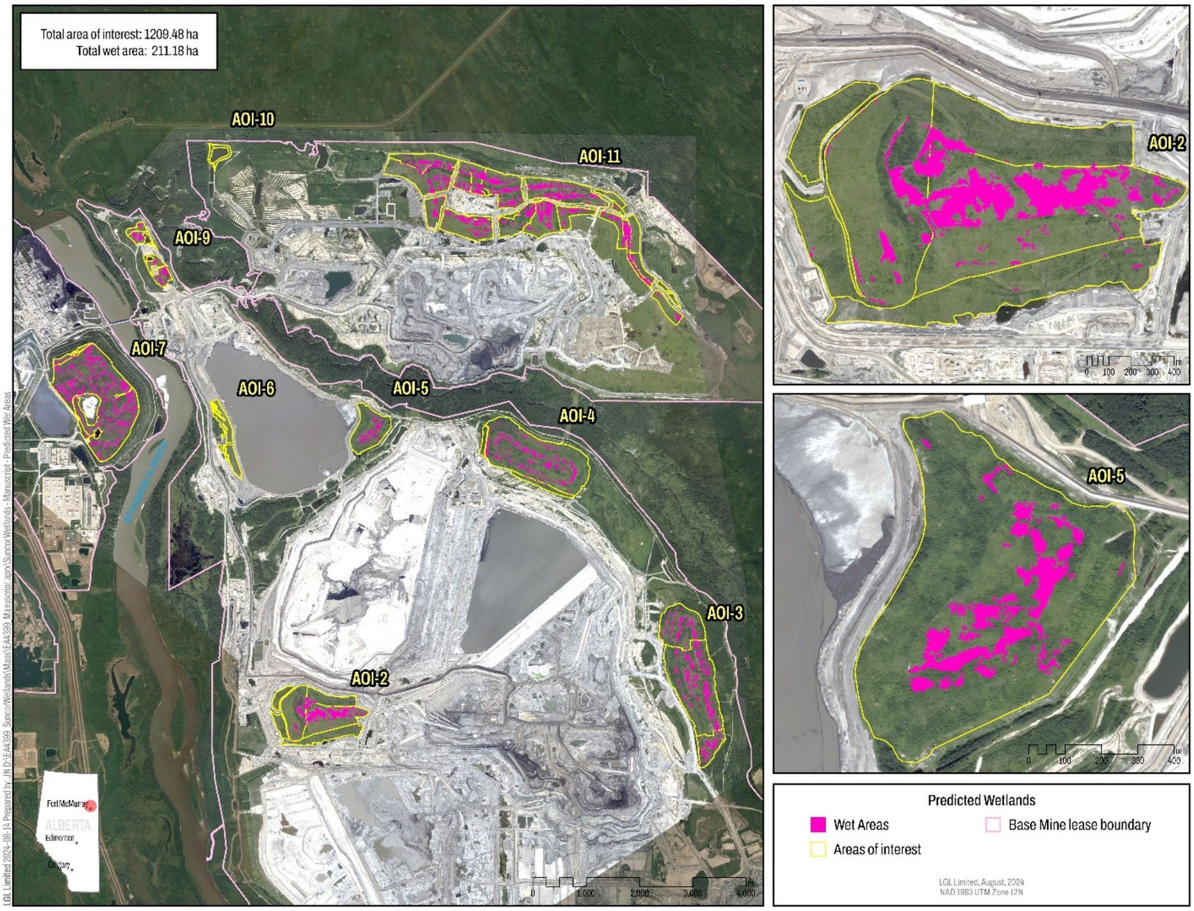

Fig. 5 Overview of predicted wetland distribution (any class) for all areas of interest on Suncor's base plant and at AOIs 2 and 5

Table 5 Area of marsh and shallow open water (SOW) habitat predicted by the 2023 model and additional wetland areas associated with field data collected in 2017 and 2023 by area of interest (AOI) on Suncor's base plant

			Marsh area (ha)			Shallow open water area (ha)		
Area of Interest	Name	AOI Area (ha)	2023 Model	Training	Total	2023 Model	Training	Total
AOI-2	MD 2	143.17	0.86	0.82	1.68	0.03	0.16	0.19
AOI-3	MD 8	201.04	9.57	1.47	10.58	0.29	0.26	0.44
AOI-4	MD 5	175.42	3.34	3.48	6.82	0.25	0.34	0.59
AOI-5	SE Dump	42.66	1.06	0.27	1.33	0.22	0	0.22
AOI-6*	Dyke 10	24.27	na	na	na	na	na	na
AOI-7	Wapisiw Lookout	195.38	19.47	0.82	19.77	0.49	1.62	0.54
AOI-9	North Dump	25.9	0.36	0.28	0.64	0.11	0.02	0.13
AOI-10*	Comp Pond	5.69	na	na	na	na	na	na
AOI-11	North Steepbank	395.95	18.54	2.49	20.9	2.2	0.33	2.53
	Total Area (ha or %)	1,179.52	53.2	9.63	61.72	3.59	2.73	4.64

*Total area excludes AOIs-6 and 10 due to a lack of training data available with which to predict the occurrence and distribution of wetlands

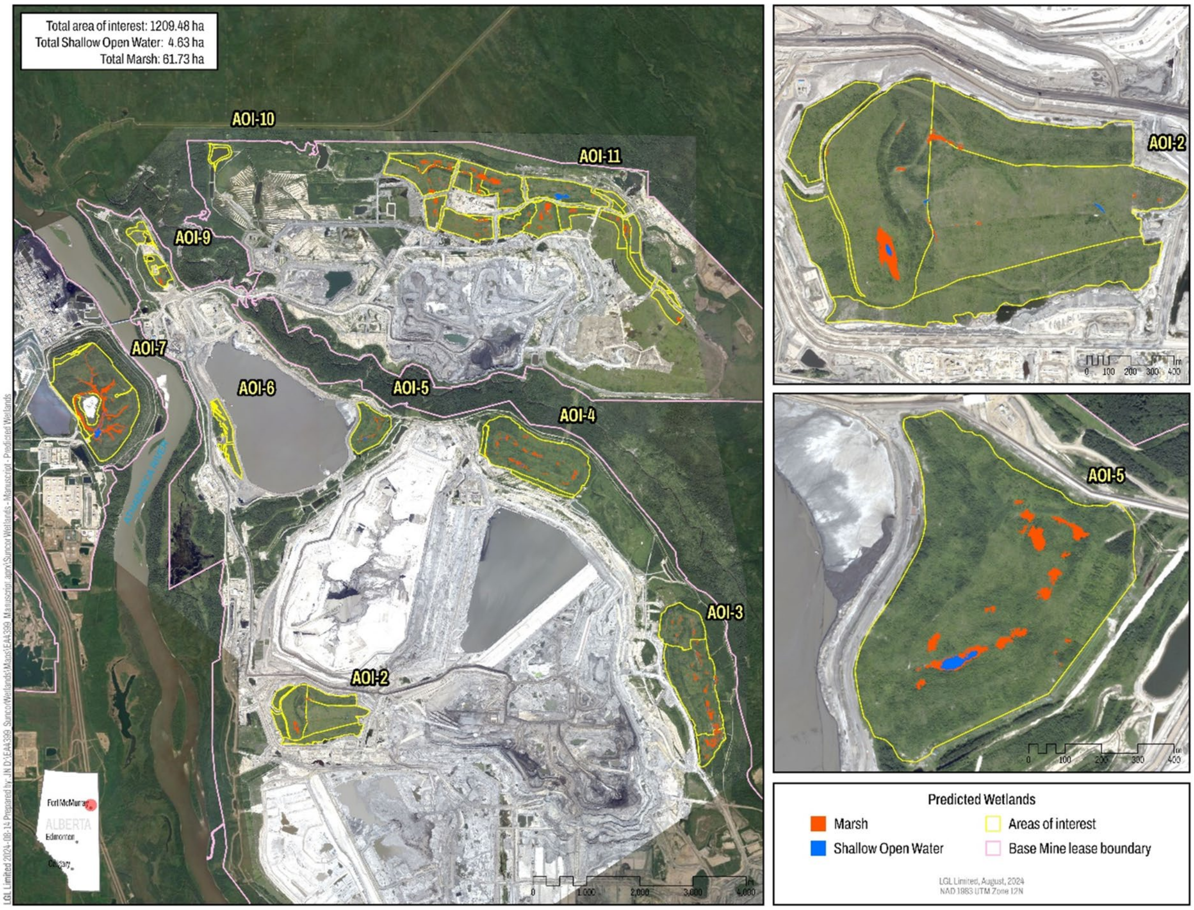

Fig. 6 Overview of predicted wetland distribution (by class: marsh and shallow open water) for all areas of interest on Suncor's base plant and at AOIs 2 and 5

Discussion

The development of opportunistic wetlands on landforms reclaimed to an upland forest type in the Athabasca Oil Sands Region has only recently received attention (Little-Devito et al. 2019; Hawkes et al. 2020, Hawkes and Novoa 2021, Zakharov et al. 2024). Multiple mechanisms for opportunistic wetland development have been postulated. These include the uneven placement of reclamation soils resulting in microsite heterogeneity (Gringas-Hill et al. 2018) and the promotion of wetter conditions for wetland plant development (Trites and Bayley 2009); placement of fine-textured top soils resulting in elevated surface saturation (Little-Devito et al. 2019); occurrence of lateral groundwater flows from surrounding uplands (Price et al. 2010); development of depressional areas with groundwater influence (Little Devito et al. 2019); and elevation gradients within the surrounding catchment (Hawkes et al. 2020). Our study, in contrast, uses remote sensing attributes and machine learning to detect and classify wetland areas, based on the view that underlying physical factors and mechanisms like soil texture, landscape position, or ecohydrological feedback will be best understood once existing wetlands have first been delineated and classified.

Heterogeneity in wetland density or likelihood is because not all landforms reclaimed to an upland forest ecosite are created the same. For example, Wapisiw Lookout (AOI-7) is established on the top of what was previously a tailings pond while most other AOIs were created using either tailings sand of finer textured overburden material, which is then covered by up to 2 m of suitable reclamation soil material (Pinno and Hawkes 2015). The composition and depth of soils placed on reclaimed landforms varies as a function of year of reclamation, availability of material, planned ecosite for the landform, and even operator experience. Collectively these factors can introduce the irregularities and heterogeneity that allow for opportunistic wetland creation but cannot be accurately recorded during placement or easily mapped post-placement, making spatial prediction of opportunistic wetlands difficult based solely on the operational history of the reclamation landscape. The present work provides a useful roadmap for using remote sensing to identify, *post-hoc*, wetland potential after reclamation work is complete.

Because of the potential to consider the inclusion of opportunistic wetlands in reclamation and closure planning, the value of wetlands to wildlife, and the potential to manipulate current reclamation practices to promote the establishment and persistence of wetlands on reclaimed landforms, a more accurate, robust, and repeatable method to predict and classify opportunistic wetlands is of substantive benefit to oil sands operators. Further, the rapid assessment of wetland areas investigated here can provide annual results that allow for investigation of wetland permanence, persistence and extent through time, which is relevant for achieving long-term reclamation goals. Finally, the increased area of wetlands on the closure landscape will better approximate 'equivalent land capability' in relation to the pre-existing landscape, and better integration with a natural surrounding landscape, as outlined in the regulatory guidelines.

Previous work on the occurrence, classification, and quantification of opportunistic wetlands was accomplished using a combination of desktop (GIS) wetland delineation and fieldwork (Miller et al. 2017). Here, the machine learning models based on topographic and spectral variables confirmed that ~ 18% (211 ha) of the upland-reclaimed area assessed develops instead to wetland (Hawkes et al. 2020). Two random forest models were ultimately developed that predicted wetland occurrence in each AOI at two levels: (1) wetland vs. non-wetland (to generate a prediction of all wet areas on reclaimed landforms); and (2) wetland class (with specific emphasis on marsh and shallow open water wetland classes). We found that the models accurately predict the occurrence of wetlands themselves (i.e., distinguish wetland and non-wetland) and often—though not always—correctly distinguish marsh from shallow open water. These results were consistent across multiple different reclaimed landforms assessed at Suncor's Base Plant with the random forest models handling the variability in reclamation approach, substrate type, and soil placement depth with a high degree of accuracy, implying that if the models are applied to novel areas (and the same variables used), they will successfully predict both the occurrence of wetland habitat and, to a lesser degree, the wetland form.

Consistent with the literature (He and Garcia 2009; López et al. 2013), our modeling process revealed a common phenomenon in imbalanced classification problems, which is that the model tends to achieve a low precision and high recall for the minority classes. While Precision measures the proportion of true positives among all positive predictions made by the model, Recall measures the proportion of true positives among all actual positive instances in the dataset. In an imbalanced dataset, achieving a high Recall for the minority class is essential, as it ensures the model can identify a substantial proportion of the minority class instances. Conversely, a low Recall for the minority class would lead to a high number of false negatives, which can have undesirable consequences, such as failing to predict wetlands where they actually exist. Moreover, over 80% of locations predicted to be a wetland were assigned to a wetland class in the field. Model performance was not as strong with respect to classifying wetlands but is able to discriminate (some of the time) between marsh and shallow open water with the former more often correctly assigned. One of the primary outcomes of this work is the need to validate the predictions made by the model. The ability to predict wetland occurrence is high, but field validation is needed to ensure the spatial extent and wetland class is accurate.

Our work continues to demonstrate that the development and application of these models can be used in a multi-temporal analysis to monitor and understand the evolution of opportunistic wetlands on Suncor's base plant (and elsewhere) through the use of topographic and spectral datasets associated with past or future dates. In a similar way, an understanding of how terrain modifications (e.g., soil settlement, slumping, and erosion) affect the formation and evolution of opportunistic wetlands is possible using a model such as the one created here by simulating different landform creation practices, revegetation strategies, and successional trajectories on reclaimed landforms. The utility of these models to define and delineate wetlands could be invaluable when it comes to reclamation and closure planning and when considering metrics of success associated with reclamation in the AOSR. Additional consideration should be given to the relatively early developmental stage of these wetlands, and how their classification may change over time as they advance successionally. For example, Mombourquette (2023) found that vegetative species richness significantly changed with age of reclaimed wetlands. Similarly, Borkenhagen et al. (2023) described rapid changes in the dominant vegetation of constructed wetlands over the course of 10 years. As such, we anticipate that many of these mineral-soil wetlands will likely transition to peat-based organic wetlands, which is based in part on observations of initial peat formation noticed at several of the wetlands visited during the field validation program.

Conclusion

Our modelling approach validates results in Hawkes et al. (2020) and provides enhanced resolution and improved accuracy in classifying marsh and shallow open water wetlands on reclaimed landforms originally designed to support upland forest ecosystems. This advancement demonstrates the capacity of machine learning GIS models to recognize and delineate opportunistically forming wetlands with speed and precision, enabling the monitoring of multiple wetland features across thousands of hectares at regular intervals. Such efficiency supports a deeper understanding of wetland ecological performance, resilience, and persistence in the context of large-scale reclamation projects.

An important future application of this approach is the quantification of the spatial and temporal persistence of opportunistic wetlands on reclaimed landscapes. By reliably tracking the total wetland area and comparing it against closure and reclamation objectives, this method offers a robust tool for assessing reclamation success and informing adaptive management practices. Additionally, the ability to monitor changes over time will provide valuable insight into how these ecosystems evolve, particularly under the influence of climate change and other environmental stressors.

The development of opportunistic wetlands indicates a spontaneous diversification of the reclaimed landscape, transitioning it toward a more dynamic and naturalized closure state. By emulating the pre-mining landscape, where frequent wetland formation was a dominant feature, such spontaneous developments represent an essential step in the ecological maturation of reclaimed mine sites within the Athabasca Oil Sands Region. By fostering conditions that support wetland formation and persistence, reclamation efforts have the potential to not only promote biodiversity and habitat complexity but also to nudge successional processes along the desired trajectory. As such, the integration of machine learning tools with reclamation practices can be an important step in ensuring that reclaimed landscapes are resilient, functional, and aligned with long-term ecological and social objectives.

Acknowledgements This project was compelted by LGL Limited under contract to Suncor Energy Inc. We thank Bryce McKinnon and Shaan Aroeste for assistance with fieldwork.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by VH, MM and JN. The first draft of the manuscript was written by VH, JH, JN, MM, and SK and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was funded by Suncor Energy Inc. through contracts to LGL Limited environmental research associates.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing Interests The authors declare no competing interests.

References

Alberta environment and parks (AEP) (2014) Guide for assessing permanence of wetland basins. Land Policy Branch, Alberta Environment and Parks, Edmonton, Alberta, Canada. 28 pp.

Alberta environment and parks (AEP) (2018) Alberta merged wetland inventory. [Online.] Available from https://maps.alberta.ca/genesis/rest/services/Alberta_Merged_Wetland_Inventory/Latest/MapServer/

Alberta environment and parks (AEP) (2023) Age of permanent reclamation in the mineable oil sands region. [Online] Available from http://osip.alberta.ca/library/Dataset/Details/28 [accessed August 13 2024]

Alberta environment and sustainable resource development (AESRD) (2015) Alberta wetland classification system. Water Policy Branch, Policy and Planning Division, Edmonton, AB

Amirpour A, Lang M, Wenzel P (2020) H3: A hexagonal hierarchical spatial index. Proceedings of the 28th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 1–4. https://doi.org/10.1145/3397533.3397536

Bohnett E, Hulse D, Ahmad B, Hoctor T (2020) Multi-level, multi-scale modeling and predictive mapping for jaguars in the Brazilian Pantanal. Open J Ecol 10:243. https://doi.org/10.4236/oje.2020.105016

Borkenhagen A, Cooper DJ, House M, Vitt DH (2023) Establishing peat-forming plant communities: a comparison of wetland reclamation methods in alberta's oil sands region. Ecol Appl 20023:e2929. https://doi.org/10.1002/eap.2929

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Brokamp C, Jandarov R, Rao MB, LeMasters G, Ryan P (2017) Exposure assessment models for elemental components of particulate matter in an urban environment: a comparison of regression and random forest approaches. Atmos Environ 151:1–11. https://doi.org/10.1016/j.atmosenv.2016.11.066

Brownlee J (2020) Imbalanced classification with python: better metrics, balance skewed classes, cost-sensitive learning. Machine Learning Mastery.

CEMA (Cumulative Environmental Management Association) (2014) Guideline for wetland establishment on reclaimed oil sands leases, 3rd ed. Alberta: Prepared by the Wetlands and Aquatics Subgroup of the Reclamation Working Group of the Cumulative Environmental Management Association.

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J (2022) SAGA GIS (Version 9.0.0) [Computer software]. University of Göttingen, Germany. https://www.saga-gis.org/

Esri (2022) ArcGIS Pro (Version 3.2), Spatial Analyst [Computer software]. Redlands, CA: Esri.

Florath J, Keller S (2022) Supervised machine learning approaches on multispectral remote sensing data for a combined detection of fire and burned area. Remote Sensing 14(3):657

Gingras-Hill T, Nwaishi FC, Macrae ML, Price JS, Petrone RM (2018) Ecohydrological functioning of an upland undergoing reclamation on post-mining landscape of the Athabasca oil sands region. Canada Ecohydrology 11(3):e1941. https://doi.org/10.1002/eco.1941

Gosselin P, Hrudey SE, Naeth MA, Plourde A, Therrien R, Van Der Kraak G, Xu Z (2010) Environmental and health impacts of Canada's oil sands industry. Royal Society of Canada

Government of Alberta (2021) Environmental protection and enhancement act. revised statutes of alberta 2000. Chapter E-12. Current as of June 2 2021. Published by Alberta Queen's Printer.

Grabs TJ, Seibert J, Bishop T, Laudon H (2009) Modeling spatial patterns of saturated areas: a comparison of the topographic wetness index and a dynamic distributed model. J Hydrol 373:15–23

Hastie T, Tibshirani R, Freidman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. Math Intell 27:83–85

Hawkes VC, Gerwing TG (2019) Wildlife usage indicates increased similarity between reclaimed upland habitat and mature boreal forest in the athabasca oil sands region of alberta. Canada Plos One 14:e0217556

Hawkes VC, Gibeau P, Challenger W (2024) Making sense of wildlife habitat use on active oil sands mines: quasi-experiments, occupancy models, trends assessments, and upland habitat reclamation. Ecol Solut Evid 5:e12367. https://doi.org/10.1002/2688-8319.12367

Hawkes VC, Miller MT, Novoa J, Ibeke E, Martin JP (2020) Opportunistic wetland formation, characterization, and quantification on landforms reclaimed to upland ecosites in the athabasca oil sands region. Wetlands Ecol Manage 28:953–970. https://doi.org/10.1007/s11273-020-09760-x

Hawkes VC, Hentze N, Gerwing TG (2021) Trends in avian use of reclaimed boreal forest habitat in Canada's oil sands. Avian Conservation and Ecology 16(2):5. https://doi.org/10.5751/ACE-01915-160205

Hawkes VC, Gibeau P, Challenger W (2024) Making sense of wildlife habitat use on active oil sands mines: quasiexperiments, occupancy models, trends assessments, and upland habitat reclamation. Ecol Solut Evid 5:e12367. https://doi.org/10.1002/2688-8319.12367

He H, Garcia EA (2009) Learning from imbalanced data. IEEE Trans Knowl Data Eng 21(9):1263–1284. https://doi.org/10.1109/TKDE.2008.230He

Ketcheson SJ, Price JS, Carey SK, Petrone RM, Mendoza CA, Devito KJ (2016) Constructing fen peatlands in post-mining oil sands landscapes: challenges and opportunities from a hydrological perspective. Earth-Sci Rev 161:130–139

Lane CR, Liu H, Autrey BC, Anenkhonov OA, Chepinoga VV, Wu Q (2014) Improved wetland classification using

eight-band high resolution satellite imagery and a hybrid approach. Remote Sensing 6(12):12187–12216
Little-Devito M, Mendoza AC, Chasmer L, Kettridge N, Devito KJ (2019) Opportunistic wetland formation on reconstructed landforms in a sub-humid climate: influence of site and landscape-scale factors. Wetl Ecol Manag 1–22:587–608
López V, Fernández A, García S, Palade V, Herrera F (2013) An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characters. Inf Sci 250:113–141. https://doi.org/10.1016/j.jns.2013.07.007
Marcoulides KM, Raykov T (2019) Evaluation of variance inflation factors in regression models using latent variable modeling methods. Educ Psychol Meas 79(5):874–882. https://doi.org/10.1177/0013164418817803
Miller MT, Novoa J, Hawkes VC, McKinnon B, Gerwing TG (2017) Wetland inventory of permanently reclaimed sites at suncor oil sands base plant. Unpublished report by LGL Limited environmental research associates, Sidney, BC for Suncor Energy, Calgary, Alberta. 35 pp + App.
Mirmazloumi SM, Moghimi A, Ranjgar B, Mohseni F, Ghorbanian A, Ahmadi SA, Amani M (2021) Brisco B (2021) status and trends of wetland studies in canada using remote sensing technology with a focus on wetland classification: a bibliographic analysis. Remote Sens 13:4025. https://doi.org/10.3390/rs13204025
Mombourquette AD (2023). Effects of landscape age and salinity on plant community composition and productivity in opportunistic and constructed wetlands in the Athabasca Oil Sands Region, Alberta (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca.
Natural Regions Committee (NRC) (2006) Natural regions and sub-regions of alberta. Compiled by DJ Downing and WW Pettapiece; Government of Alberta. Pub. No. T/852. https://www.albertaparks.ca/media/2942026/nrsrcomplete_may_06.pdf. Accessed 14 August 2019.
Novoa J, Hawkes VC (2021) Wetland delineation and classification of opportunistic wetlands at wapisiw lookout on suncor's base plant. Unpublished report by LGL Limited environmental research associates, Sidney, BC, for Suncor Energy, Fort McMurray, AB. 49 pp + Appendices.
O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Quality Quantity 41:673–690. https://doi.org/10.1007/s11135-006-901806
O'Malley WC, Elbroch LM, Zeller KA, Beier P, Beale MM, Beausoleil RA, Kerston B, Knopff K, Kunkel K, Maletzke BT, Martins Q, Mathcett MR, Wilmers CC, Wittmer HU, Vickers W, Sager-Fradkin K (2024) Robinson H (2024) Machine learning allows for large-scale habitat prediction of a wide-ranging carnivore across diverse ecoregions. Landsc Ecol 39:106. https://doi.org/10.1007/s10980-024-01903-2
Pinno BD, Hawkes VC (2015) Temporal trends of ecosystem development on different site types in reclaimed boreal forests. Forests 6:2109–2124. https://doi.org/10.3390/f6062109
Price JS, McLaren RG, Rudolph DL (2010) Landscape restoration after oil sands mining: Conceptual design and hydrological modelling for fen reconstruction. Int J Min Reclam Env 24:109–112
Shoemaker KT, Heffelfinger LJ, Jackson NJ, Blum ME, Wasley T, Stewart KM (2018) A machine-learning approach for extending classical wildlife resource selection analyses. Ecol Evol 8:3556–3569. https://doi.org/10.1002/ece3.3936
Trites M, Bayley SE (2009) Vegetation communities in continental boreal wetlands along a salinity gradient: implications for oil sands mining reclamation. Aquat Bot 91(1):27–39
Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T (2014) scikit-image contributors (2014) scikit-image: Image processing in Python. PeerJ 2:e453. https://doi.org/10.7717/peerj.453
Volik O, Elmes M, Petrone R, Kessel E, Green A, Cobbaert D, Pirce J (2020) Wetlands in the athabasca oil sands region: the nexus between wetland hydrological function and resource extraction. Environ Rev. https://doi.org/10.1139/er-2019-0040
Zakharov I, Kohlsmith S, Hornung J, Charbonneau F, Bobby P, Howell M (2024) Surface soil moisture estimation from time series of RADARSAT constellation mission compact polarimetric data for the identification of water- saturated areas. Remote Sens. 2024, 16