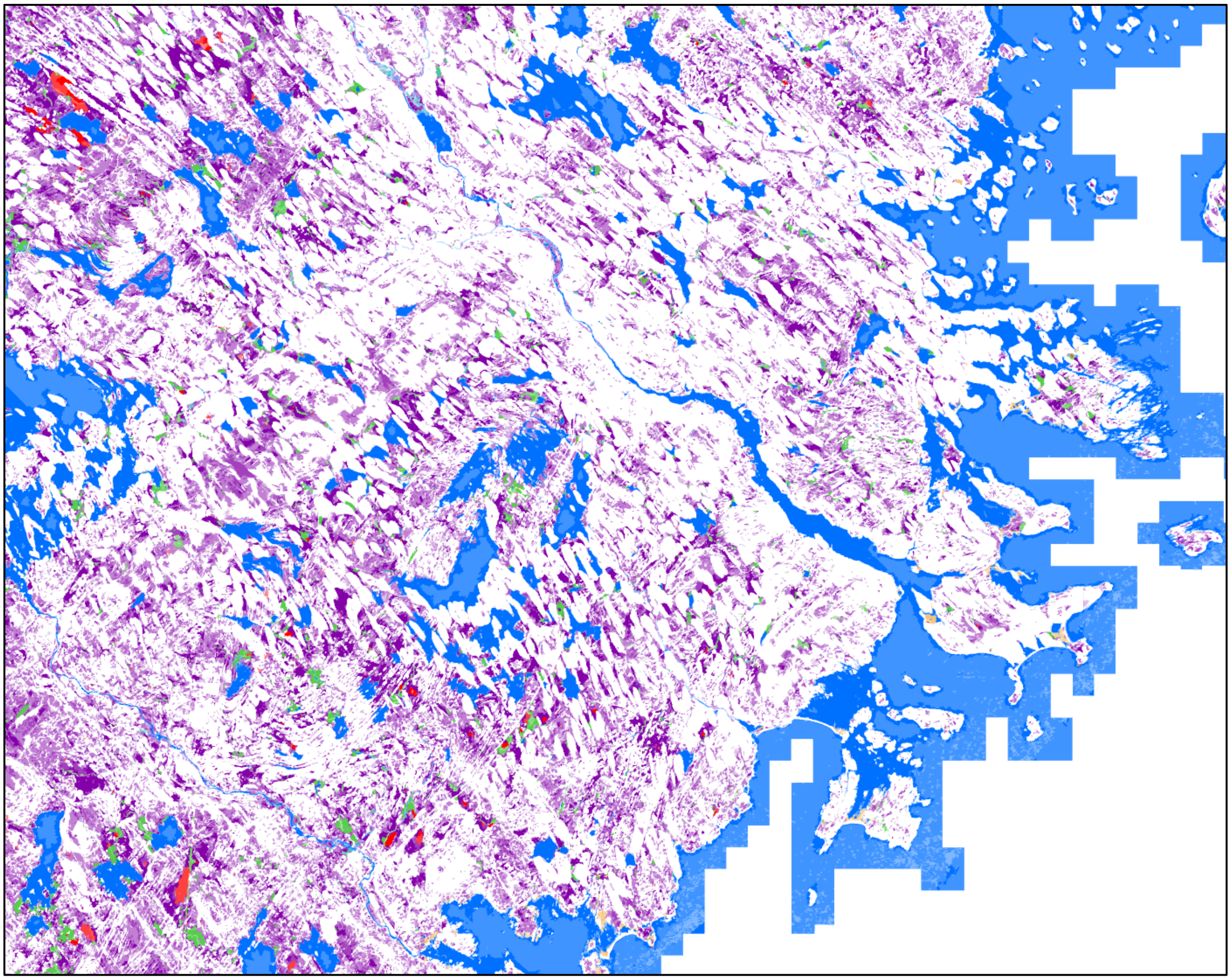
# Improved Methods for Wetland Identification



Nova Scotia Community College
Applied Geomatics Research Group
NSCC, Middleton, NS
Tel. 902 825 5475
Email: kevin.mcguigan@nscc.ca

Colin MacDonald
Director, Geographic Information Services
ICT Services, Internal Services Department
Government of Nova Scotia
5161 George St.
Halifax, NS B3J 2Y1
March 31st, 2024

How to cite this work and report:

NSCC Applied Geomatics Research Group. 2024. Improved Methods for Wetland Identification. Technical Report delivered to Internal Services Department, Government of Nova Scotia, Applied Geomatics Research Group, NSCC Middleton, NS.

# CONTENTS

Table of Figures......v
Executive Summary......ix
Key elements of the proposed approach include:......ix
1 Introduction......1
1.1 Background......1
1.2 Definitions......2
1.3 Techniques for Wetland Mapping......3
1.3.1 Detect Objects Using Deep Learning......4
1.3.2 Forest-Based Classification and Regression (FBCR)......4
1.3.3 Presence-only Prediction (MaxEnt)......4
1.3.4 AutoML in ArcGIS......5
1.4 Features for Wetland Mapping......5
1.4.1 Lidar Derived Metrics......5
1.4.2 Satellite Derived Metrics......12
1.4.3 Features Not Implemented......18
1.4.4 Existing Wetland Map Products of Nova Scotia......21
2 Methodology......22
2.1 Data Sources and Inputs......22
2.1.1 Provincial Lidar Data......22
2.1.2 Satellite Data......23
2.1.3 Training/Validation Data......24
2.1.4 Ancillary Data......25
2.2 Overview of the Proposed Methodology......26
2.3 Data Collection Methods......27
2.3.1 Lidar Data......27
2.3.2 Satellite Data......28
2.4 Data Analysis Procedures......28
2.4.1 Compilation of desired Feature Layers......28
2.4.2 Training AutoML ensemble Model......31
2.4.3 Raster Segmentation......32
2.4.4 Predict Model to Segments......33

2.4.5 Methods Overview ....................................................................................34

3 Results ..................................................................................................34

3.1 Detailed Analysis ......................................................................................36

4 Discussion ..............................................................................................46

4.1 Interpretation of Results ............................................................................47

4.1.1 Implications of Forest Canopy ..................................................................51

4.2 Comparison of ArcGIS tools ........................................................................51

4.3 Feature Layers ..........................................................................................53

4.3.1 Relative variable importance ...................................................................53

4.4 Implications for Flood Mitigation and Ecosystem Services ...............................54

5 Recommendations ...................................................................................55

5.1 Strategic Recommendations for Wetland Management .....................................55

5.2 Suggestions for Methodology Refinement .....................................................55

6 Conclusions ...........................................................................................57

6.1 Summary of Key Findings ..........................................................................57

6.2 Final Thoughts ..........................................................................................57

7 Acknowledgments ...................................................................................59

7.1 Contributions ...........................................................................................59

7.2 Support ...................................................................................................59

7.3 References ...............................................................................................59

8 Appendices ............................................................................................62

8.1 Processing Tools .......................................................................................62

8.2 Detailed Methods ......................................................................................64

8.2.1 Verbose Methods ...................................................................................64

# TABLE OF FIGURES

Figure 1 shows the basic workflow of a maxEnt type presence only model. Top right, it accepts point location presence inputs. Bottom right and middle - the model can be provided a selection of prepared rasters such as satellite images, lidar elevations, or other special indexes (middle). The resulting model (leftmost) can produce a 0-100% probability estimation of occurrence when provided with additional data similar in nature to the inputs. ....................................................4
Figure 2 shows a lidar derived DEM of central south-western Nova Scotia (Springfield). Drumlin glacial till mounds are seen as topographic highs, whereas flat, possible wetland and water areas are visible as well. Lidar DEMs are typically shaded to accentuate surface features as is presented here.............................................................................................................................6
Figure 3 shows a direct computation of the previous lidar DEM extent showing slope. Slope is displayed in, for example here, 0-90 degree (Springfield Area, NS). ..............................................7
Figure 4 shows (top) the category output for the geomoprhon tool in ArcGIS. Geomprhon categories include flat (1), Peak (2), Ridge (3), Shoulder (4), Spur (5), Slope (6), Hollow (7), Foot slope (8), Valley (9), Pit (10). Below shows the scalar output. Note that flat and low features become highly apparent (Springfield Area, NS). ...................................................................................................8
Figure 5 shows DEV calculations of various kernel sizes. Shows here are a more high-resolution DEM of 100m kernel (left) and DEV processed with 1000m kernel (right). Altering the DEV kernel highlights differing landform features. (Tusket Lake Area, NS) .....................................................9
Figure 6 shows an example Topographic Wetness Index (TWI). Note that water bodies and low laying (poor draining areas) are preferentially highlighted. This analysis is done by computing the drainage across lidar data. (Springfield Area, NS) ......................................................................10
Figure 7 shows how the fill difference (FILLD) analysis can help indicate specific discrete water and low depressions. This technique is computed as a requirement for the TWI data and is included due to its convenience. (Springfield Area, NS)........................................................................11
Figure 8 show the results of the CHROME analysis, which is computed from lidar slope. Note that low lying flat features are highly visible and differentiated generally from higher lying flat areas. (Springfield Area, NS) .....................................................................................................12
Figure 9 shows an example a location for a cloud-free median RGB composite of all available Sentienl-2 fall time images available on GEE collection 'Harmonized Sentinel-2 MSI Level-2A'. Clouds are further filtered using the cloud score plus deep learning feature. Near Infrared is available but not shown. (Springfield Area, NS) ..................................................................14
Figure 10 shows the same as previous but filtered to summertime images. This is done by filtering scenes to an appropriate range of Julian days before computing the median observed pixel values. (Springfield Area, NS) .....................................................................................................14
Figure 11 again shows an example Sentinel-2 median mosaic of the spring-time. This is computed in a similar way to both fall and summer by adjusting Julian day ranges. (Springfield Area, NS)......15
Figure 12 shows an example of a standard NDVI calculation derived from Near Infrared and Red satellite bands. Note that vegetation becomes highly differentiable from water and roads. Subtle variations are also important across vegetated areas – generally indicating variations in health, growth state, or species. (Springfield Area, NS) ..................................................................16

Figure 13 shows that the NARI index is capable of further differentiating vegetated environments. Note for example the prevalence of bight areas bounding some low laying lake perimeter. (Tusket Lake Area, NS) ....................17
Figure 14 shows the overview of both the lidar derived elevation model (left) and the canopy height model (right). These data are the culmination of many individual aerial lidar scans across the province conducted between 2016 and 2020. ....................22
Figure 15 shows province wide satellite mosaics built using GEE. Each data are compiled in using seasonal median across all scenes as described, and advanced cloud filtering is applied where appropriate. First are sentinel-2 RGB NIR products at a native 10 m including (S2P, top left) springtime, (S2S, top right) summer, (S2F mid-left) fall [NIR bands not shown]. Mid-right shows 20m native fall season Sentinel-MSI composite (S2F20, bands 5,6,7 shown). Bottom shows Sentinel-1 radar spring medians for Vertical-Vertical polarization (S1PVV, left) and Vertical-Horizontal (S1PVH, right)....................23
Figure 16 show the Nova Scotia Wetland Inventory consisting of 101,579 total wetland polygons which closely adheres to the Canadian wetland classification schema (NSE Wetlands). Wetland extents found in the finest ecological units of the Nova Scotia Ecological Classification of 2017, provide additional context and information which proved useful for a wider consideration of our results. ....................26
Figure 17 shows the overall proposed wetland mapping procedure. Each of the input datasets were resampled to a consistent 10-meter resolution and formed a total of 38 generated feature layers. These layers then form the basis for training an ensemble classification model. The model is predicted on polygon features formed from segmenting a composite of layers highly sensitive to wetland delineation (TWI, CHROME, NDVI). ....................27
Figure 18 shows a graphical summary of all included explanatory rasters used for the wetland classification model. All rasters are derived lidar or satellite seasonal mosaics generated using google earth engine (GEE). Depth to water (DTW) was provided by NSECC at a native 5m resolution and resampled 10 m to be consistent with all other raster layers. ....................31
Figure 19 shows the calculated log loss (accuracy and confidence) improve for the ensemble model as the classifier trains iteratively. ....................32
Figure 20 shows the Receiver Operating Characteristic (ROC) Curve which indicates the model's ability to differentiate between classes vs random, while the Precision-Recall Curve indicates % correct of the class (precision) and % of the class correct (recall) ....................32
Figure 21 shows the resulting segmentation as performed on the 10 m chrome, NDVI, and topographic wetness index feature rasters....................33
Figure 22 our method poduces classified segments, where each segment has been assigned the average raster value for each of the 38 input layers. ....................33
Figure 23 shows the general graphical overview of the process. The main inputs are used to derive a total of 38 explanatory variables. 3 of these (CHROME, TWI, NDVI) are used to perform a province wide segmentation, producing 10 million + segments. These segments are assigned the average of the 38 layer values which are then used to predict the wetland classification type based on an appropriately trained model. The model presented was trained using 2000+ wetland type identifications points provided by NSECC, which were sampled to the same 38 input layers and trained using AutoML ADVANCED mode in ArcGIS. The prediction applied to the segments in performed in batches. ....................34

Figure 24 shows the overall results of the AutoML prediction applied to each of the ~10 million segments across the province. ....................................................................................... 35
Figure 25 shows one of various views by which the results have been inspected. Each segment, having been assigned a wetland classification through the prediction of our model – is overlain with the best available known wetland locations for Nova Scotia, namely, the wetland inventory (NSE Wetlands) and the ecological land classification wetlands (NS ELC Wetlands). Further we inspect the in context of the overlay with various predictor values including the high prediction importance CHROME value, here shows as the average CHROME value per segment, sampled from the 10m raster source. ........................................................................................................... 36
Figure 26 shows the prediction results around the Kentville/Minas Basin area. Locally we see a greater abundance of salt marshes predicted with is in line with the attributes of the Nova Scotia wetland inventory (NSE). We see some agricultural land being depicted as slat marsh as well which may have implications for agricultural monitoring though further investigation as such is outside of the scope of this report. Note that while the delineation of fens (green) and bogs (red) do conform well to the NSE wetlands, swamps present a systematic increase. This increase is generally tied to tree covered areas – though the accuracy of this in real world terms generally unknown. .............. 38
Figure 27 shows the south of Nova Scotia, south lake Rossignol. This area of our results presents largest concentration of predicted wetlands (generally swamps) which deviate from the coverage presented in the existing Nova Scotia wetlands Inventory (NSE wetlands). This discrepancy is notably less substantial when considering the wetland coverage presented by the ecological land classification wetland ecosites (NS ELC wetlands). Further inspection details that this discrepancy maybe accounted for in part by previous works inability to detect features below the forest canopy. This overall result is yet not scrutinized by any field validation. .................................................. 39
Figure 28 shows a portion of southwestern Nova Scotia in the Weymouth area. Here we see a vast distribution of predicted swamps (purple) which are not included in the existing Nova Scotia wetland inventory or ecological land classification wetland polygons (hatched polygons). These swamps to conform to low laying features are glacial drumlins are clearly visible across the area showing notably less wetlands present. ......................................................................... 40
Figure 29 shows the northern shore and isthmus area of Nova Scotia. We see a vast increase in the overall predicted presence of swamp type wetlands (purple) vs existing wetland maps (hatched lines). Note that our model predicts salt marshes in agricultural areas – though generally seems to perform across varied tide ranges. The prevalence of predicted swamp wetlands does decrease with geological and elevation change visible to the south of the map where the Cobequid mountains are visibly less dense with swamps. ................................................................. 41
Figure 30 shows the Halifax regional municipality and area. Here we can see some success in detecting key wetland areas bounding various rivers and lakes, whereas urbanized locations such as the Halifax peninsula shows little existing wetland areas. Our prediction suggests much of the region includes small low laying swamp type wetlands which may not presently be present in the Nova Scotia wetlands inventory. These results have not been independently verified. .................. 42
Figure 31 shows the eastern shore area of Nova Scotia. As with other area, our model agrees strongly with bog type wetlands when considering the Nova Scotia wetland inventory of the ecological land classification wetland polygons. We systematically do predict a wider distribution of swamp type wetlands (purple) in low laying areas. ................................................................. 43

Figure 32 shows the southern portion of Cape Breton. Comparing our results to the existing wetland inventory and ecological land classification wetland polygons, we see generally a strong agreement with bog type wetlands (red). As with much of our results, a generally greater coverage of swamp type wetlands has been predicted (purple). ....44
Figure 33 shows the Cape Breton highlands and region. Here we see that our model does predict significant areas of new swamps (purple) generally bound to river flood plane areas. In the highlands themselves however, we do see systematic underprediction of bogs. This discrepancy appears unique to this region based on inspecting our results across the province. ....45
Figure 34 shows the normalized confusion matrix of the wetland prediction model trained with ArcGIS AutoML on 2,041 training points with 10% withheld for validation. ....46
Figure 35 shows the numerical Ecodistricts as depicted in the Nova Scotia ecological land classification (ELC). Also indicated are the wetland areas designated in that analysis (Neily et al. 2017). ....49
Figure 36 highlights the deviation between, **A1**; the coarse wetland mapping product presents in the NS ELC (Neily and 2017), **A2**; all wetland types as depicted by the by existing wetland inventory, and **A3**; Our results. As shown in the above table, the cape Breton highland is the only Ecodistrict where we detect an overall lower abundance of wetlands. This is likely due to the unique characteristics of the highlands which are not adequately captured in the model. Perhaps additional training or values may help correct this perceived discrepancy. ....49
Figure 37 shows Ecodistrict 760 (Sable), directly to the south of Lake Rossignol, NS. This area presents the largest systematic discrepancy in terms of our greater abundance of swamp type wetlands. Considering all wetlands, we see that the wetland extents presented in the NS ELC (**B1**) are in support of our results (**B3**) in that there are similarly more overall wetlands depicted than currently in the Nova Scotia wetland Inventory (**B2**). ....50
Figure 38 shows the same area Ecodistrict 760 (Sable) south of Lake Rossignol. Here our results are filtered to segments where the mean canopy height value (CHM) is less than 1.5 meters. This indicates wetlands which do not have significant vegetation. Note that with this restriction, our then results (black) are more in line with what is presented by the current Nova Scotia wetland Inventory in terms of all wetland types (red). ....50
Figure 39 Python 3.10.12 xgBoost version 2.0.3 feature importance. ....53
Figure 40 shows the SHAP feature importance score for the AutoML BASIC model, which was run with the same inputs as the ADVANCED model used to generate out results. ....54

# EXECUTIVE SUMMARY

This report presents a comprehensive machine learning based approach to wetland identification within Nova Scotia. This work is in support of the province's commitment under the Climate Change Plan for Clean Growth to leverage nature-based solutions for mitigating coastal and inland flood risks (Action 12). Recognizing the pivotal role of wetlands in flood regulation alongside their myriads of ecosystem services, this initiative seeks to enhance provincial resilience through enabling improved environmental stewardship.

The inadequacy of the current Nova Scotia Wetland Inventory in accurately representing forested wetland ecosystems underscores the necessity for a refined inventory to better gauge climate change impacts and opportunities, notably with respect to flooding. Consequently, this report outlines the development of a desktop methodology employing lidar and advanced remote sensing/classification techniques at the provincial scale.

Although this report is focused on methods and presents a preliminary wetland assessment, our classification results do suggest a significantly higher overall proportion of wetland areas (27.5% not including water), specifically swamp type wetlands, when compared to existing provincial wetland inventories (6.8%). When restricting our results to areas to low vegetation height we get a more similar 7.2%. Though these results are not intended as an update without further refinement and validation, this increase is in line with the perceived bias in the current inventory against tree covered swamps due to limitations in the collection methods.

This report details the effectiveness of our approach and suggests a potential for future expansion to include climate change projections. Overall, it presents an adaptable classification method at the provincial scale which can be applied more generally. This proposed method is intended for the Nova Scotia Department of Environment and Climate Change's consideration amongst others, to assist in various efforts to facilitate a more automated classification approach to ecosystem monitoring.

## KEY ELEMENTS OF THE PROPOSED APPROACH INCLUDE:

- An analysis of available, provincial scale, remotely sensed datasets and various derived indexes in a context of wetland mapping including lidar and Sentinel-1/2 satellite data.
- A brief review of various machine learning tools available in ESRI ArcGIS software.
- Development of 38 selected 10m resolution province-wide explanatory rasters
- Supervised wetland AutoML classification model trained on 2,041 sample locations
- Specialized approach producing 10+ million polygon segments spanning the province
- Wetland type prediction calculated for each polygon segment
- A comparison against existing wetland inventories
- A suite of python-based tools encompassing the entire workflow

# 1 INTRODUCTION

## 1.1 BACKGROUND

Wetlands are a highly valuable ecosystem for several reasons, including water purification through runoff buffering, carbon sequestration, flood control, and pollution mitigation (Gallant et al., 2020; Reis et al, 2017; Tozer et al. 2018). They are typically highly productive and biodiverse, creating critical hotspots for biodiversity conservation.

It is estimated that over 3 million $km^2$, or over 20%, of global wetlands have been lost since 1700, primarily due to human land usage (Fluet-Chouinard et al. 2023). Negative effects of wetland loss are numerous, including habitat loss, flooding, and increased greenhouse gas emissions from land reclamation (Li et al. 2018).

Canada contains over 1 million $km^2$ of wetlands, which cover approximately 13% of the country's terrestrial area. Canadian wetlands account for almost one quarter of the remaining global wetlands (ECCC 2016). Almost 4,000 $km^2$ of these wetlands are found in the province of Nova Scotia (NSECC 2024). Chignecto, Musquodoboit Harbour, and Southern Bight-Minas Basin are three wetlands in Nova Scotia which are considered of International Importance by the Ramsar Convention on Wetlands and are considered "significant for humanity as a whole" (Convention on Wetlands Secretariat, 2024). As wetland loss continues globally, it will be increasingly important to monitor changes to wetlands in Canada at the federal and provincial level.

Wetland delineation and classification is traditionally done in-situ, which can be labour intensive, time-consuming, and inaccessible (Amani et al. 2017). For large scale classification, remote sensing offers the ability to collect high resolution datasets quickly and cost effectively (Mahdavi et al. 2017; Amani et al. 2019). Remotely sensed wetland data can be used for baseline establishment, short and long-term change detection, and mapping at high resolutions (Ballanti et al. 2017).

The Nova Scotia Department of Natural Resources and Renewables (NRR) conducted a comprehensive provincial wetland inventory in 2004, utilizing aerial photographs captured between 1985 and 1997. This inventory was later updated by the reclassification of LANDSAT imagery from 2000 to 2002 (Government of Nova Scotia, 2019). Wetlands smaller than 0.5 hectares were excluded from this inventory, which poses a limitation on our ability to fully understand the abundance of smaller isolated or headwater wetlands, potentially crucial for hydrological and ecological functions.

As per the current inventory, freshwater wetlands cover approximately 360,462 hectares, constituting roughly 6.8% of the total land area, while salt marshes span 17,060 hectares, representing 0.3% of the total land area in Nova Scotia (Government of Nova Scotia, 2019). Despite these findings, there remains a significant challenge in accurately identifying and quantifying forested wetlands, particularly swamps. The existing dataset falls short in adequately capturing the extent of forested wetlands due to limitations in aerial interpretation techniques, compounded by similar spectral signatures exhibited by forested uplands and wetlands (Jahncke et al., 2018). The underrepresentation of wetlands, notably swamps, highlights a well-documented challenge inherent in the current inventory and photo-interpretation methodologies at large (Jahncke et al., 2018). These limitations hinder our ability to effectively manage and conserve wetland resources across the province, highlighting the need for more advanced and precise mapping techniques to address these shortcomings.

## 1.2 DEFINITIONS

Wetlands are defined federally in Canada as "areas of marsh, fen, peatland or water, whether natural or artificial, permanent or temporary, with water that is static or flowing, fresh, brackish or salt" (ECCC 2016). The Canadian Wetland Classification System also provides the definition of "terrain affected by water table at, near or above the land surface and which is saturated for sufficient time to promote wetland or aquatic processes" (National Wetlands Working Group 1997).

The following table is adapted from National Wetlands Working Group (1997) and contains information on wetland classification:

*Table 1. Wetland classification.*

| Wetlands | | | |
|---|---|---|---|
| Peatland | | Mineral Wetland | |
| Bog | Peatland receiving water exclusively from precipitation and not influenced by groundwater; *Sphagnum*-dominated vegetation. | Marsh | Periodic or persistent standing water or slow-moving surface water which is circumneutral to alkaline and generally nutrient-rich. Vegetation is dominated by graminoids, shrubs, forbs, or emergent plants. |
| Fen | Peatland receiving water rich in dissolved minerals; vegetation cover composed dominantly of graminoid species and brown mosses. | Shallow Water | Wetlands with free surface water up to 2 m deep, present for all or most of the year, with less than 25% of the surface water area occluded by standing emergent or woody plants. Submerged or floating aquatic plants usually dominate the vegetation. |

| | | | |
|---|---|---|---|
| Swamp | Peatland dominated by trees, shrubs, and forbs; waters are rich in dissolved minerals. | Swamp | Periodically standing surface water and gently moving, nutrient-rich groundwater, with vegetation dominated by woody plants often more than 1 m high. |

- "**Bogs** are peat-covered wetlands (peatlands), in which the vegetation shows the effects of a high water table and a general lack of nutrients."
- "**Fens** are peatlands characterized by a high water table, but with very slow internal drainage by seepage. "
- "**Marshes** are wetlands that are periodically inundated by standing or slowly moving water and hence are rich in nutrients."
- "**Swamps** are wetlands where standing or gently moving waters occur seasonally or persist for long periods, leaving the subsurface continuously waterlogged."
- "**Shallow water** wetlands, [...] a transition stage between lakes and marshes"

Note that this method and report does not include mapping of the wetland class 'shallow water' which are mixed with marshes in the provided training data. These features are typically more complex to map adequately due to their inherent association with neighboring wetlands.

## 1.3 Techniques for Wetland Mapping

Traditional techniques for wetland mapping have primarily relied on manual expert delineation and Random Forest algorithms for classification and analysis. Random Forest is a versatile and robust machine learning algorithm that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification), offering high accuracy without overfitting the data (Breiman 2001). Recently, there has been a growing interest in exploring the potential of fully convolutional methods for environmental mapping, indicating a significant shift towards more automated and sophisticated approaches (Long 2015, Mainali 2023). Even still, many recent successful wetland studies rely on random forest-based classification (LaRocque 2020, Amani 2019, DeLancey 2021).

Previous similar work to this study has been undertaken in the region, specifically surrounding HRM which employed a random forest classifier of wetlands using satellite observations including radar, high resolution lidar, and several lidar derived metrics such as local texture and roughness (Jahncke, 2018). Federally, the Canadian government has an ongoing program to amalgamate such wetland mapping data into the Canadian National Wetlands Inventory (CNWI). This dataset follows a standard classification schema and is sourced from various levels of government though no standard methodology for deriving these data is known to the authors of this report.

In this study, we focus first on experimenting with several methods available in ArcGIS Pro 3.1.1, leveraging its comprehensive suite of tools for spatial analysis and developing a final repeatable workflow. Concurrently, a parallel project is investigating the capabilities of tools available directly in Google Earth Engine, aiming to compare the effectiveness and applicability of different platforms in enhancing wetland inventory accuracy and efficiency. This dual approach allows us to assess a broad spectrum of methodologies and identify the most promising strategies for modernizing wetland mapping in Nova Scotia.

In the context of mapping spatial extents of features such as wetlands, there exists a suite of tools and functions available in ArcGIS as described below.

### 1.3.1 Detect Objects Using Deep Learning

This tool leverages deep learning algorithms within ArcGIS to automatically identify and delineate wetland features from high-resolution imagery, facilitating efficient and scalable mapping efforts by learning from labeled examples to improve accuracy over traditional manual methods. Though deep learning approaches have been showing increasing utility, their added complexity proved an impediment to pursuing these tools for province wide results at this time.

### 1.3.2 Forest-Based Classification and Regression (FBCR)

The Forest-Based Classification and Regression tool in ArcGIS employs an ensemble of decision trees (Random Forests) to analyze spatial data, providing a robust means for classifying wetland areas by incorporating a variety of input features, including lidar and satellite imagery, to predict wetland presence with high accuracy. Extensive testing indicated very poor and erroneous results from this tool specifically and ultimately it was abandoned in favor of other options.

### 1.3.3 Presence-only Prediction (MaxEnt)

Utilizing the MaxEnt (Maximum Entropy) model within ArcGIS, this approach focuses on presence-only location data, making it particularly useful for areas with limited information (Tesfamariam, 2022). It estimates the distribution of wetland types by maximizing entropy to predict the likelihood of wetland presence across the landscape, based on environmental and spatial variables. Initial test with this technique showed very promising results and should be considered further.

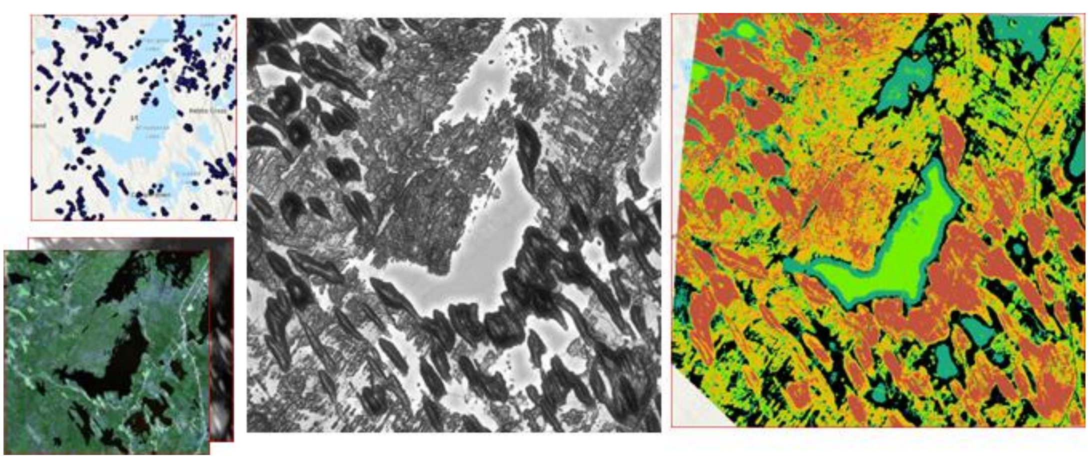


*Figure 1 shows the basic workflow of a maxEnt type presence only model. Top right, it accepts point location presence inputs. Bottom right and middle - the model can be provided a selection of prepared rasters such as satellite images, lidar* *elevations, or other special indexes (middle). The resulting model (leftmost) can produce a 0-100% probability estimation of occurrence when provided with additional data similar in nature to the inputs.*

### 1.3.4 AutoML in ArcGIS

AutoML in ArcGIS streamlines the application of machine learning algorithms to spatial data, automating the process of model selection, training, and optimization. It allows for the efficient identification of the most effective algorithm and parameter settings for specific wetland mapping tasks, significantly reducing the expertise and time required to deploy advanced analytics for environmental monitoring and assessment.

AutoML can perform in multiple modes, including BASIC, and ADVANCED, whereby the overall complexity and amount of automated parameter tuning may increase. This can provide a more robust cross-validated model where accuracy is most important.

The training process within the AutoML framework automatically tests, tunes, and incorporates various supervised machine learning algorithms to construct the best performing ensemble model. The selection includes **Linear regression**, primarily for smaller datasets, as well as a series of various implementations of decision tree-based classification models including **Random Trees**, **Decision Trees**, **XGBoost**, **Light GBM**, and **Extra Tree** algorithms, each offering distinct advantages in terms of speed, interpretability, and handling complex data structures. An additional tree-based model, **CatBoost** is particularly versatile, effectively processing mixed data types. AutoML seamlessly integrates these algorithms, evaluating and optimizing them to collectively form an ensemble model that delivers optimal performance for specific analytical tasks. This automated ensemble approach ensures that the final model leverages the combined strengths of each algorithm, maximizing accuracy and efficiency.

When tested, in general, against each of the previously indicated options available in ArcGIS, AutoML was determined to be the best available option both in terms of sophistication and reliability. The AutoML classification tool is therefore central to our proposed wetland mapping approach.

## 1.4 Features for Wetland Mapping

All machine learning type classifiers rely on suitable input datasets for optimal results. This is especially true with classifiers such as random forest, which perform no alterations to input data directly. As such, the success in rendering a high-quality province wide wetland inventory depends upon sourcing and preparing specialized map feature layers which specifically aid in differentiating both wetlands from non-wetlands (uplands), as well as individual wetland types. A selection out of the following metrics was compiled for the complete extent of Nova Scotia and included in our prototype wetland mapping process of the province. A complete list is included in the methods section.

### 1.4.1 Lidar Derived Metrics

The province wide lidar inventory provides a significant and unique advantage for upgrading the provincial wetland inventory. The occurrence of wetlands is naturally related to elevation and elevation derivatives such as local slope and metrics which can be derived from lidar elevation including various wet area indexes.

##### *1.4.1.1 Digital Elevation Model (DEM)*

The digital elevation model (DEM), also known as the digital terrain model (DTM) is a foundational form of lidar data in which the lidar point cloud has been classified as valid ground returns and rasterized (Aguilar et al. 2010). The DEM has been reported as the highest performing classification feature for detecting wetlands in some cases (LaRocque 2020). This dataset is the basis for many additional lidar derived metrics used in wetland mapping.

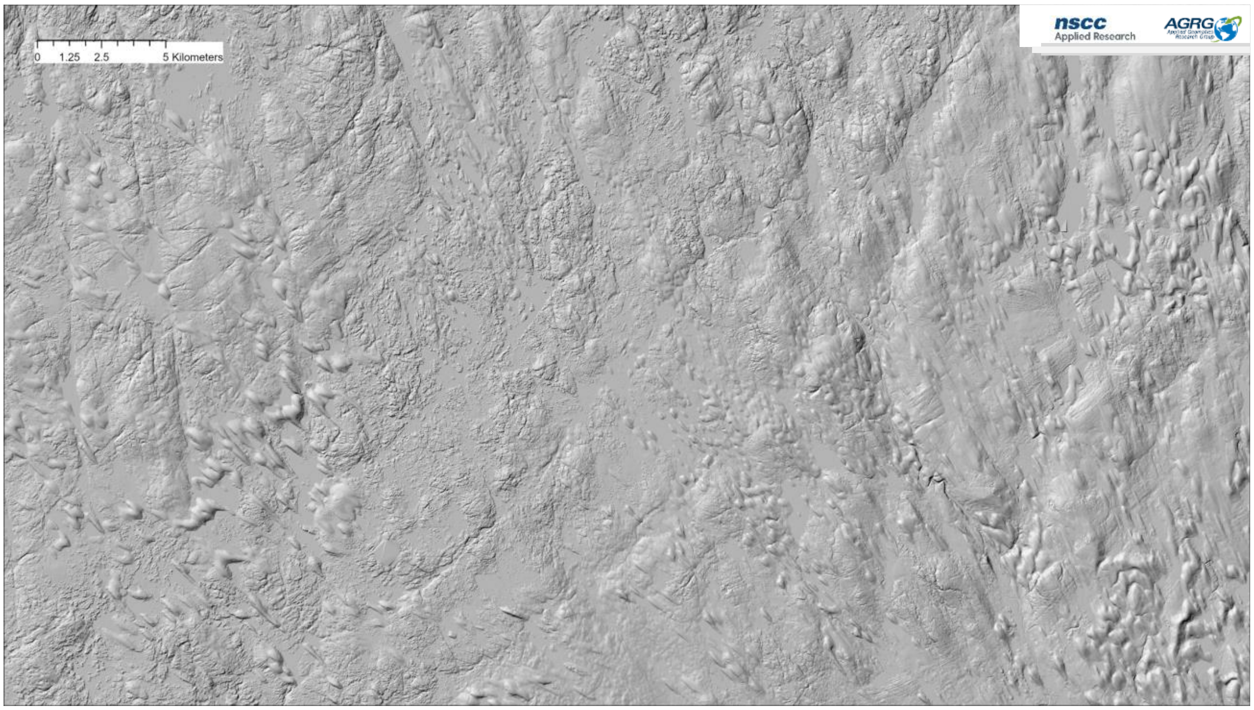



*Figure 2 shows a lidar derived DEM of central south-western Nova Scotia (Springfield). Drumlin glacial till mounds are seen as topographic highs, whereas flat, possible wetland and water areas are visible as well. Lidar DEMs are typically shaded to accentuate surface features as is presented here.*

##### *1.4.1.2 Canopy Height Model (CHM)*

Canopy height model (CHM) is critical for providing accurate tree canopy characteristics, such as growth and volume estimations (Khosravipour et al. 2015). These data are typically computed as the vertical difference between the DEM and the Digital Surface Model (DSM). The resulting value represents the normalized height from the ground to the, typically, highest valid lidar return. For a CHM to be considered a true canopy model, typically the analysis is limited to vegetation by masking or removing structures such as buildings. CHMs are valuable for further characterizing wetland structures such as treed swamp forest canopy and vegetation characteristics. Wang et al. (2017) have demonstrated, for example, that CHM values can have a high correlation with certain short standing wetland vegetation species.

#### *1.4.1.3 Slope*

Slope serves as a critical factor in wetland identification. Areas with lower slope angles are typically more conducive to wetland formation, as they facilitate water accumulation and reduced drainage, compared to regions with steeper slopes (Uuemaa et al., 2018).

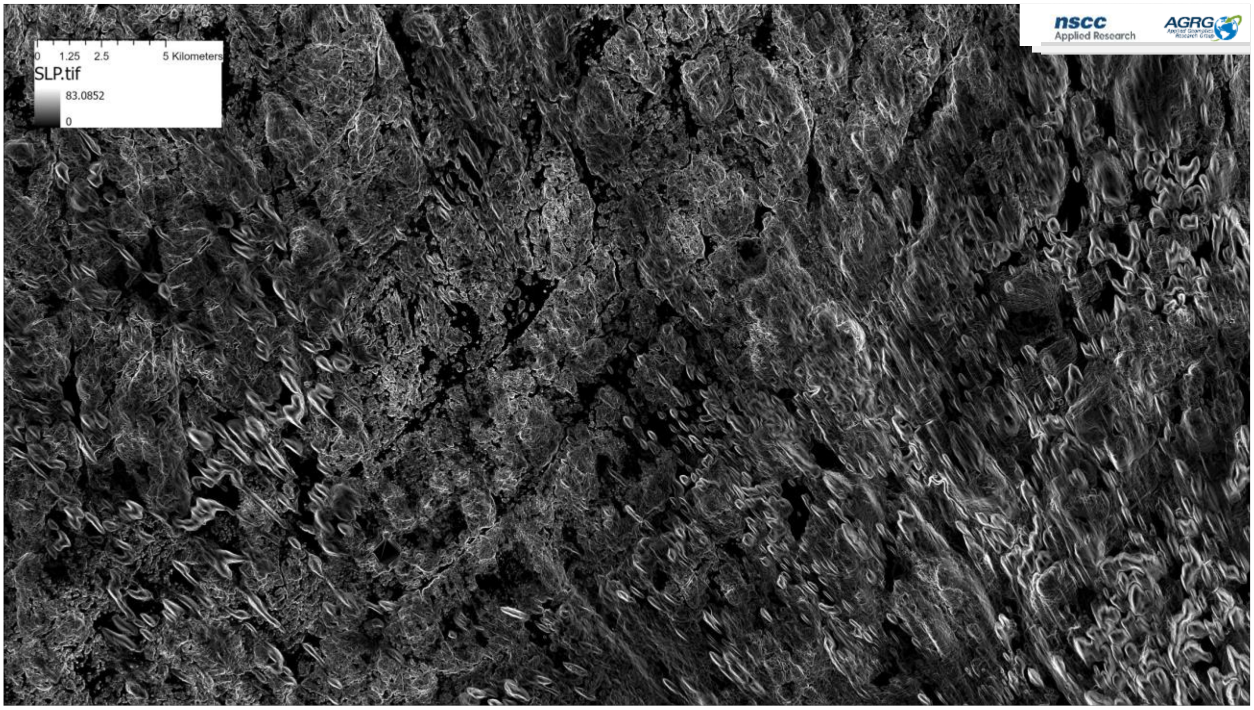



*Figure 3 shows a direct computation of the previous lidar DEM extent showing slope. Slope is displayed in, for example here, 0-90 degree (Springfield Area, NS).*

The slope is calculated as the rate of change in elevation (z) over a distance (x), which can be directly computed from the lidar DEM where:

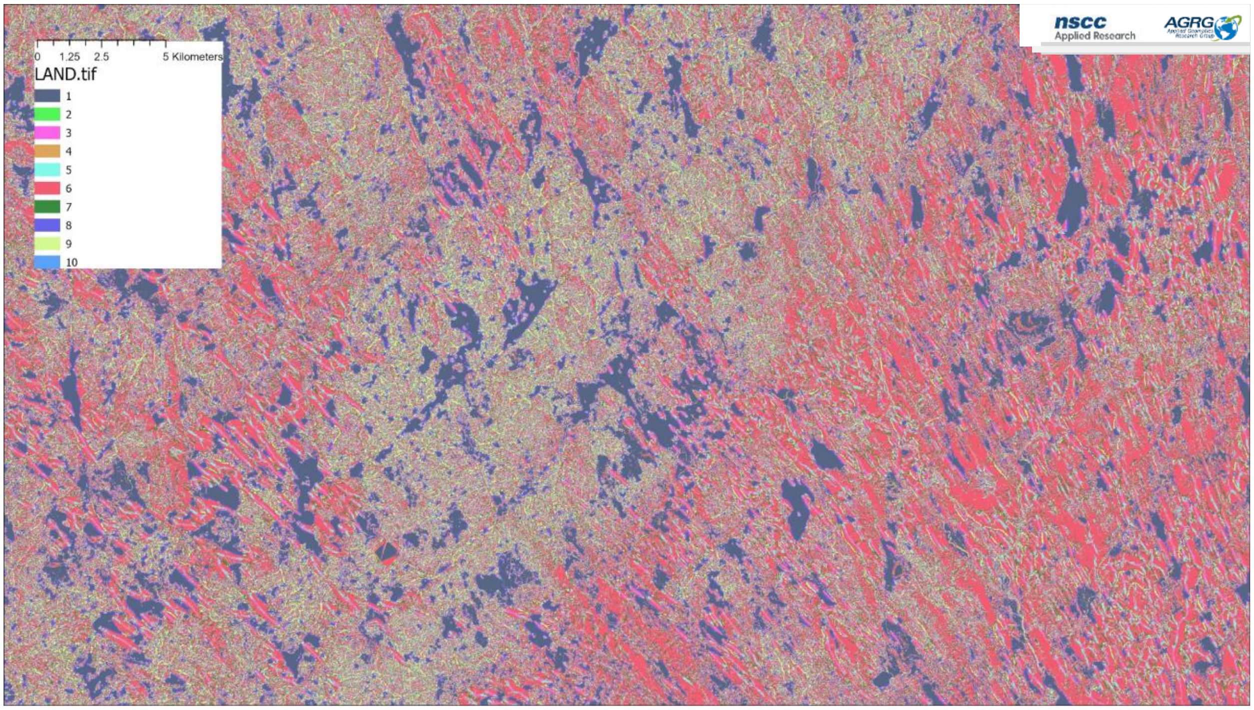
$$slope = \frac{dz}{dx}$$

#### *1.4.1.4 Depth To Water (DTW)*

Depth to Water (DTW) is an analytical derivative of the lidar elevation model which estimates an elevation difference between a given raster location and hydraulically least-cost connected water or drainage location elevation. Water locations can be defined as lakes, streams, or rivers (Murphy et al., 2007). Functionally this metric highlights areas which exhibit greater levels of water saturation and less drainage and are thus a very powerful metric for determining the presence of wetlands.

#### *1.4.1.5 Geomorphon*

Geomorphons are a form of pattern-based classification which describes each cell of a DEM according to one of the following: flat, peak, ridge, shoulder, convex slope, slope, concave slope, foot slope, valley, pit/depression (Stepinski, 2011; Gioia et al. 2021). Geomorphons have been shown to successfully detect wetlands when paired with deep-learning methods (Mainali 2023).



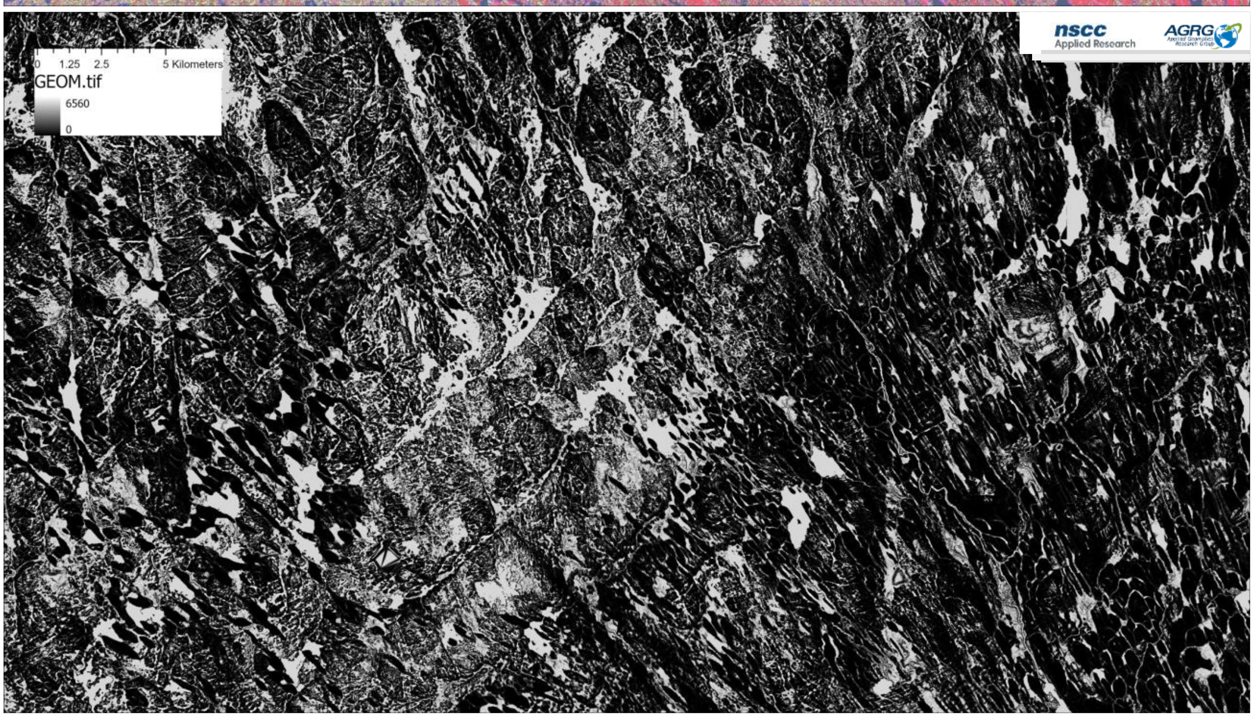




*Figure 4 shows (top) the category output for the geomoprhon tool in ArcGIS. Geomprhon categories include flat (1), Peak (2), Ridge (3), Shoulder (4), Spur (5), Slope (6), Hollow (7), Foot slope (8), Valley (9), Pit (10). Below shows the scalar output. Note that flat and low features become highly apparent (Springfield Area, NS).*

#### *1.4.1.6 Deviation of the Mean Elevation (DEV)*

Deviation of the Mean Elevation (DEV) is a variation of the Topographic Position Index (TPI), which refines the analysis by focusing on the elevation deviation of a specific pixel from the mean elevation within a specified window, normalized by the standard deviation of that window. This approach enhances the model's ability to highlight flat areas in regions of low topographical relief, making it particularly effective over the conventional TPI for delineating wetland areas. DEV is highly useful when analyzing heterogeneous landscapes, as it is scaled by local ruggedness (Lindsay et al. 2015).

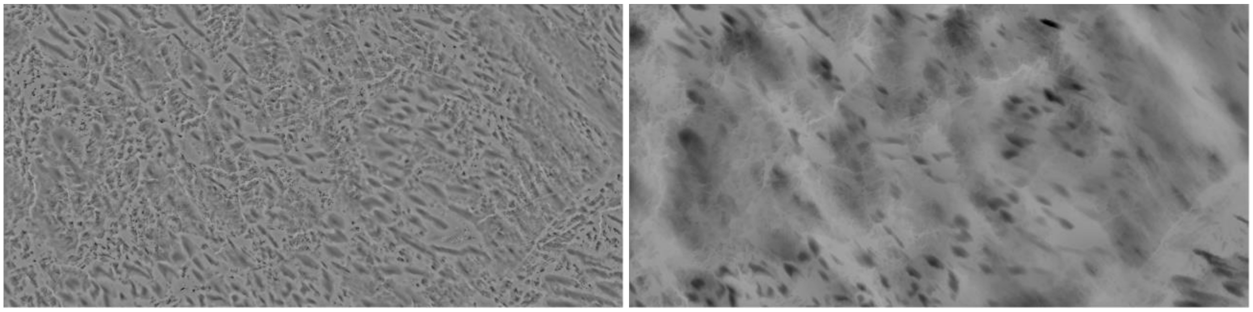


*Figure 5 shows DEV calculations of various kernel sizes. Shows here are a more high-resolution DEM of 100m kernel (left) and DEV processed with 1000m kernel (right). Altering the DEV kernel highlights differing landform features. (Tusket Lake Area, NS)*

$$DEV(D) = \frac{z_0 - \bar{z}_0}{S_D}$$

Where:

- D = window size
- $z_0$ = elevation of the window centre cell
- $\bar{z}_0$ = window mean elevation

#### *1.4.1.7 Topographic Wetness Index (TWI)*

The Topographic Wetness Index (TWI) is a quantitative measure that assesses the potential for soil saturation and water accumulation in each area, based on topography (Kopecký et al., 2021).

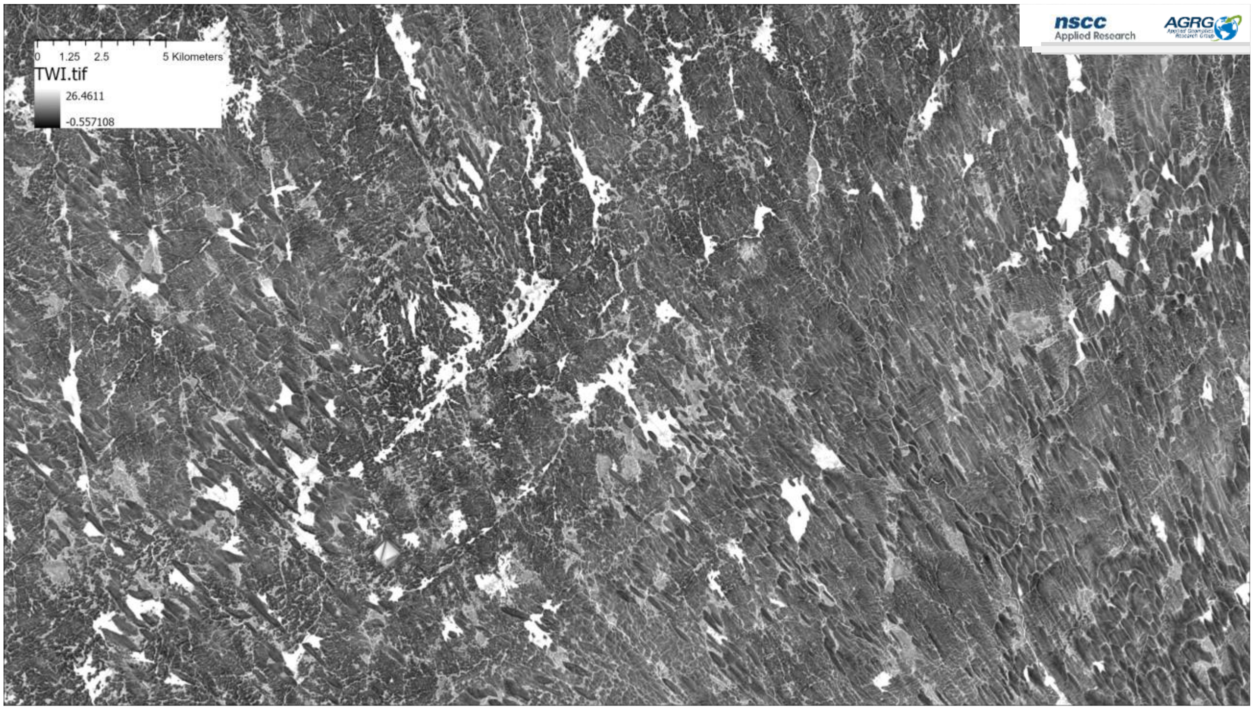



*Figure 6 shows an example Topographic Wetness Index (TWI). Note that water bodies and low laying (poor draining areas) are preferentially highlighted. This analysis is done by computing the drainage across lidar data. (Springfield Area, NS)*

TWI is calculated as the natural logarithm of the ratio of the total catchment area to the flow width, divided by the tangent of the slope. This index is pivotal for identifying areas with higher moisture levels, which are indicative of wetland presence.

$$TWI = ln \frac{\frac{total\ catchment\ area}{flow\ width}}{\tan\left(slope\right)}$$

#### 1.4.1.8 Hydrological Fill Difference

The Hydrological Fill process is essential in preparing elevation models for accurate computation of topographic wetness indices (TWI) and depth to water (DTW). It involves simulating the filling of depressions in the terrain to ensure hydraulic pathways are appropriately represented in the elevation model. The Hydrological Fill Difference metric specifically identifies areas that may not be adequately drained or do not drain well in the DEM. This identification is crucial for accurately analyzing areas prone to wetland conditions by highlighting potential discrepancies in hydraulic connectivity, thus ensuring more accurate wetland mapping.

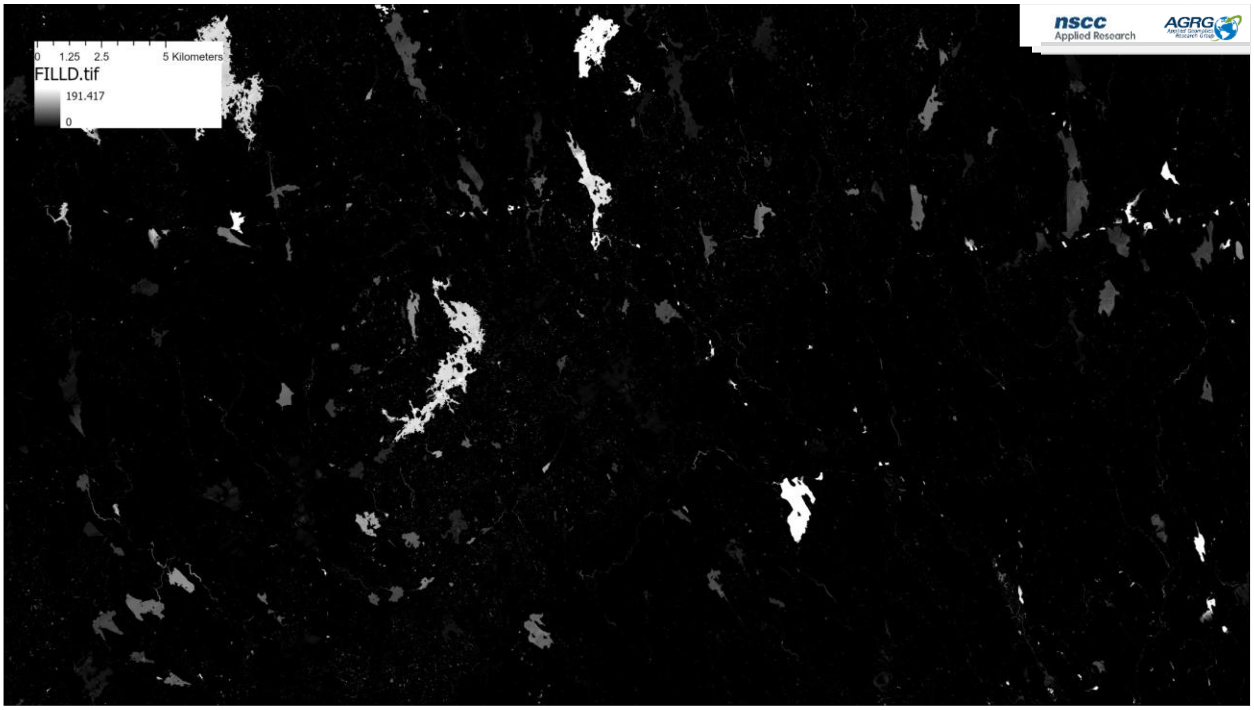



*Figure 7 shows how the fill difference (FILLD) analysis can help indicate specific discrete water and low depressions. This technique is computed as a requirement for the TWI data and is included due to its convenience. (Springfield Area, NS)*

#### 1.4.1.9 Chrome DEM

In this study, we introduce the novel Chrome DEM metric, which employs a distinctive approach by calculating the maximum of the mean slope across an increasing scale which effectively normalizes and scales the index appropriately. In effect, this balances the detection of both micro-habitat variations and broader landscape features. In practice this analytical layer performs only marginally better than slope directly in high slope areas, whereas it tends to differentiate flat 'higher' areas from flat 'lower' areas uniquely. This layer has also proven useful for visually inspecting low laying areas to highlight possible forested wetlands, for example, through inspection directly.

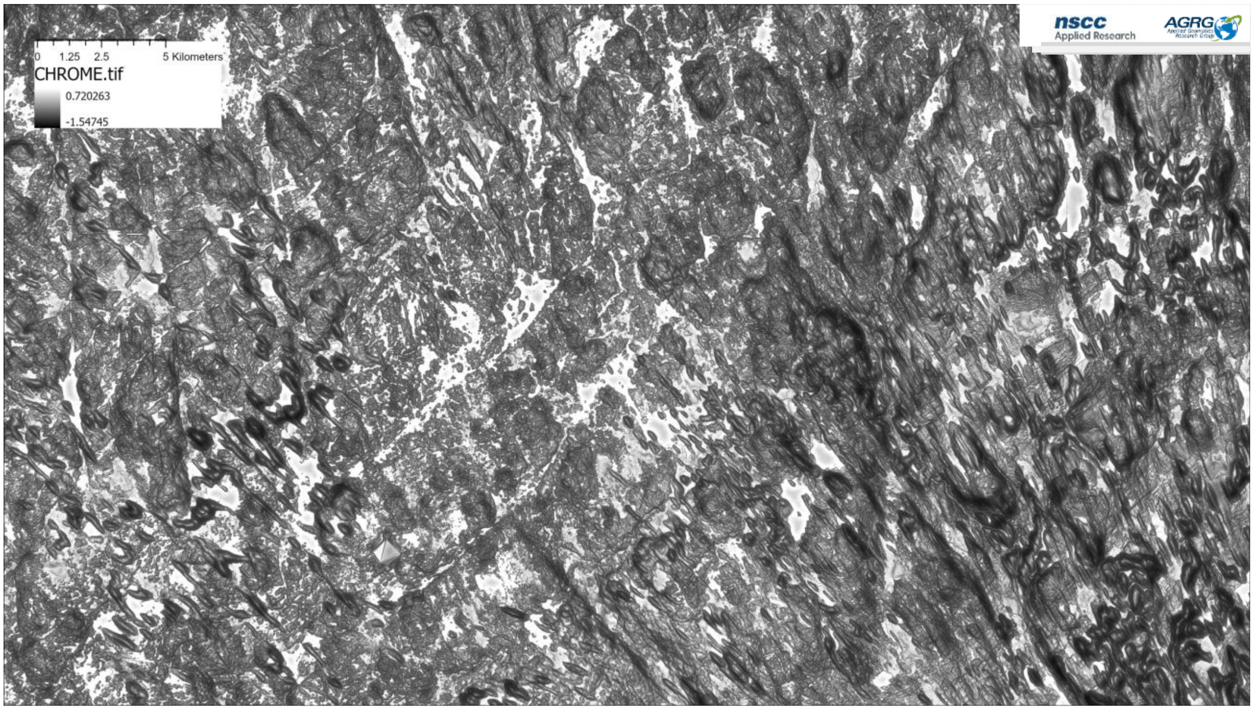



*Figure 8 show the results of the CHROME analysis, which is computed from lidar slope. Note that low lying flat features are highly visible and differentiated generally from higher lying flat areas. (Springfield Area, NS)*

This technique seems to ensure a comprehensive and comparably uniform metric that significantly enhances the accuracy of wetland mapping across diverse geographical landscapes.

$$ChromeDEM = \mu_{30m}\{max_{15m}[\sigma_{5m}(atan_S)]\} - atan_S$$

Where:

- μ = average
- σ = standard deviation
- S = slope

### 1.4.2 Satellite Derived Metrics

Satellite-derived metrics, particularly from Sentinel-2 bands, provide essential spectral information for wetland mapping. This data, capturing a range of wavelengths, enables the precise identification and classification of wetland areas through the analysis of vegetation health and moisture content.

Sentinel-2 bands and their associated wavelengths can be found in the following table from Kaplan & Avdan (2017):

*Table 2. Sentinel-2 bands (from Kaplan & Avdan 2017).*

| Sentinel-2 Bands | Central Wavelength (µm) | Resolution (m) |
|---|---|---|
| Band 1 - Coastal aerosol | 0.443 | 60 |
| Band 2 - Blue | 0.490 | 10 |

| | | |
|---|---|---|
| Band 3 - Green | 0.560 | 10 |
| Band 4 - Red | 0.665 | 10 |
| Band 5 - Vegetation Red Edge | 0.705 | 20 |
| Band 6 - Vegetation Red Edge | 0.740 | 20 |
| Band 7 - Vegetation Red Edge | 0.783 | 20 |
| Band 8 - NIR | 0.842 | 10 |
| Band 8A - Vegetation Red Edge | 0.865 | 20 |
| Band 9 - Water vapor | 0.945 | 60 |
| Band 10 - SWIR - Cirrus | 1.375 | 60 |
| Band 11 - SWIR | 1.610 | 20 |
| Band 12 – SWIR | 2.190 | 20 |

#### *1.4.2.1 Sentinel-2 Cloud Free Mosaics*

Satellite data can be, and typically is, incorporated into a classification model directly as bands (Jahncke 2018; Amani 2019; LaRocque 2020; DeLancey 2021). However, special care must be made to ensure that clouds are removed from such data. This can be done by carefully selecting and compositing cloud free scenes, though this proves to be a challenge over large geographies, while maintaining consistent conditions such as lighting and ground changes. A popular approach to facilitate this conveniently is to integrate a large number of satellite images over time such that the individual changes scene to scene can be averaged out in some way. Typically, this technique is employed on externally hosted satellite processing systems such as Google Earth Engine (GEE, Gorelick et al 2017). These scenes can be composited using the average (mean) or perhaps median observed value, and clouds can be further suppressed with special compensation or masking. A particularly robust method of cloud masking is specifically available on the GEE platform, namely, the Cloud Score Plus feature (Pasquarella 2023). This specialized masking value, derived from deep learning of cloud occurrences, can be specifically applied to the Harmonized Sentinel-2 MSI Level-2A catalog of satellite images to very competently suppress cloud occurrences to produce very high-quality cloud-free mosaics. For the methods of this paper, we produce, using GEE platform, 10m resolution Nova Scotia scale Sentinel-2 mosaics which contain the median cloud filtered observations of Red, Green, Blue, and NIR bands for each season (spring, summer, and fall) across the entire available Sentinel-2 catalog (approximately March 2017 to Dec. 2023). We further generate resampled 10m resolution for all additional bands (see above table) specifically for the fall season.

This data collectively provides a remarkably robust analytical tool for ecological mapping across Nova Scotia. Although compiling these types of data are made much more feasible and high quality in systems such as GEE, a similar procedure could be undertaken to compile the data from free sources. Whereas GEE is only available for free to academic and research purposes, it is worth considering other similar options which may be more readily available to the provincial government.

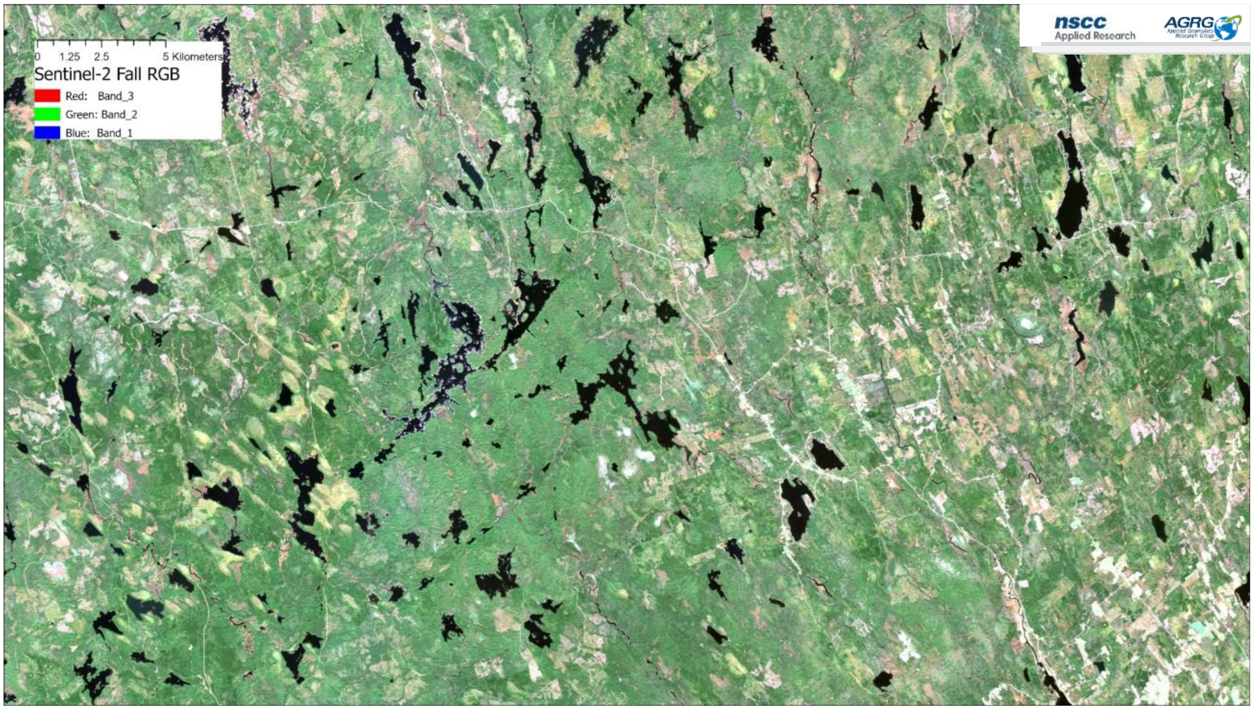



*Figure 9 shows an example a location for a cloud-free median RGB composite of all available Sentienl-2 fall time images available on GEE collection 'Harmonized Sentinel-2 MSI Level-2A'. Clouds are further filtered using the cloud score plus deep learning feature. Near Infrared is available but not shown. (Springfield Area, NS)*


*Figure 10 shows the same as previous but filtered to summertime images. This is done by filtering scenes to an appropriate range of Julian days before computing the median observed pixel values. (Springfield Area, NS)*


*Figure 11 again shows an example Sentinel-2 median mosaic of the spring-time. This is computed in a similar way to both fall and summer by adjusting Julian day ranges. (Springfield Area, NS)*

#### 1.4.2.2 Normalized Digital Vegetation Index (NDVI)

To further increase the signal of vegetation, specifically health and wetness, the Normalized Digital Vegetation Index (NDVI) is perhaps the most popular index to be computed directly from satellite bands. Today NDVI is a commonly used index, which was first described by Deering (1978).

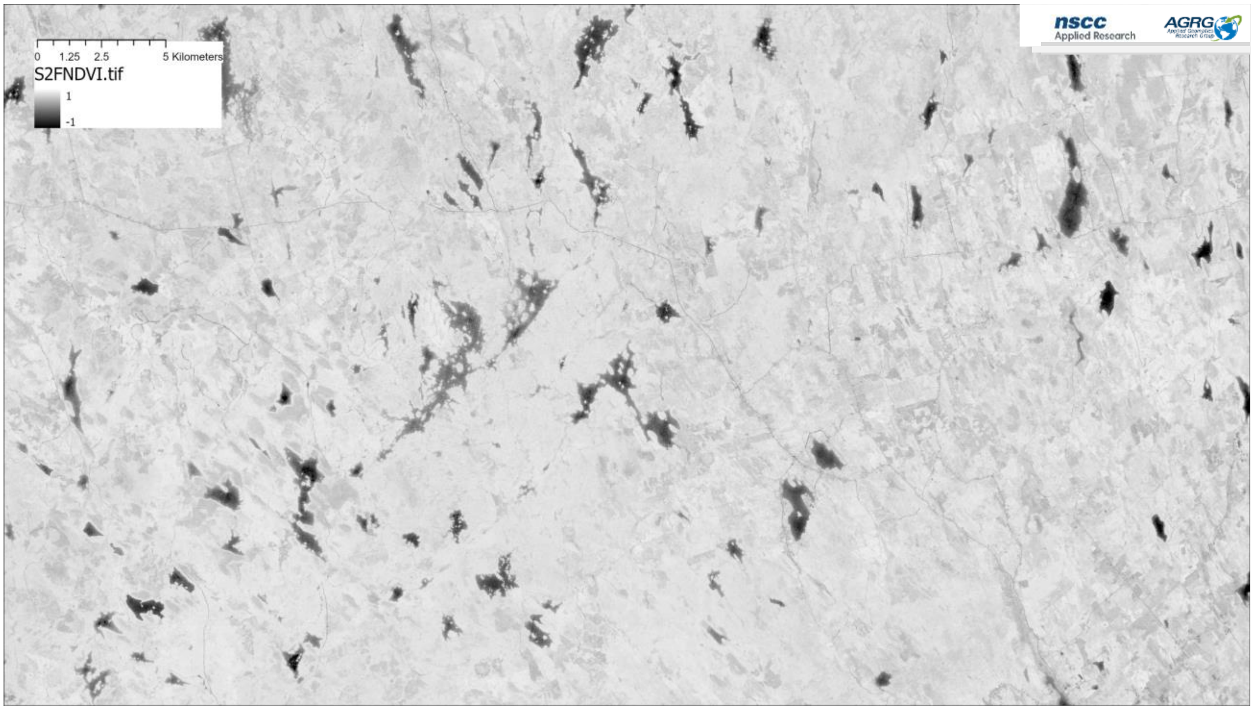



*Figure 12 shows an example of a standard NDVI calculation derived from Near Infrared and Red satellite bands. Note that vegetation becomes highly differentiable from water and roads. Subtle variations are also important across vegetated areas – generally indicating variations in health, growth state, or species. (Springfield Area, NS)*

The following equation is from Deering (1978):

$$NDVI = \frac{Rnir - Rred}{Rnir + Rred}$$

#### 1.4.2.3 Normalized Anthocyanin Reflectance Index (NARI)

The Normalized Anthocyanin Reflectance Index (NARI) effectively quantifies the autumnal anthocyanin accumulation in canopies of Ericaceae, a family of flowering plants in acidic and nutrient-poor environments. NARI outperforms the Normalized Difference Vegetation Index (NDVI) for mapping specific wetland species, significantly improving the estimated area of Ericaceae-dominated shrublands, for example. NARI's sensitivity to anthocyanin, a pigment that can be prevalent in Ericaceae species during the autumn, allows for the detection of these shrubs when they are most distinct from other vegetation. This is particularly useful in peatlands where these species might indicate more acidic conditions (bogs) or less acidic, nutrient-rich conditions (fens) (Bayle et al 2019). Note that NARI is dependent on coarser resolution Sentienl-2 bands, and this is not natively 10m in resolution as RGBNIR, or NDVI.

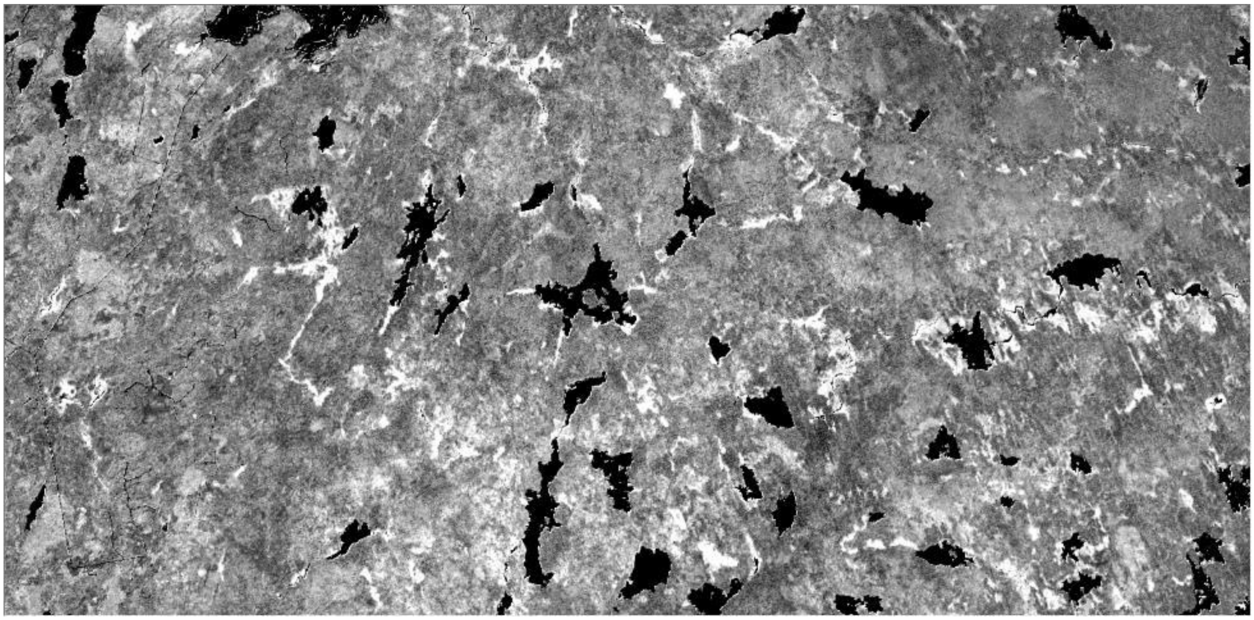


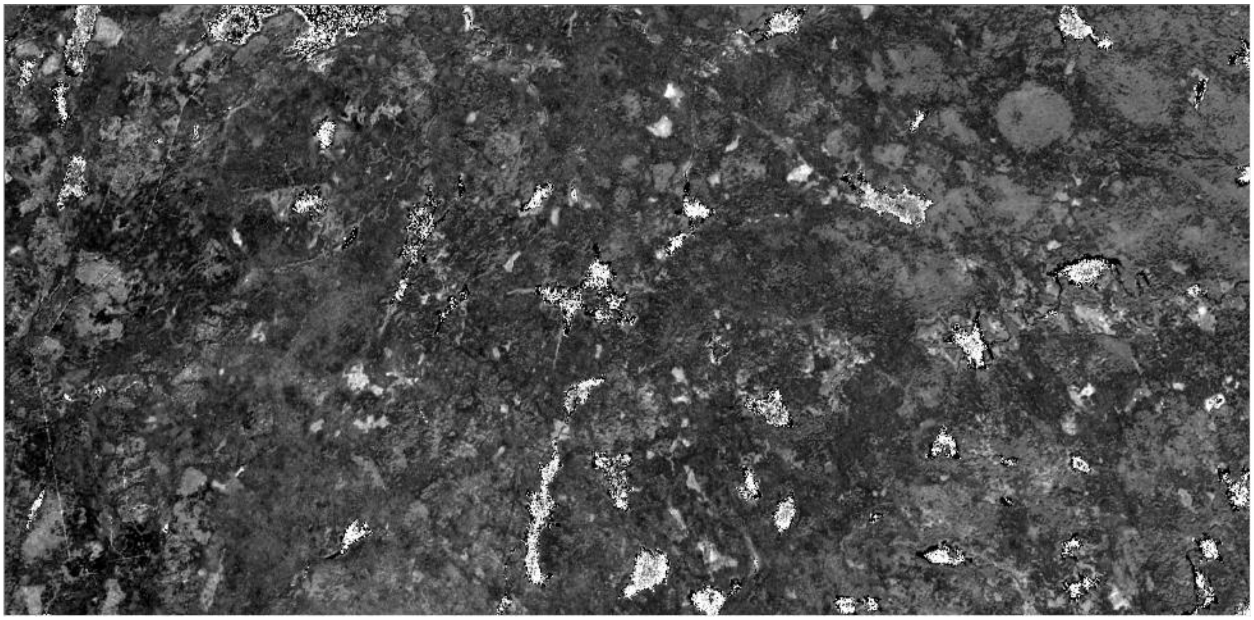
*Figure 13 shows that the NARI index is capable of further differentiating vegetated environments. Note for example the prevalence of bight areas bounding some low laying lake perimeter. (Tusket Lake Area, NS)*

The following equation for NARI is provided by Bayle et al. (2019):

$$NARI = \frac{\frac{1}{B3} - \frac{1}{B5}}{\frac{1}{B3} + \frac{1}{B5}}$$

#### 1.4.2.4 Red Edge Inflection Point (REIP)

The Red-Edge Inflection Point (REIP) is particularly valuable for wetland classification because it captures the precise transition between the red (R) and near-infrared (NIR) parts of the spectrum where chlorophyll absorption features can suddenly change. This shift not only reflects changes in chlorophyll content but also indicates various plant conditions such as Leaf Area Index (LAI), nutrient levels, and water content. These are all crucial for analyzing the health and, importantly, composition of wetland ecosystems. As chlorophyll content increases, the red-edge shifts to longer wavelengths, and vice versa, allowing detailed monitoring of vegetation dynamics within these environments (Herrmann et al. 2010). Similar to NARI, this index relies on coarser resolution Setntinel-2 bands (B5, red edge, 20m) and this produces a coarser index than RGBN or NDVI.



The following equation for REIP is provided by Herrmann et al. (2019):

$$REIP = 700 + 40\left[\frac{\left(\frac{B4+B7}{2}\right) - B5}{B6 - B5}\right]$$

#### 1.4.2.5 Sentinel-1 VV and HV

Sentinel-1, a European Space Agency satellite equipped with C-band Synthetic Aperture Radar (SAR), utilizes VV (vertical-vertical) and VH (vertical-horizontal) polarizations to distinguish surface features in wetland areas. These Sentinel-1 images are similarly available on GEE and can be made into seasonal median composites much the same way as was done with Sentinel-2 data, though without complication due to clouds which are generally not visible in C-band radar. VV polarization is particularly sensitive to variations in soil moisture and surface roughness, making it highly effective for detailed soil moisture mapping. VH polarization, on the other hand, is responsive to vegetation structure and water content, which is understood to be valuable for the classification and monitoring of wetland vegetation (Ma et al 2020).

To compensate for the systematic bias of terrain angle common to active radar system such as Sentinel-1, various ratios of observations (such as VV, VH) may be computed. In addition to VV and VH directly, we include a normalized difference ratio value calculated as such:

$$S1ND = \frac{VV - VH}{VV + VH}$$

### 1.4.3 Features Not Implemented

The following additional metrics have been identified to be valuable but have not been included in this study to date. They are provided here as a future reference for continued analysis. Generally, these layers are not included due to various processing constraints and/or prioritization, but each should be considered for future work.

#### 1.4.3.1 Normalized Difference Water Index (NDWI)

The Normalized Difference Water Index (NDWI) is effective for enhancing water features and assessing water clarity but is less suitable than NDVI for differentiating vegetation types in wetlands (McFeeters, 1996). This is computed by replacing the red band of NDVI with the green band, which has greater penetration in water. While it otherwise may not have significant improvement versus NDVI for terrestrial land types, it may be worth integrating for shallow water-type wetlands.

NDWI is described in McFeeters (1996) as such:

$$NDWI = \frac{B3 - B8}{B3 + B8}$$

#### 1.4.3.2 Moisture Stress Index (MSI)

The Moisture Stress Index (MSI), calculated as the ratio of near-infrared (NIR) to shortwave-infrared (SWIR) reflectance, is effective for identifying moisture content in vegetation and has been shown to improve the accuracy of peatland identification (Pang et al., 2020). While not incorporated into our study, MSI is noted for its potential to distinguish peatlands from marshes specifically and could be valuable for refinements to mapping wetland types (Hunt & Rock, 1989).

$$MSI = \frac{Rswir}{Rnir}$$

#### 1.4.3.3 Normalized Difference Infrared Index (NDII)

Normalized Difference Infrared Index (NDII) is recognized for its robust performance in detecting moisture content in wetland vegetation (Hardisky et al., 1983; Pang et al., 2020). NDII's application in differentiating marsh and peatland areas has been validated in studies and practical implementations in Nova Scotia, as noted by J. Gallop, NSECC (pers. comm., 2024). The following adapted equation is described by Hardisky et al. (1983):

$$NDII = \frac{Rnir - Rswir}{Rnir + Rswir}$$

#### 1.4.3.4 Enhanced Vegetation Index (EVI)

Enhanced Vegetation Index (EVI) is a variation of NDVI where the green vegetation signal is enhanced (Huete et al. 1997), and the atmospheric conditions and canopy background are reduced (Wardlow et al. 2007). EVI is used to monitor vegetation growth and coverage (Pang et al. 2020). The following EVI formula is taken from Wardlow et al. (2007):

$$EVI = G\left[\frac{\rho NIR - \rho red}{(\rho NIR + C1)(\rho red - C2)(\rho blue + L)}\right]$$

Where:

- $G$ = gain factor (2.5)
- $\rho$ = atmospherically corrected surface reflectance

- $C_1$ = aerosol resistance coefficient (6)
- $C_2$ = aerosol resistance coefficient (7.5)
- $L$ = canopy background adjustment (1)

#### *1.4.3.5 dual-pol Radar Vegetation Index (DpRVI)*

The effectiveness of the dual-pol Radar Vegetation Index (DpRVI) in quantifying vegetation biophysical parameters, as demonstrated by Mandal et al. (2020), indicates its potential in ecological monitoring applications. This represents one of many possible variants on methods to normalize Sentinel-1 to compensate for terrain angle and other biases. Mandal et al. (2020) describe DpRVI as:

$$\mathrm{DpRVI} = \frac{4\sigma\mathrm{VH}^{\circ}}{\sigma\mathrm{VV}^{\circ} + \sigma\mathrm{VH}^{\circ}}$$

Where σ is a dimensionless backscatter coefficient.

#### *1.4.3.6 Significant Water-Depth Increases (SWDI)*

Zhang et al. (2022) present a framework to detect Significant Water-Depth increases (SWDI) directly observable from Sentienl-1 (S1) SAR backscatter. This may be of specific use when considering the role of wetland features with respect to flood water abatement. The SWDI is calculated in part using Normalized Difference Backscatter Index (NDBI) which considers S1 observations before and after a flood events, with the equation:

$$\mathrm{NDBI} = (\sigma^{\circ}\mathrm{t} - \sigma^{\circ}\mathrm{p}) \,/\, \mathrm{SDp}$$

Where:

- $\sigma^{\circ}t$ is the target image pixel backscatter
- $\sigma^{\circ}p$ is the mean pre-event backscatter
- and SDp is the standard deviation of the pre-event backscatter values.

#### *1.4.3.7 Topographic Position Index (TPI)*

The Topographic Position Index (TPI) is a measure calculated by dividing the elevation of a specific cell in a Digital Elevation Model (DEM) by the mean elevation of a designated neighborhood around that cell in various ways. This metric is used to identify the relative position of landforms, while not distinguishing between ridges, valleys, and flat areas. TPI values can indicate areas likely to accumulate water, thus aiding in the identification of potential wetland locations (De Reu et al., 2013; Vinod, 2017). For our study, the desired effect of TPI is captured with various DEV calculations as described. Generally, though variants exist:

$$TPI = \frac{cell\ elevation}{mean\ elevation\ of\ surrounding\ neighborhood}$$

#### *1.4.3.8 Curvature*

Halabisky et al. (2023) demonstrate the effectiveness of integrating multi-scale topographic indices, including curvature, for identifying wetland formation areas. The study utilized curvature to highlight landscape depressions where water may accumulate, which is valuable for inferring wetlands obscured by forest canopy or that are ephemeral in nature. This may allow for the enhancement of wetland mapping accuracy by delineating areas prone to water accumulation and retention. Curvature at a point on a surface, generally, can be calculated using partial derivatives of the elevation (Z) with respect to the horizontal coordinates in various methods (such as a plan or profile curvature). In practice, curvature metrics tend to be highly sensitive to subtle errors in lidar and may appear complex over otherwise subtle terrain variations.

#### *1.4.3.9 Valley Bottom Flatness*

Gallant and Dowling (2003) describe multiresolution index of valley bottom flatness (MRVBF) which leverages the DEM to delineate depositional areas. This index may have significant correlation to the described geomorphon scalar value, but likely may still warrant further exploration.

#### *1.4.3.10 SAGA Topographic Wetness Index*

While our methodology utilizes a streamlined calculation of the wetness index as outlined, Mattivi et al. (2019) present an extensive comparison of more complex algorithms, like the SAGA GIS tools, to derive more nuanced water accumulation metrics. These approaches were not examined in detail for this study.

### 1.4.4 Existing Wetland Map Products of Nova Scotia

#### *1.4.4.1 Nova Scotia Wetland Inventory*

As noted previously, Nova Scotia Department of Natural Resources has an existing inventory of wetlands which does span the entire province. This assessment was digitized based on large-scale photography and updated with satellite imagery. The schema of wetland classifications used conform to the the Canadian Wetland Classification System, and as such, it does appear this this dataset has been integrated into the larger Canadian National Wetlands Inventory (CNWI). These data can be most easily accessed through the Provincial Landscape Viewer (https://nsgi.novascotia.ca/plv/). Modified versions of this layer are available in various forms though the Nova Scotia department of Environment and Climate Change directly.

#### *1.4.4.2 Ecological Land Classification for Nova Scotia*

An alternative to the official Nova Scotia Wetland Inventory exists as part of the wider Ecological Land Classification (ELC) for Nova Scotia (Neily 2003 and 2017). This analysis was conducted across Nova Scotia whereby the province is divided into 9 distinct Ecoregions, and further divided into 39 Ecodistricts. This extensive mapping exercise was based on Nova Scotia’s Bio-physical Land Classification (BLC) whereby the province was segmented into small distinct topographical base units where topographic position, soil drainage, and soil texture attributes were considered. This was refined with extensive expert knowledge, forest inventory data, and field validation to provide a meaningful and comprehensive ecological categorization across the province. The smallest ecological unit, Ecosections, covers the province at approximately 1:10,000 scale. This product includes wetland classifications which are noted as poorly drained Ecosections of smooth topography (Neily 2017).

# 2 METHODOLOGY

## 2.1 DATA SOURCES AND INPUTS

The following is a precise list of all data utilized in the methods of this report. Additional information related to each can be found in the previous section. The coverage of each listed data includes the entirety of Nova Scotia.

| *Name* | *Description* | *Details* | *Source* | *Resolution* | *Period* |
|---|---|---|---|---|---|
| *DEM* | Lidar Elevation Model | CGVD2013 m | GeoNova/AGRG | 1m | 2016-2020 |
| *CHM* | Lidar Canopy Height Model | Height in m | GeoNova/AGRG | 1m | 2016-2020 |
| *S1P* | Sentinel 1 Spring Mosaic | Bands 2,3,4,8 | GEE | 10m | 2014-2023 |
| *S2P* | Sentinel 2 Spring Mosaic | Bands 2,3,4,8 | GEE | 10m | 2017-2023 |
| *S2F* | Sentinel 2 Fall Mosaic | Bands 2,3,4,8 | GEE | 10m | 2017-2023 |
| *S2F20* | Sentinel 2 Fall MSI Mosaic | Bands 5,6,7,8A,11,12 | GEE | 20m | 2017-2023 |
| *DTW* | Depth to Water | Height in cm | NSECC | 5m | 2016-2020 |
| *PTS* | Wetland Class ID | Expert Identified | NSECC | 2,041 points | |

### 2.1.1 Provincial Lidar Data

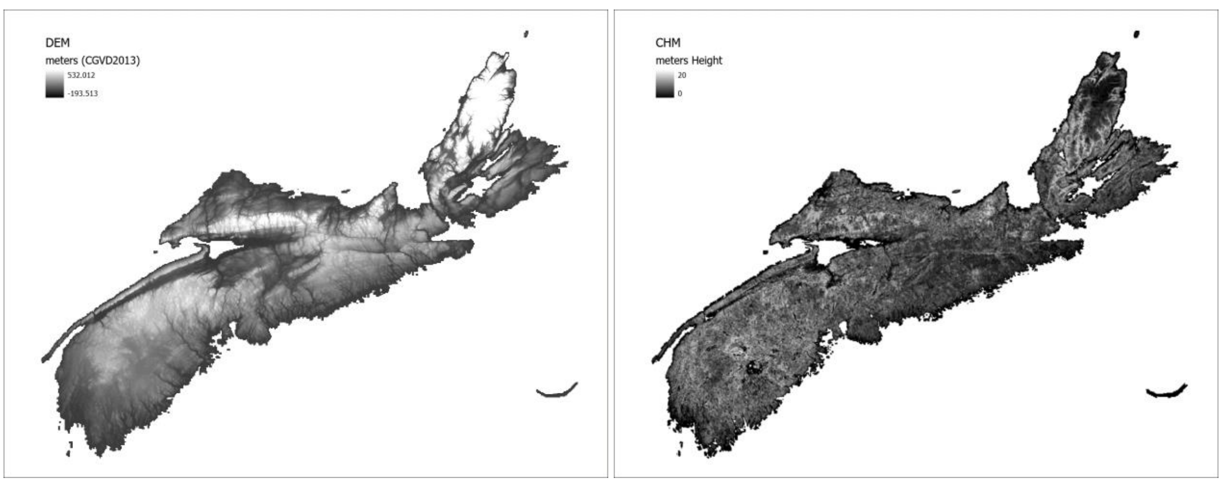



*Figure 14 shows the overview of both the lidar derived elevation model (left) and the canopy height model (right). These data are the culmination of many individual aerial lidar scans across the province conducted between 2016 and 2020.*

The Applied Geomatics Research Group (AGRG) has access to lidar DEM and CHM tiles provided by GeoNOVA. Through various research activities, the AGRG routinely refines aspects of the combined DEM product, which is utilized here. Both this and the CHM model (which has been mosaiced directly from GeoNova tiles) are native 1m resolution rasters in vertical meters.

### 2.1.2 Satellite Data

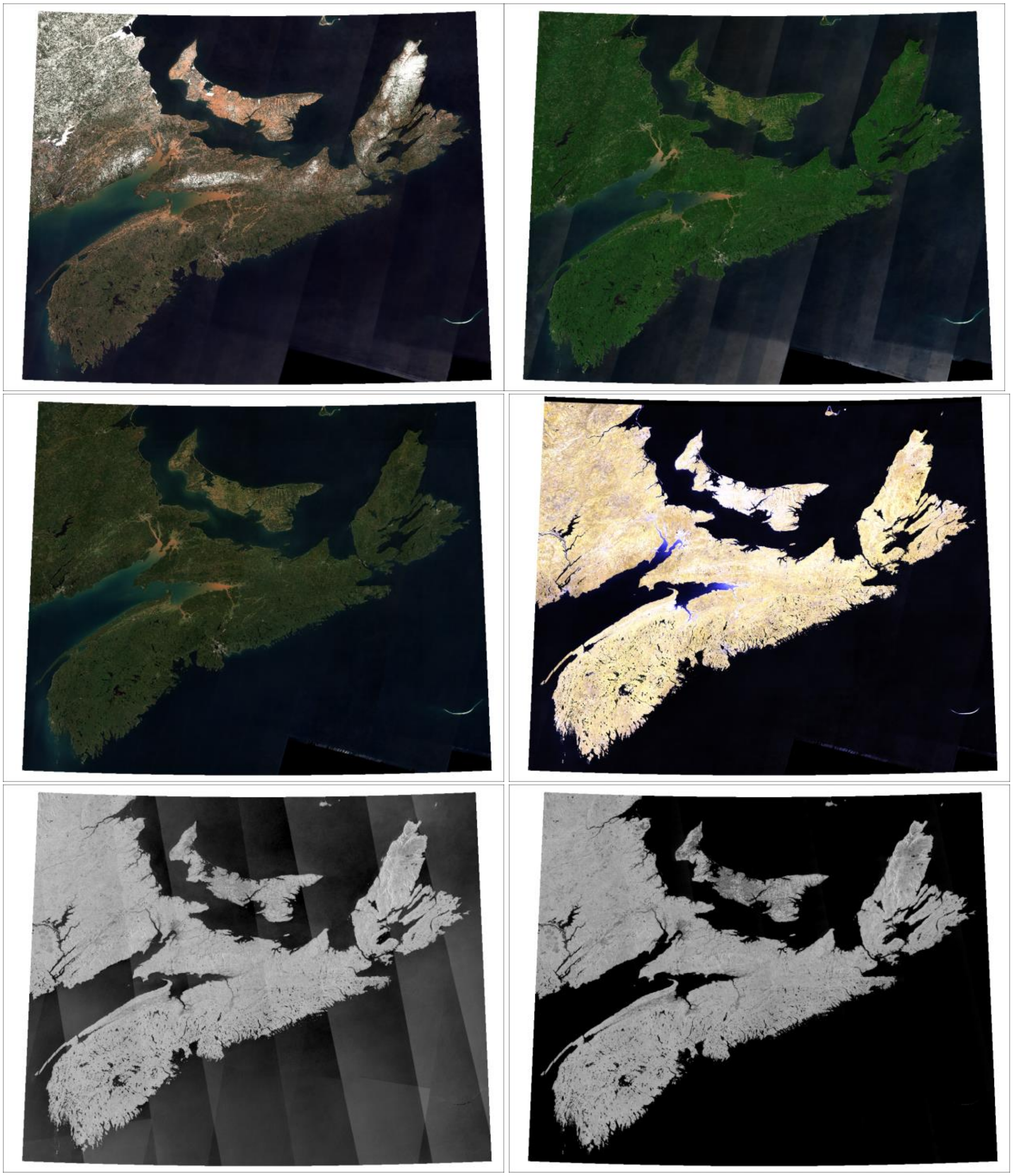


*Figure 15 shows province wide satellite mosaics built using GEE. Each data are compiled in using seasonal median across all scenes as described, and advanced cloud filtering is applied where appropriate. First are sentinel-2 RGB NIR products at a native 10 m including (S2P, top left) springtime, (S2S, top right) summer, (S2F mid-left) fall [NIR bands not shown]. Mid-right shows 20m native fall season Sentinel-MSI composite (S2F20, bands 5,6,7 shown). Bottom shows Sentinel-1 radar spring medians for Vertical-Vertical polarization (S1PVV, left) and Vertical-Horizontal (S1PVH, right).*

Satellite data (indicated in the figure and table above) were processed with the considerations described in the previous section. This includes advanced cloud filtering, and median pixel calculation of all available scenes, per band, for each satellite. Cloud filtering was not required for Sentienl-1 radar data. Seasonal mosaics were computed by filtering collections by Julian day. Overall, satellite datasets are generated from Google Earth Engine where:

**Sentinel-2 Median Cloud-Filtered Composites spanning 2017-03-28 to 2023-12-08:**

Derived from Sentinel-2 imagery within the 'COPERNICUS/S2_SR_HARMONIZED' dataset in GEE, representing the median, cloud-filtered value using a cloud mask from 'GOOGLE/CLOUD_SCORE_PLUS/V1/S2_HARMONIZED' with a threshold of 0.7.

Where ‘spring’ includes data spanning March 1st to May 31st (Julian days 60-151) annually, ‘summer’ includes June 1st to August 31st (Julian days 152-243), and ‘fall’ September 1st to November 30th (Julian days 244-334). Each seasonal mosaic include Sentinel-2 bands 2,3,4 and 8 (R,G,B,NIR) at a native 10m resolution whereas only a fall mosaic was generated for Bands 5, 6, 7, 8A, 11, 12 (Red-edge, SWIR etc) at a native 20m resolution.

**Sentinel-1 Median Composites spanning 2014-10-03 to 2023-12-08:**

Derived from Sentinel-1 imagery within the 'COPERNICUS/S1_GRD' dataset in GEE, representing the median ‘spring’ includes data spanning March 1st to May 31st (Julian days 60-151) annually, including both Vertical-Vertical (VV) and Vertical-Horizontal (VH) polarization image bands at a native 10m resolution.

All satellite data are exported from GEE to be utilized in ArcGIS with the extent based on the GEE dataset 'FAO/GAUL_SIMPLIFIED_500m/2015/level1' to represent full coverage of Nova Scotia, expanded by a 10km buffer.

### 2.1.3 Training/Validation Data

Wetland identifying training points were provided by the Nova Scotia Department of Environment and Climate Change (NSECC). These data, totaling 2,041 data geographic points, constituted high-quality expert derived wetland locations. These data were used for both training and validation of the classification model.

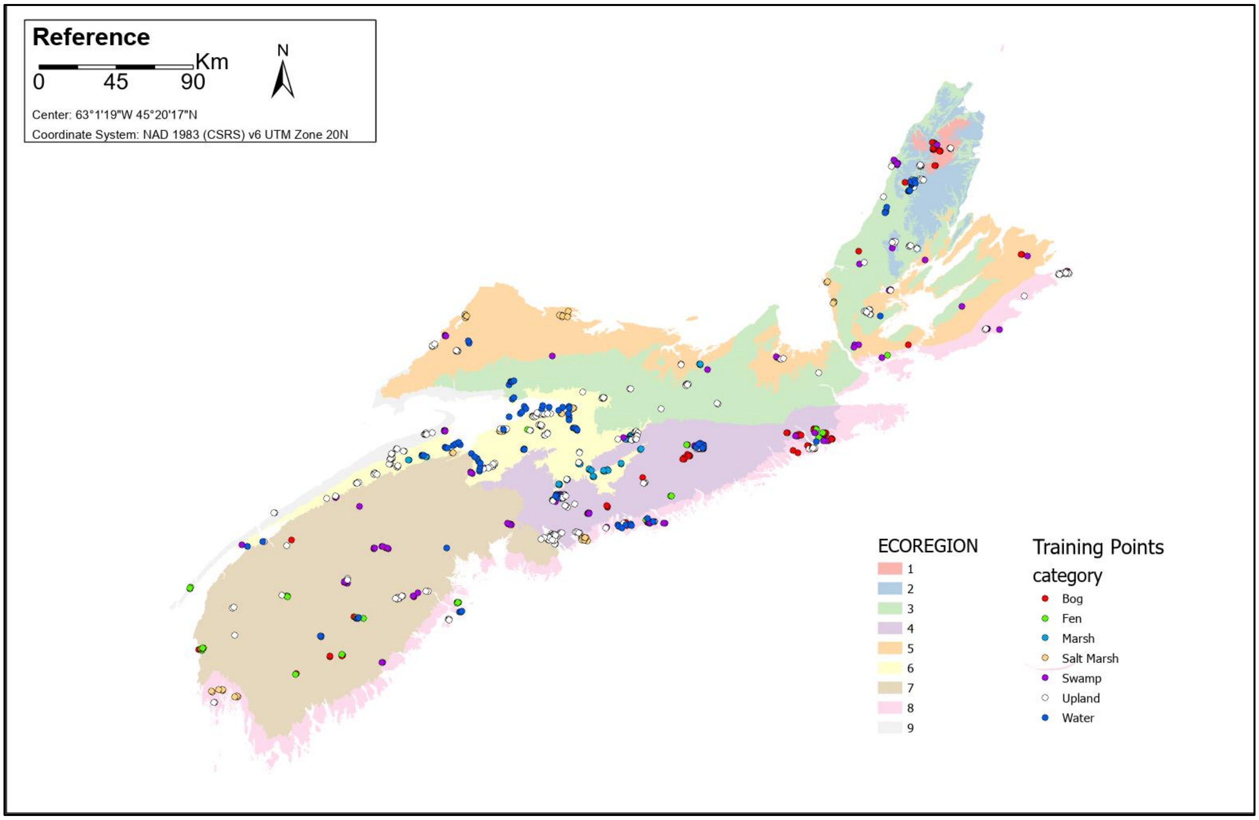



*Table 3 shows the total count of training/validation data provided by NSECC.*

| *Wetland Category* | *Count* |
|---:|---:|
| *Bog* | 342 |
| *Fen* | 218 |
| *Marsh* | 196 |
| *Salt Marsh* | 179 |
| *Swamp* | 303 |
| *Upland* | 630 |
| *Water* | 173 |

High quality training data such as this with a consistent source from an expert interpretation are critical to the success of such classification routines. That these data points are extensive and cover the entirety of the province, and span each of the Nova Scotia Ecoregions - provides our model with the foundation for general applicability across the province in various ecological contexts when trained.

### 2.1.4 Ancillary Data

Published locations for wetlands in Nova Scotia were provided by the NSECC in the form of the official Nova Scotia Wetland Inventory. This data proved useful as an overall benchmark and spatial comparison to our results. Additionally, the wetlands identified in the Nova Scotia Ecological Land Classification (2017) proved to be a variable reference to compare against our results in a broad sense.

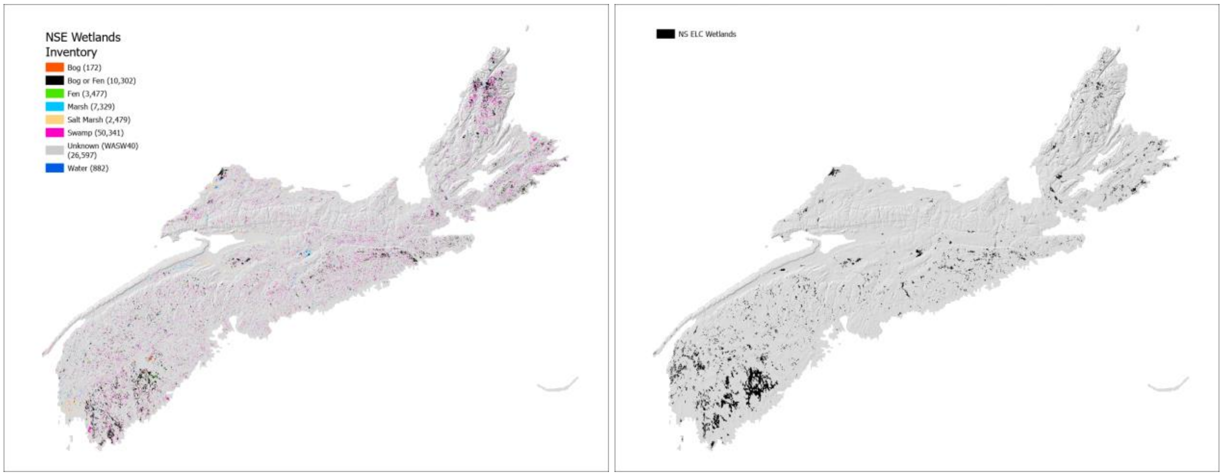



*Figure 16 show the Nova Scotia Wetland Inventory consisting of 101,579 total wetland polygons which closely adheres to the Canadian wetland classification schema (NSE Wetlands). Wetland extents found in the finest ecological units of the Nova Scotia Ecological Classification of 2017, provide additional context and information which proved useful for a wider consideration of our results.*

## 2.2 Overview of the Proposed Methodology

For our approach, we constructed 38 key wetland classification rasters chosen based on expert discussion and literature review. The overall process then relies on the AutoML tools found in the GeoAI suite of tools in ArcGIS 3.1.1. Utilizing the best available source for lidar ground and canopy elevations, as well as optical and radar satellite information, a complete provincial wide suite of image composites was constructed at 10 meter spatial resolution. The **'Train Using AutoML'** ArcGIS tool is then used to compute a best ensemble classification model based on wetland classification points provided by NSECC. A segmentation method was then conducted on selected layers wherein a full prediction and classification was performed on each segment based on the average value of for each of the 38 layers.

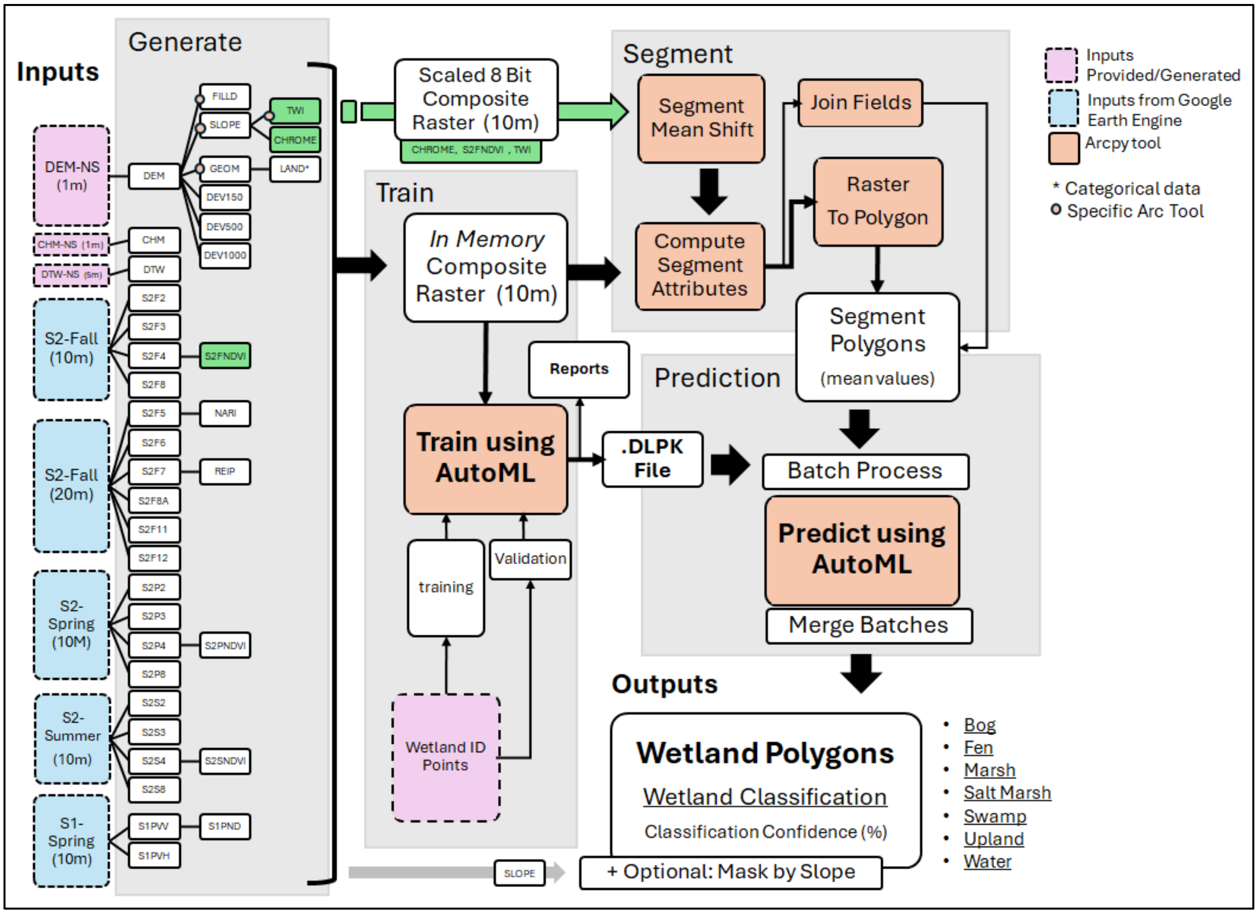



*Figure 17 shows the overall proposed wetland mapping procedure. Each of the input datasets were resampled to a consistent 10-meter resolution and formed a total of 38 generated feature layers. These layers then form the basis for training an ensemble classification model. The model is predicted on polygon features formed from segmenting a composite of layers highly sensitive to wetland delineation (TWI, CHROME, NDVI).*

## 2.3 Data Collection Methods

While there is a large list of elevation and satellite data types and sources, this project attempted to consolidate these to a minimum set for simplicity while also maintaining high quality and province wide applicability. As such this study relies heavily on the existing province wide lidar raster data, made available by GeoNOVA, and sources all satellite information from Google Earth Engine (GEE). While this access to GEE was conducted under a research license and therefore data is not directly shareable in all cases, this source was selected for rapid and high-quality generation of province wide seasonal cloud free mosaics.

### 2.3.1 Lidar Data

Lidar data, specifically Digital Elevation Model (DEM) and Canopy Height Model (CHM) tiles, were acquired from GeoNOVA for the years 2016 to 2020 at a spatial resolution of 1 meter. The lidar DEM is referenced to the Canadian standard vertical datum CGVD2013 whereas the CHM was referenced as relative height. Buildings and some artifacts such as tile and water boundaries were present in the CHM data and significant cleaning of these data were outside the scope of this analysis. These tiles were then mosaiced together using ArcGIS to create comprehensive province-wide datasets for each. Special attention was paid to ensure best available data was preferentially mosaiced into the final dataset.

### 2.3.2 Satellite Data

Seasonal mosaics of Nova Scotia were generated using Google Earth Engine's Sentinel-2 collection for the years 2017 to 2023, employing the Cloud Score+ filter on harmonized imagery for cloud-free, median pixel value composites:

- Spring: March 1st to May 31st annually.
- Summer: June 1st to August 31st annually.
- Fall: September 1st to November 30th annually.
- Winter: December 1st to February 28th/29th annually.

This process was applied to all 10m resolution bands of Sentinel-2. Additionally, a separate mosaic for the fall season was created for all 20m resolution bands, ensuring comprehensive coverage and detail of the landscape across all available spectral bands for each season. These data consist of 16-bit top of atmosphere reflectance values. Additional information including a total list of Sentinel-2 bands is available in the references (Google Developers, 2024). Sentinel-2 data was the basis for additional computed metrics as described including NARI, REIP, and NDVI. Additionally, each Sentienl-2 band for each mosaic is included directly in the analysis.

Sentinel-1 data was also acquired in a similar manner (as the median 10 m pixel) for the described spring timeframe and otherwise covering the entire timespan. Two radar polarization options were included, vertical-vertical (VV), and cross-polarized vertical-horizontal (VH), which each provide a differing noise and reflectance profile. While C-band radar such as Sentinel-1 is generally not sensitive to clouds, reflectance can be heavily impacted by incident angle such as terrain angle. To attempt to compensate these effects, a normalized difference was also computed between VV and VH.

## 2.4 Data Analysis Procedures

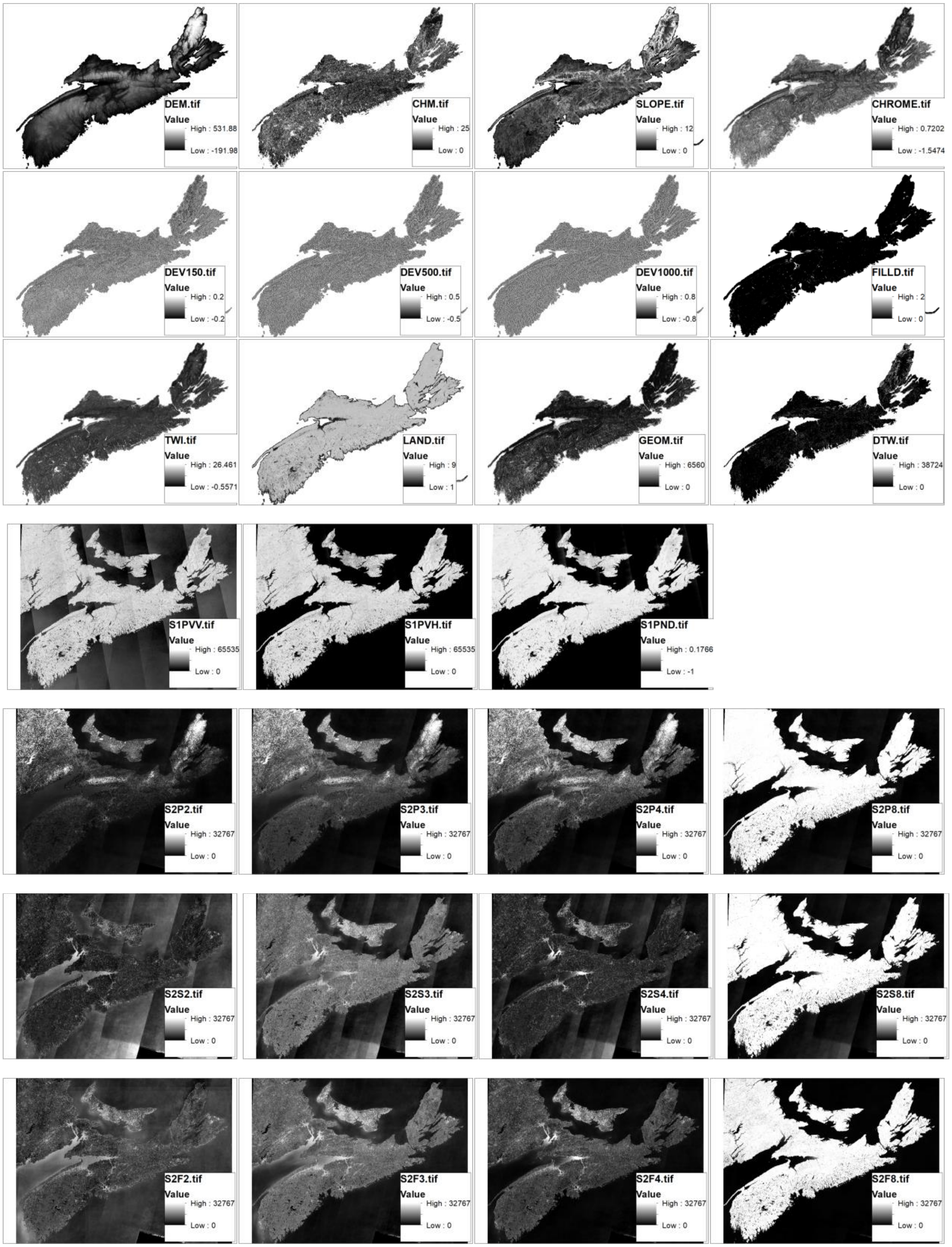
### 2.4.1 Compilation of desired Feature Layers

A series of Python scripts were constructed to generate each of the desired rasters, selected from the list of desirable features for wetland mapping. Each layer was resampled to 10 meters using bilinear or nearest neighbor as appropriate. A complete list of the included layers is as follows and form the inputs features or 'explanatory raster' information for later machine learning classification steps:

*Table 4 complete list of feature layers which are directly resampled or computed from input datasets. A full description for each can be found in the previous section. Details for the computation can be found in the accompanying python script.*

| *Number* | ***ID Name*** | ***Full Name*** |
|---|---|---|
| *1* | DEM | Digital Elevation Model |

| | | |
|---|---|---|
| 2 | CHM | Canopy Height Model |
| 3 | SLOPE | Lidar Derived Terrain Slope |
| 4 | CHROME | Chrome DEM (Custom Metric) |
| 5 | DEV150 | Deviation of Mean Elevation (150 meter) |
| 6 | DEV500 | Deviation of Mean Elevation (500 meter) |
| 7 | DEV1000 | Deviation of Mean Elevation (1000 meter) |
| 8 | FILLD | Hydrological Sinks, Filled Depth |
| 9 | TWI | Topographic Wetness Index |
| 10 | LAND | Geomorphon Land Cover Index |
| 11 | GEOM | Geomorphon Scalar Value |
| 12 | FILLD | Hydrological Sinks, Filled Depth |
| 13 | S1PVV | Sentinel-1 Spring Vertical-Vertical Polarization |
| 14 | S1PVH | Sentinel-1 Spring Vertical-Horizontal Polarization |
| 15 | S1PND | Sentinel-1 Spring VV/VH Normalized Difference |
| 16 | S2P2 | Sentinel-2 Spring, Band 2 (Blue) |
| 17 | S2P3 | Sentinel-2 Spring, Band 3 (Green) |
| 18 | S2P4 | Sentinel-2 Spring, Band 4 (Red) |
| 19 | S2P8 | Sentinel-2 Spring, Band 8 (NIR) |
| 20 | S2S2 | Sentinel-2 Summer, Band 2 (Blue) |
| 21 | S2S3 | Sentinel-2 Summer, Band 3 (Green) |
| 22 | S2S4 | Sentinel-2 Summer, Band 4 (Red) |
| 23 | S2S8 | Sentinel-2 Summer, Band 8 (NIR) |
| 24 | S2F2 | Sentinel-2 Fall, Band 2 (Blue) |
| 25 | S2F3 | Sentinel-2 Fall, Band 3 (Green) |
| 26 | S2F4 | Sentinel-2 Fall, Band 4 (Red) |
| 27 | S2F8 | Sentinel-2 Fall, Band 8 (NIR) |
| 28 | S2PNDVI | Sentinel-2 Spring, NDVI Composite |
| 29 | S2SNDVI | Sentinel-2 Summer, NDVI Composite |
| 30 | S2FNDVI | Sentinel-2 Fall, NDVI Composite |
| 31 | S2F5 | Sentinel-2 Fall, Band 5 |
| 32 | S2F6 | Sentinel-2 Fall, Band 6 |
| 33 | S2F7 | Sentinel-2 Fall, Band 7 |
| 34 | S2F8A | Sentinel-2 Fall, Band 8A |
| 35 | S2F11 | Sentinel-2 Fall, Band 11 |
| 36 | S2F12 | Sentinel-2 Fall, Band 12 |
| 37 | NARI | Normalized Anthocyanin Reflectance Index |
| 38 | REIP | Red Edge Inflection Point |


DEM.tif
Value
High : 531.88
Low : -191.98
CHM.tif
Value
High : 25
Low : 0
SLOPE.tif
Value
High : 12
Low : 0
CHROME.tif
Value
High : 0.7202
Low : -1.5474
DEV150.tif
Value
High : 0.2
Low : -0.2
DEV500.tif
Value
High : 0.5
Low : -0.5
DEV1000.tif
Value
High : 0.8
Low : -0.8
FILLD.tif
Value
High : 2
Low : 0
TWI.tif
Value
High : 26.461
Low : -0.5571
LAND.tif
Value
High : 9
Low : 1
GEOM.tif
Value
High : 6560
Low : 0
DTW.tif
Value
High : 38724
Low : 0
S1PVV.tif
Value
High : 65535
Low : 0
S1PVH.tif
Value
High : 65535
Low : 0
S1PND.tif
Value
High : 0.1766
Low : -1
S2P2.tif
Value
High : 32767
Low : 0
S2P3.tif
Value
High : 32767
Low : 0
S2P4.tif
Value
High : 32767
Low : 0
S2P8.tif
Value
High : 32767
Low : 0
S2S2.tif
Value
High : 32767
Low : 0
S2S3.tif
Value
High : 32767
Low : 0
S2S4.tif
Value
High : 32767
Low : 0
S2S8.tif
Value
High : 32767
Low : 0
S2F2.tif
Value
High : 32767
Low : 0
S2F3.tif
Value
High : 32767
Low : 0
S2F4.tif
Value
High : 32767
Low : 0
S2F8.tif
Value
High : 32767
Low : 0

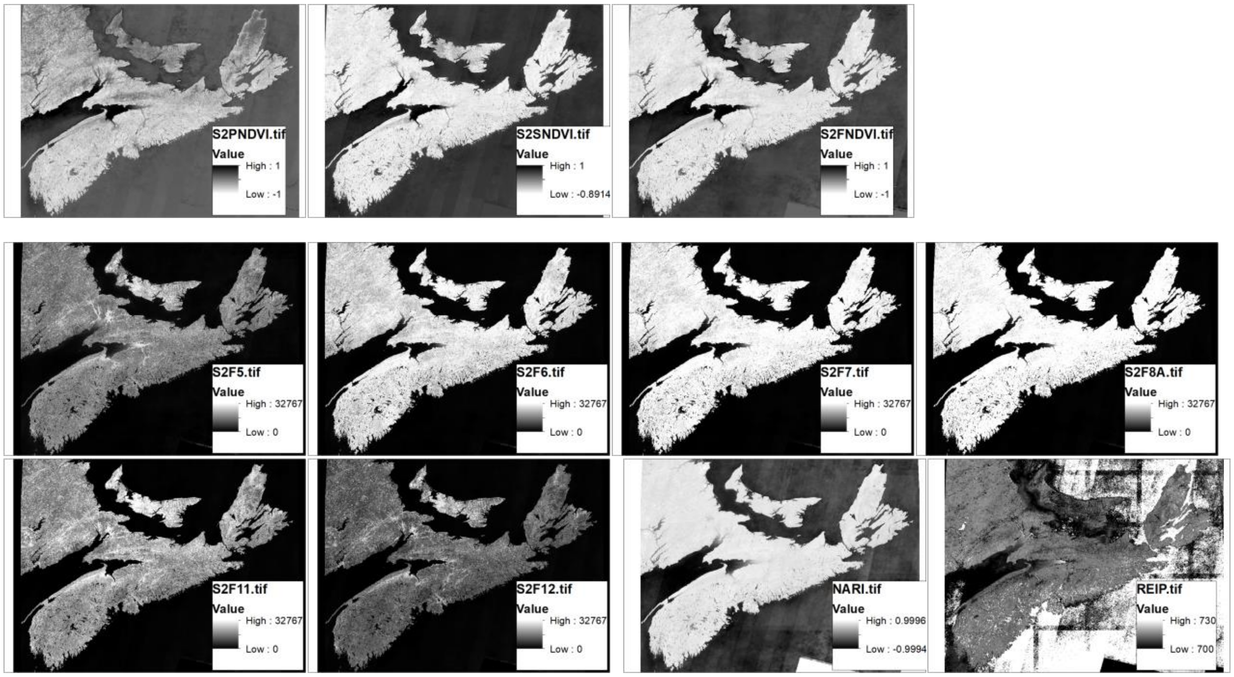



*Figure 18 shows a graphical summary of all included explanatory rasters used for the wetland classification model. All rasters are derived lidar or satellite seasonal mosaics generated using google earth engine (GEE). Depth to water (DTW) was provided by NSECC at a native 5m resolution and resampled 10 m to be consistent with all other raster layers.*

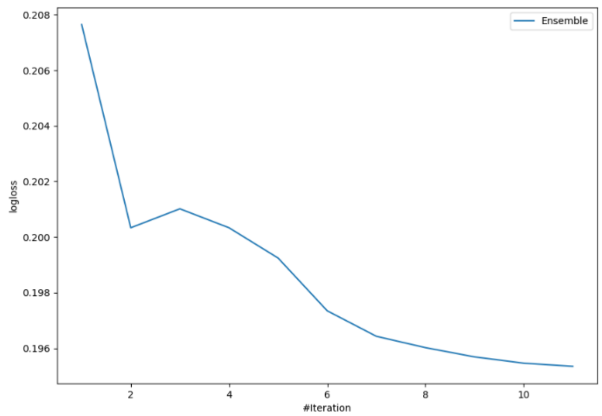
### 2.4.2 Training AutoML ensemble Model

Using the ArcGIS GeoAI ***Train Using AutoML*** tool, we extract point values for each of the training data points provided by NSECC for all explanatory features. A random sample of 10% of the provided wetland identification points were withheld from the training to validate the model, which is the default validation percentage for the AutoML tool. This validation step is critical for assessing the model's accuracy and generalizability to unseen data. The performance of each classifier is tested using a standard Log loss function which tests both accuracy and confidence. The resulting model forms the best ensemble classifier based on a suite of attempted classification routines. Higher performing classifiers are weighted into the final ensemble classification model including LightGBM, CatBoost, and Linear classifiers.

*Table 5 shows the total scope of models utilized by the AutoML classifier in ArcGIS pro. Top performing classifiers are used to generate a single Ensemble model*

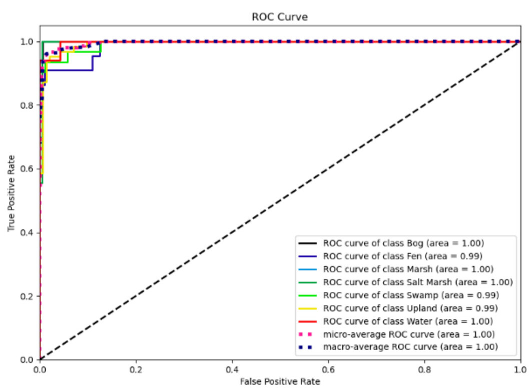
| *Model name* | *Model type* | *Metric type* | *Metric value* | *Training time (Seconds)* |
|---|---|---|---|---|
| 1_DecisionTree | Decision Tree | logloss | 1.23063 | 0.64 |
| 2_DecisionTree | Decision Tree | logloss | 0.991596 | 0.65 |
| 3_DecisionTree | Decision Tree | logloss | 0.991596 | 0.66 |
| 4_Linear | Linear | logloss | 0.325585 | 0.91 |
| 5_Default_LightGBM | LightGBM | logloss | 0.276918 | 3.61 |
| 6_Default_Xgboost | Xgboost | logloss | 0.24405 | 4.68 |
| 7_Default_CatBoost | CatBoost | logloss | 0.207643 | 31.53 |

| | | | | | |
|---|---|---|---|---|---|
| | 10_Xgboost | Xgboost | logloss | 0.217611 | 4.11 |
| | 11_Xgboost | Xgboost | logloss | 0.247786 | 4.26 |
| | 12_LightGBM | LightGBM | logloss | 0.274184 | 2.15 |
| | 15_Xgboost | Xgboost | logloss | 0.225407 | 5.83 |
| *the best* | **Ensemble** | **Ensemble** | **logloss** | **0.195354** | **0.25** |




The AutoML tool in ArcGIS provides several valuable analysis outputs to gauge the performance of the classification. Especially valuable are the Log Loss over each iteration, and the resulting Receiver Operating Characteristic (ROC) and precision-recall curves of the ensemble model.

*Figure 19 shows the calculated log loss (accuracy and confidence) improve for the ensemble model as the classifier trains iteratively.*



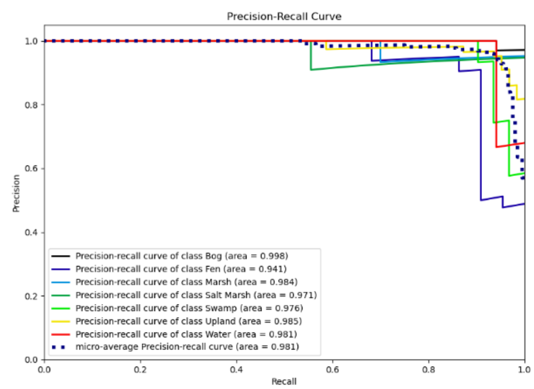




*Figure 20 shows the Receiver Operating Characteristic (ROC) Curve which indicates the model's ability to differentiate between classes vs random, while the Precision-Recall Curve indicates % correct of the class (precision) and % of the class correct (recall)*

The AutoML tool generates a DLPK (deep learning model package file) which can then be utilized with several ArcGIS tools to generate an output prediction of each wetland type for a given set of explanatory inputs. Note that all explanatory values used in training the model are required to generate a prediction.

### 2.4.3 Raster Segmentation

To apply the model across the province, the region was segmented into detailed vector polygons to provide the basis wetland prediction. This was accomplished using the ArcGIS Segment Mean Shift tool. To provide a good balance between processing time and the ability to discriminate wetlands finely, segmentation was performed on a 10 m resampling. The tool can accept a 3-band raster input so a composite of Chrome, NDVI, and TWI was selected as it was visually assessed across the province to best differentiate between various features as indicated by the existing Nova Scotia wetland Inventory.

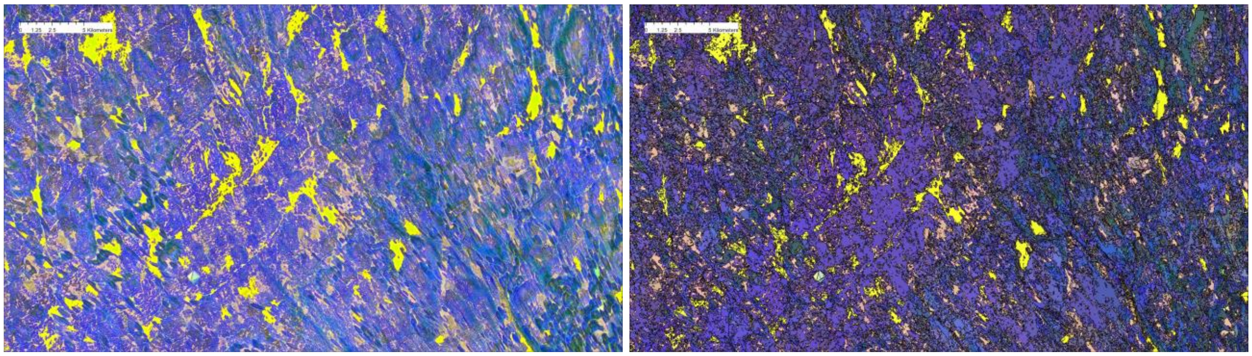


*Figure 21 shows the resulting segmentation as performed on the 10 m chrome, NDVI, and topographic wetness index feature rasters.*

In total this process generated 10,101,784 discrete segments. These segments were then attributed with average values for each of the indicated explanatory rasters using the ArcGIS tool ***Compute Segment Attributes***. Finally, a polygon feature class of all segments was generated using the ***Raster To Polygon*** tool and features for all explanatory rasters were added using the ***Join*** tool.

### 2.4.4 Predict Model to Segments

Once a single segment feature class covering the entire province including all and explanatory values was generated, the ***Predict Using AutoML*** was run in a batch script. This process was performed iteratively to ensure stability and to track progress.

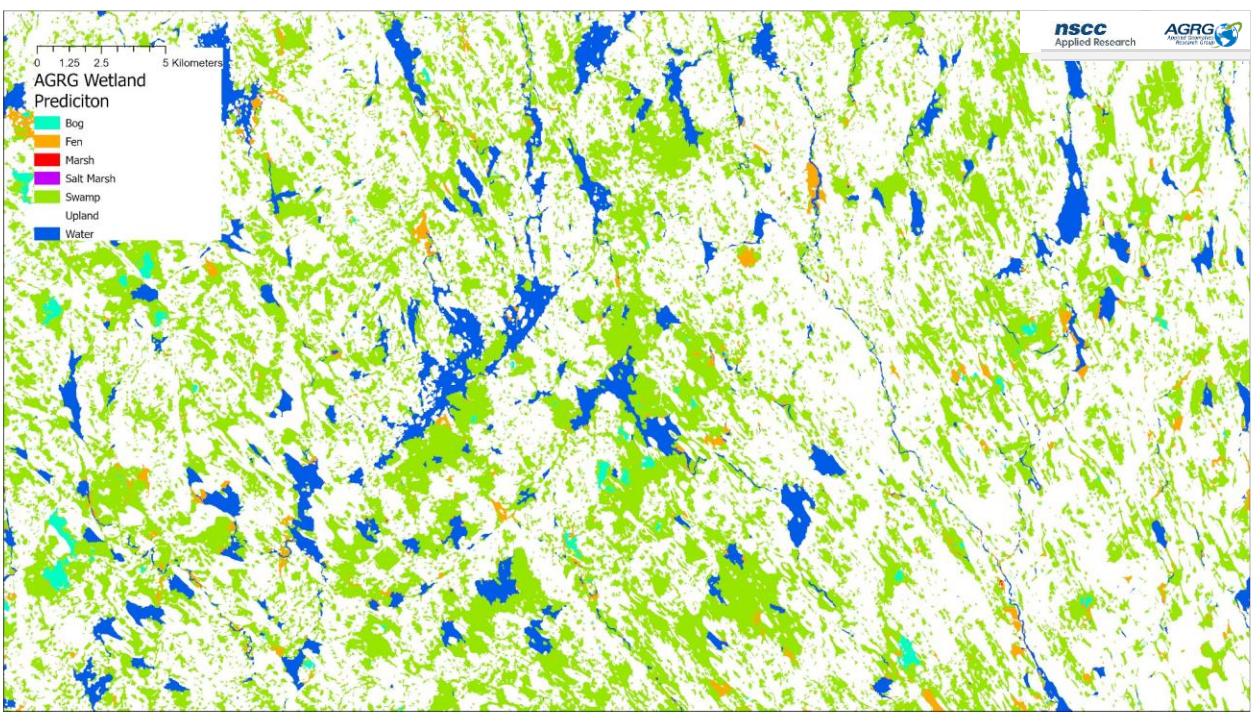



*Figure 22 our method poduces classified segments, where each segment has been assigned the average raster value for each of the 38 input layers.*

### 2.4.5 Methods Overview

The final set of wetland prediction results were then merged into a single wetland prediction feature class. At this point, data can be inspected, mapped, summarized, and disseminated. These data have been selected for wetland features (marsh, bog, fen, saltmarsh, swamp) and stripped of non-wetland features (uplands, water) to reduce the file size.

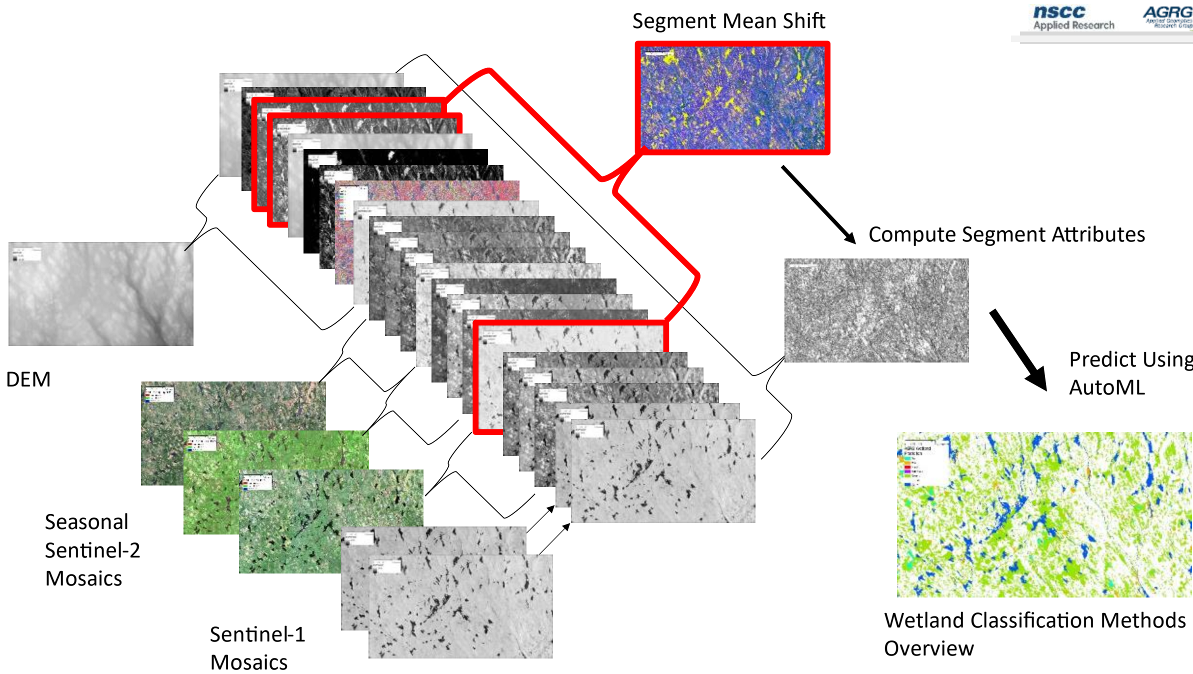



*Figure 23 shows the general graphical overview of the process. The main inputs are used to derive a total of 38 explanatory variables. 3 of these (CHROME, TWI, NDVI) are used to perform a province wide segmentation, producing 10 million + segments. These segments are assigned the average of the 38 layer values which are then used to predict the wetland classification type based on an appropriately trained model. The model presented was trained using 2000+ wetland type identifications points provided by NSECC, which were sampled to the same 38 input layers and trained using AutoML ADVANCED mode in ArcGIS. The prediction applied to the segments in performed in batches.*

# 3 Results

Utilizing the training data points, provided by NSECC, and the AutoML classifier tool available in ArcGIS GeoAI tools, we were able to generate a wetland prediction for 10,101,784 discrete segments with a combined estimated accuracy of approximately 94.1%. This combined area of 62,228 sq km covers the entire land area of Nova Scotia. This analysis is performed with an approximate spatial resolution of 10 m.

Of the total segments classified, approximately 61% of the area was determined ‘Upland’ (non-wetland), 15% water, and the remaining 24% as wetland. Of the total wetland area, swamps formed the vast majority at 87.9% (20.5% of the total province), followed distantly by Bogs (at 6.5% of wetlands, 1.2% of the province). The remaining Fens, Marshes, and Salt Marshes constitute less than 2% of the province collectively.

*Table 6 shows the total segments modelled across the province, including their total area per classification*

| Type | Segments | Area (Sq. Km) | Area (%) |
|---|---|---|---|
| Bog | 144,160 | 723.94 | 1.16% |
| Fen | 140,713 | 666.87 | 1.07% |
| Marsh | 29,404 | 114.24 | 0.18% |
| Salt Marsh | 53,799 | 231.86 | 0.37% |
| Swamp | 3,118,780 | 12,767.99 | 20.52% |
| Upland | 5,820,115 | 38,156.96 | 61.32% |
| Water | 794,813 | 9,566.54 | 15.37% |
| **Total** | **10,101,784** | **62,228.40** | **100.00%** |

*Table 7 shows the total number of segments province wide which are considered wetlands.*

| Type | Segments | Area (Sq. Km) | Area (%) |
|---|---|---|---|
| Bog | 42,797 | 321.10 | 2.46% |
| Fen | 136,895 | 842.13 | 6.44% |
| Marsh | 32,024 | 175.39 | 1.34% |
| Salt Marsh | 46,025 | 244.18 | 1.87% |
| Swamp | 2,151,865 | 11,494.33 | 87.90% |
| **Total** | **2,409,606** | **13,077.15** | **100.00%** |

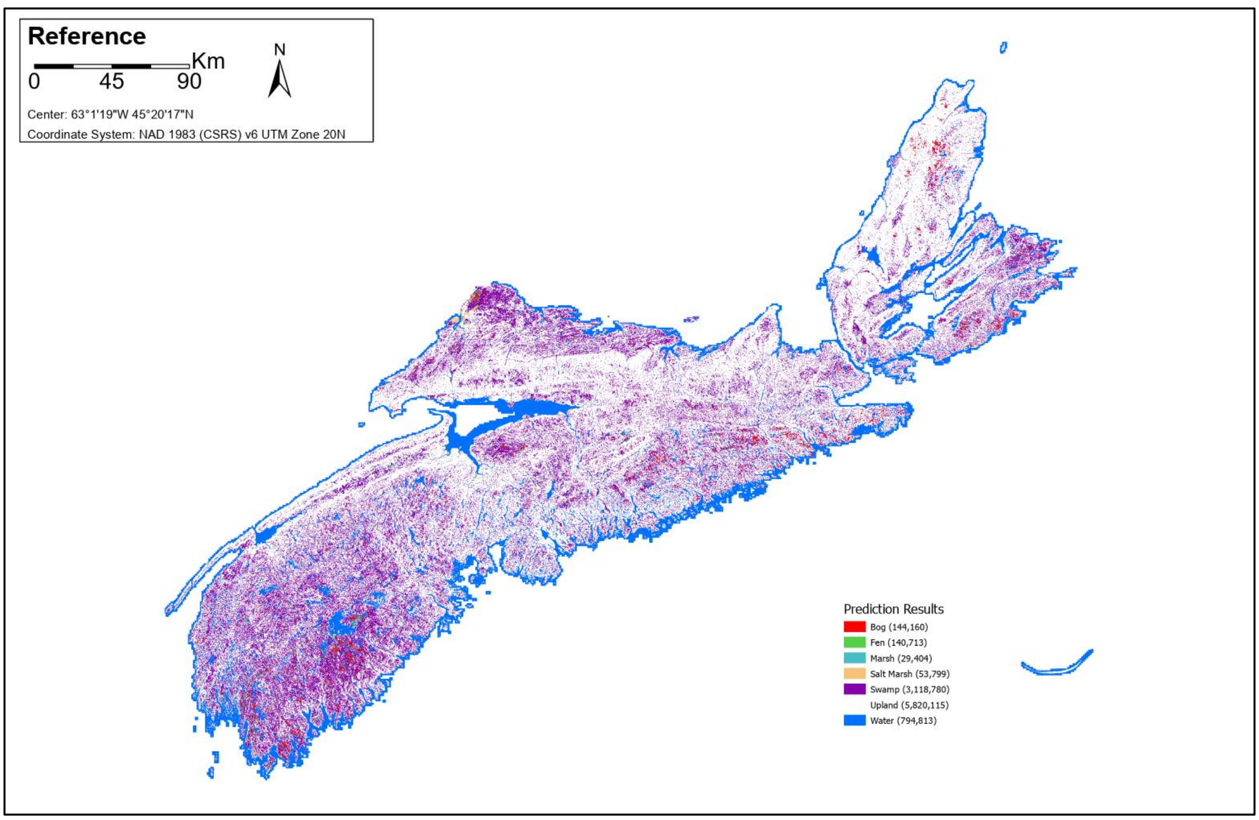



*Figure 24 shows the overall results of the AutoML prediction applied to each of the ~10 million segments across the province.*

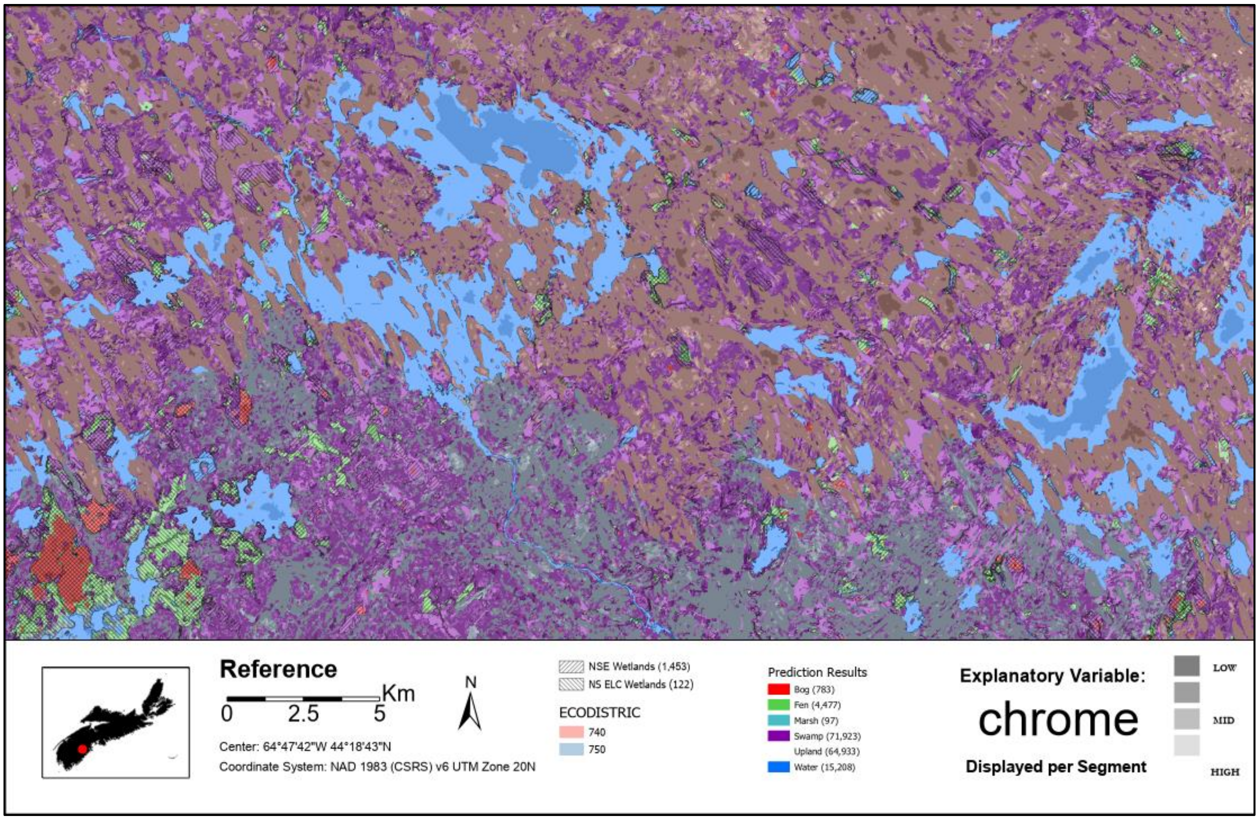



*Figure 25 shows one of various views by which the results have been inspected. Each segment, having been assigned a wetland classification through the prediction of our model – is overlain with the best available known wetland locations for Nova Scotia, namely, the wetland inventory (NSE Wetlands) and the ecological land classification wetlands (NS ELC Wetlands). Further we inspect the in context of the overlay with various predictor values including the high prediction importance CHROME value, here shows as the average CHROME value per segment, sampled from the 10m raster source.*

## 3.1 Detailed Analysis

While the overall accuracy of the model, based on 10% validation of the total 2,041 total training points reported **~ 94%** overall accuracy, the mean prediction confidence across all **10,101,784** segments tells a slightly different story, yielding a total mean Prediction confidence of **89.3%** with a standard deviation of 13.9%. This represents an improvement over previous model runs which achieved an overall average prediction confidence across all segments of 85.8% (std. 16.0%). This previous model run did not include critical feature layers such as NARI, REIP, CHM, DEV, or DTW – indicating their overall value to the results.

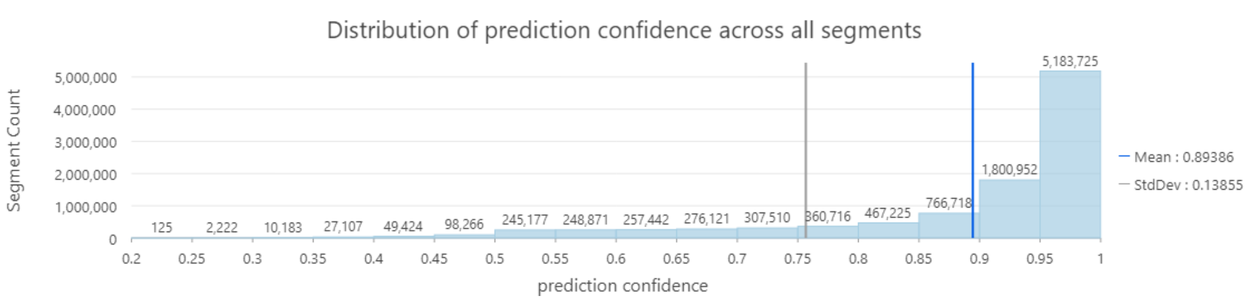


*Table 8 shows the Mean confidence % per category (and standard deviation) across all 10.1 million segments. Square bracket values indicate the mean percent confidence of previous model runs which did not include NARI, REIP, CHM, DEV, or DTW feature layers.*

| *Category* | *Mean Prediction Confidence (%)* | *Standard Deviation (%)* |
|---|---|---|
| *Bog* | **77.9** [59.4] | 18.0 |
| *Fen* | **72.7** [60.6] | 18.6 |
| *Marsh* | **66.4** [63.4] | 19.4 |
| *Salt Marsh* | **72.5** [70.1] | 17.6 |
| *Swamp* | **88.0** [85.0] | 14.9 |
| *Upland* | **90.8** [85.9] | 12.1 |
| *Water* | **91.8** [95.3] | 14.3 |

What this likely indicates is that, while the classifier produces highly accurate and validated prediction on all the expert selected wetland locations for a given set of explanatory pixel values, it becomes less confident when faced with the overall distribution of all explanatory pixel averaged into segments. This suggests that the model, while accurate by all accounts, could either be made more robust by training on segment data as opposed to directly on point samples, or by refining segments to be more precise. The latter will likely improve the model on all accounts.

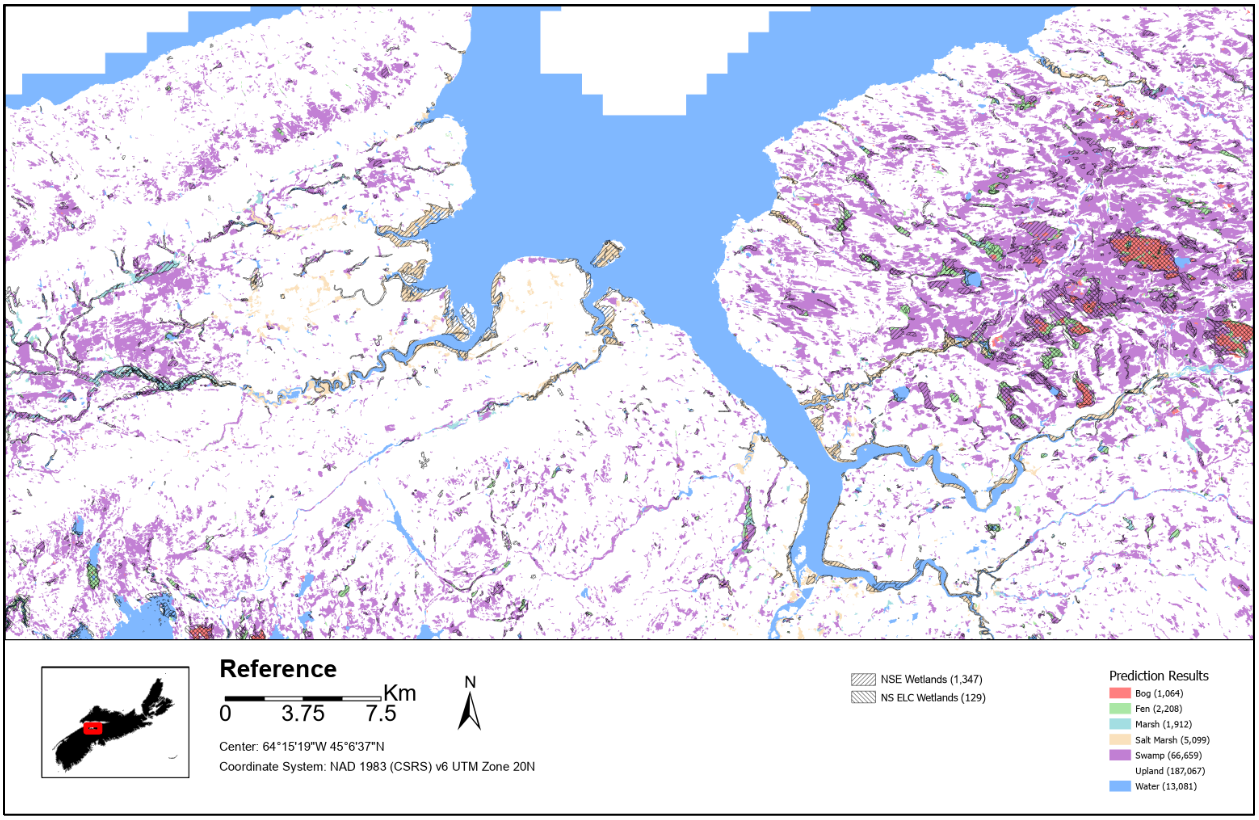



*Figure 26 shows the prediction results around the Kentville/Minas Basin area. Locally we see a greater abundance of salt marshes predicted with is in line with the attributes of the Nova Scotia wetland inventory (NSE). We see some agricultural land being depicted as slat marsh as well which may have implications for agricultural monitoring though further investigation as such is outside of the scope of this report. Note that while the delineation of fens (green) and bogs (red) do conform well to the NSE wetlands, swamps present a systematic increase. This increase is generally tied to tree covered areas – though the accuracy of this in real world terms generally unknown.*

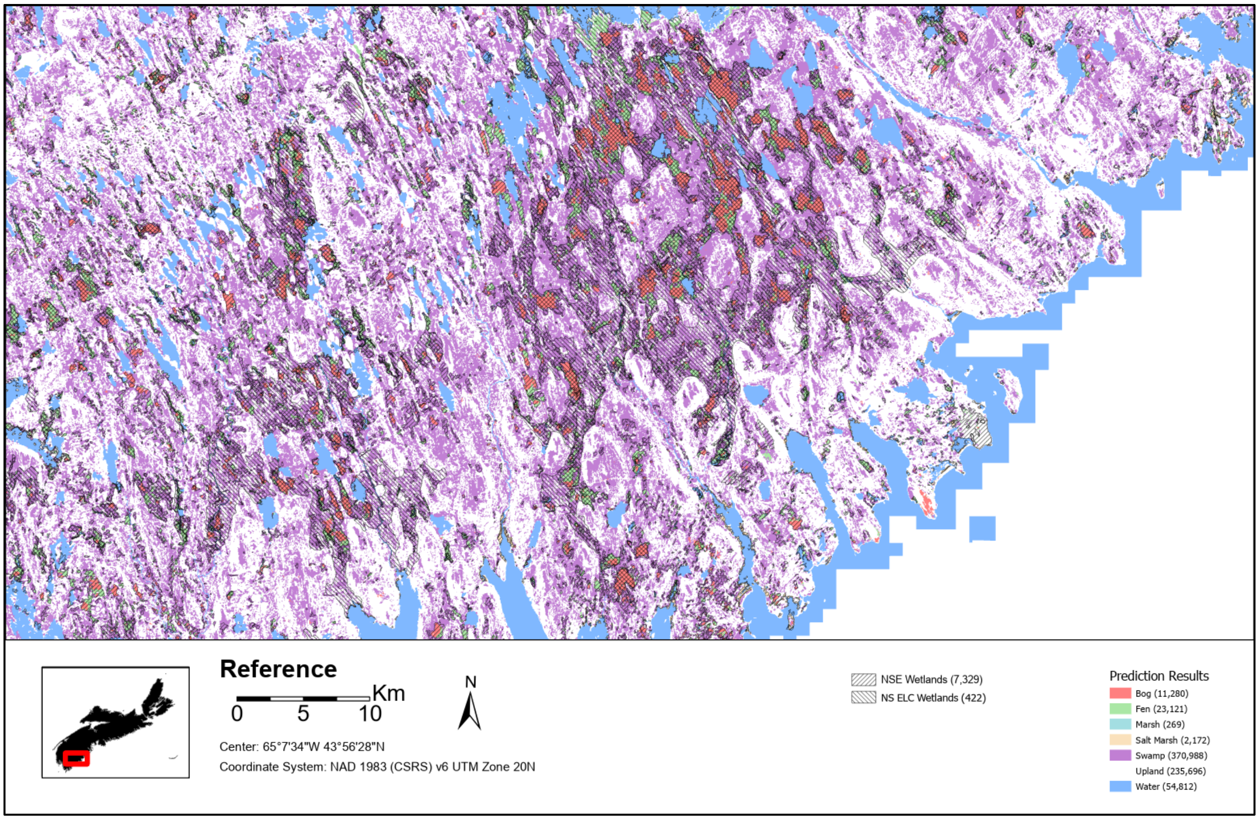



*Figure 27 shows the south of Nova Scotia, south lake Rossignol. This area of our results presents largest concentration of predicted wetlands (generally swamps) which deviate from the coverage presented in the existing Nova Scotia wetlands Inventory (NSE wetlands). This discrepancy is notably less substantial when considering the wetland coverage presented by the ecological land classification wetland ecosites (NS ELC wetlands). Further inspection details that this discrepancy maybe accounted for in part by previous works inability to detect features below the forest canopy. This overall result is yet not scrutinized by any field validation.*

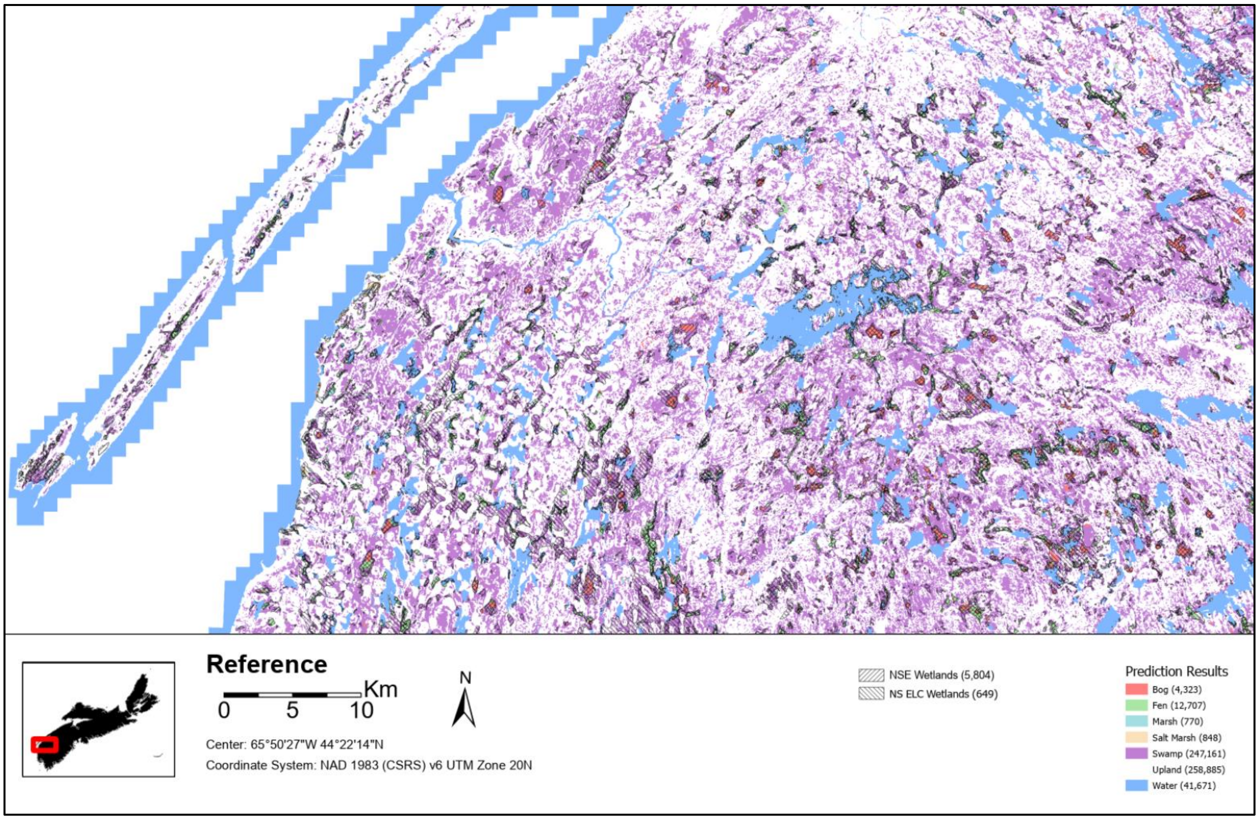



*Figure 28 shows a portion of southwestern Nova Scotia in the Weymouth area. Here we see a vast distribution of predicted swamps (purple) which are not included in the existing Nova Scotia wetland inventory or ecological land classification wetland polygons (hatched polygons). These swamps to conform to low laying features are glacial drumlins are clearly visible across the area showing notably less wetlands present.*

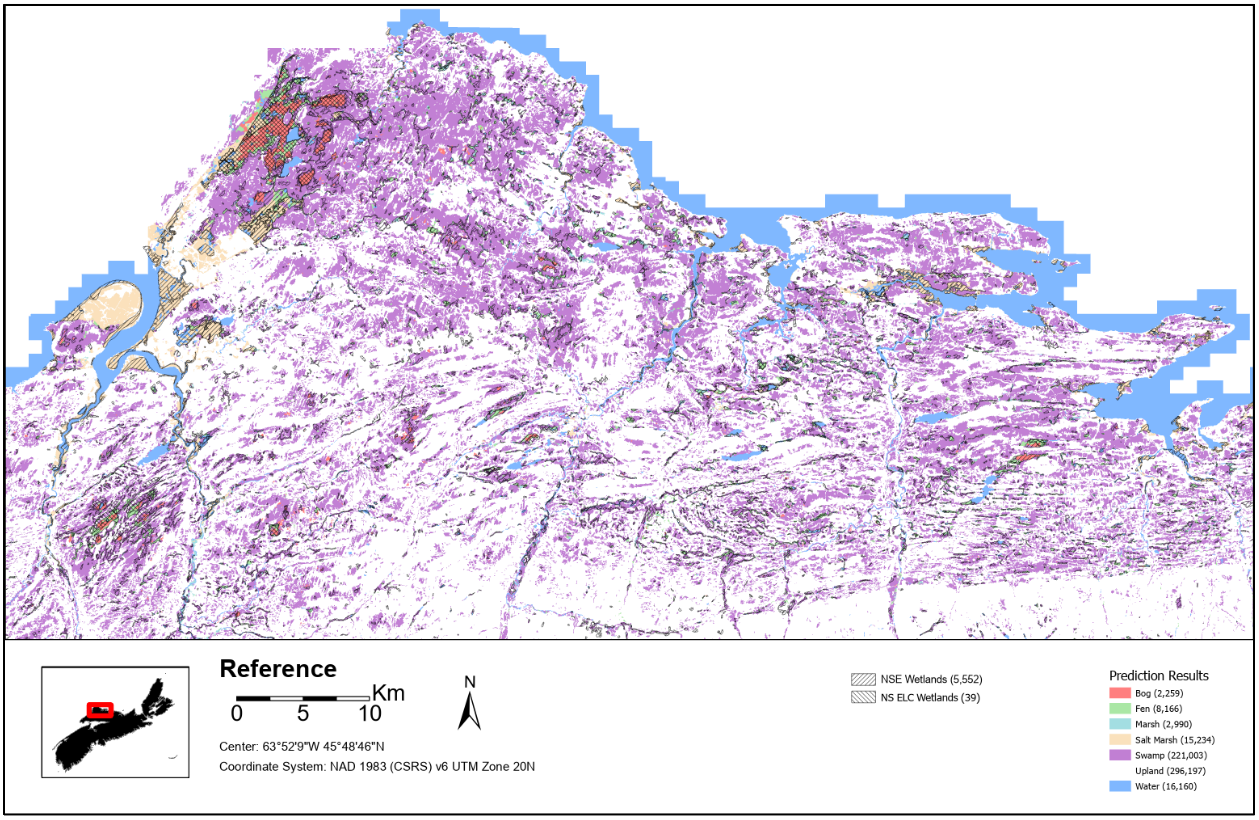



*Figure 29 shows the northern shore and isthmus area of Nova Scotia. We see a vast increase in the overall predicted presence of swamp type wetlands (purple) vs existing wetland maps (hatched lines). Note that our model predicts salt marshes in agricultural areas – though generally seems to perform across varied tide ranges. The prevalence of predicted swamp wetlands does decrease with geological and elevation change visible to the south of the map where the Cobequid mountains are visibly less dense with swamps.*

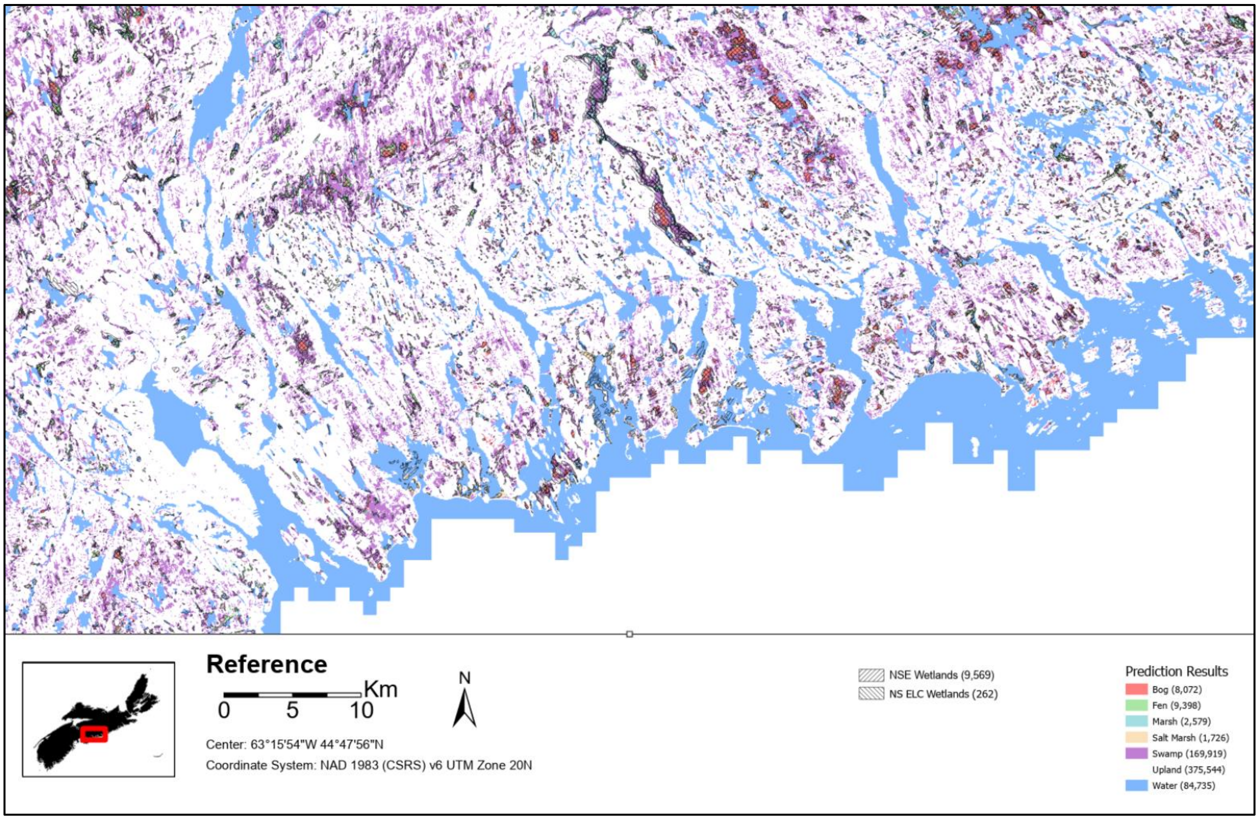



*Figure 30 shows the Halifax regional municipality and area. Here we can see some success in detecting key wetland areas bounding various rivers and lakes, whereas urbanized locations such as the Halifax peninsula shows little existing wetland areas. Our prediction suggests much of the region includes small low laying swamp type wetlands which may not presently be present in the Nova Scotia wetlands inventory. These results have not been independently verified.*

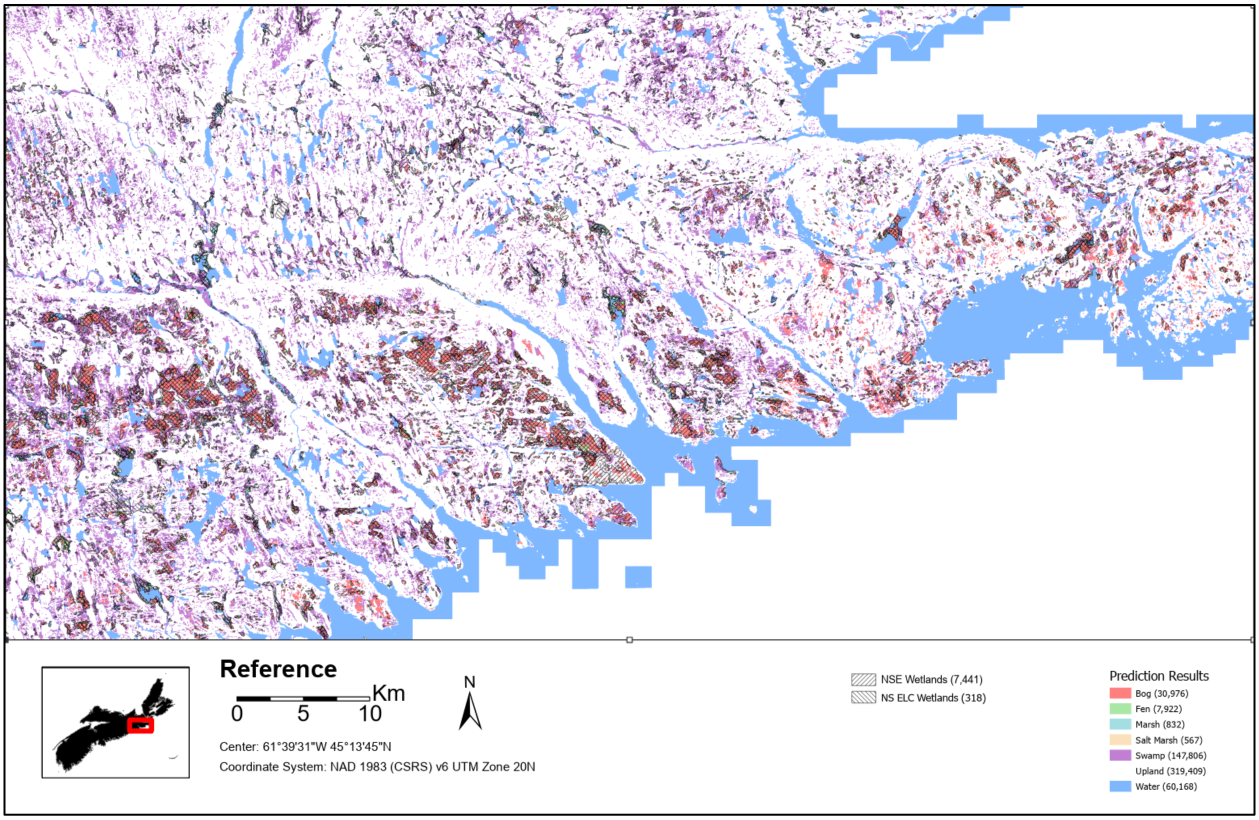



*Figure 31 shows the eastern shore area of Nova Scotia. As with other area, our model agrees strongly with bog type wetlands when considering the Nova Scotia wetland inventory of the ecological land classification wetland polygons. We systematically do predict a wider distribution of swamp type wetlands (purple) in low laying areas.*

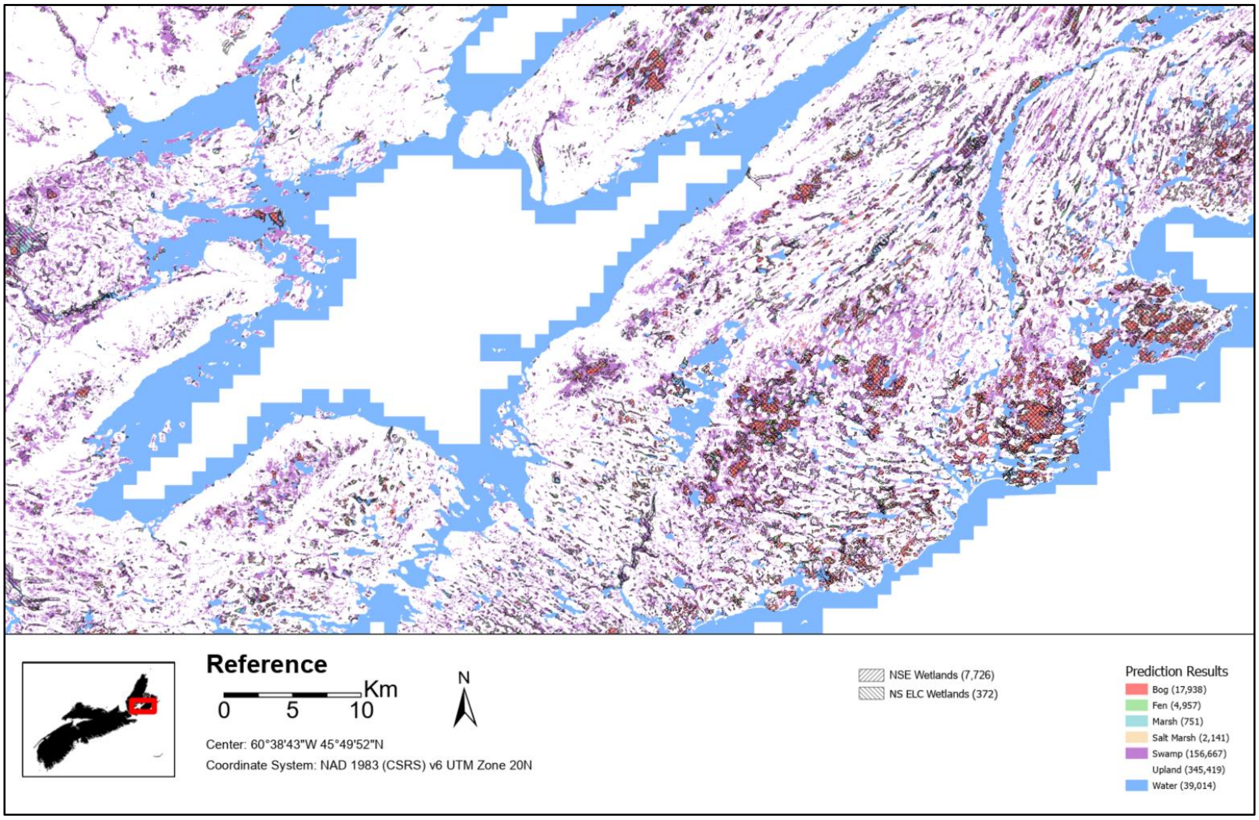



*Figure 32 shows the southern portion of Cape Breton. Comparing our results to the existing wetland inventory and ecological land classification wetland polygons, we see generally a strong agreement with bog type wetlands (red). As with much of our results, a generally greater coverage of swamp type wetlands has been predicted (purple).*

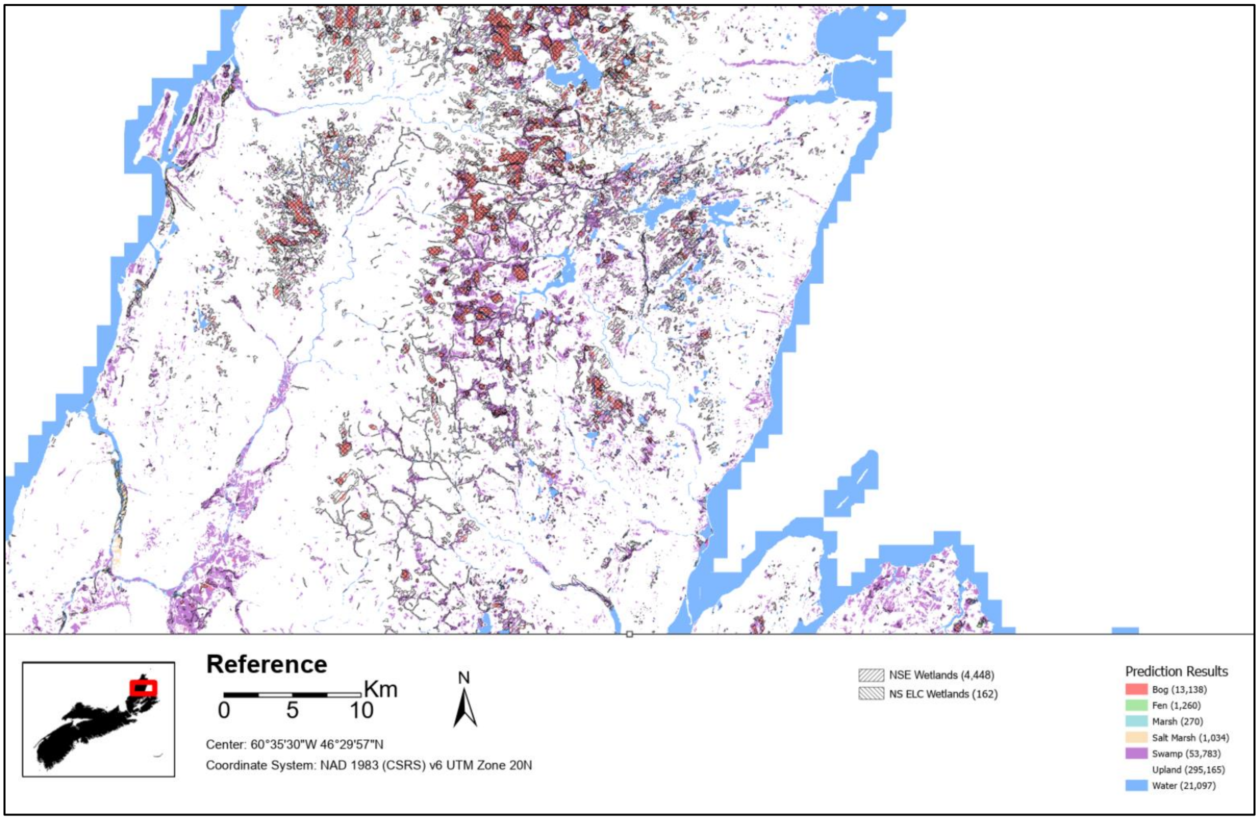



*Figure 33 shows the Cape Breton highlands and region. Here we see that our model does predict significant areas of new swamps (purple) generally bound to river flood plane areas. In the highlands themselves however, we do see systematic underprediction of bogs. This discrepancy appears unique to this region based on inspecting our results across the province.*

On visual analysis, segmentation using the "Chrome", NDVI, and topographic wetness index (TWI) indicates that the segmentation is performing better than similar attempts. The Chrome layer specifically does a reasonably good job of segmenting wetlands from non-wetlands, whereas the NDVI and TWI assist in differentiating subtle changes within a given wetland, allowing the classifier to assess each segmented section individually.

In terms of the internal validation of the model, based on the 10% withheld training data, the model performs highly accurately. With a weighted average total accuracy of 94%.

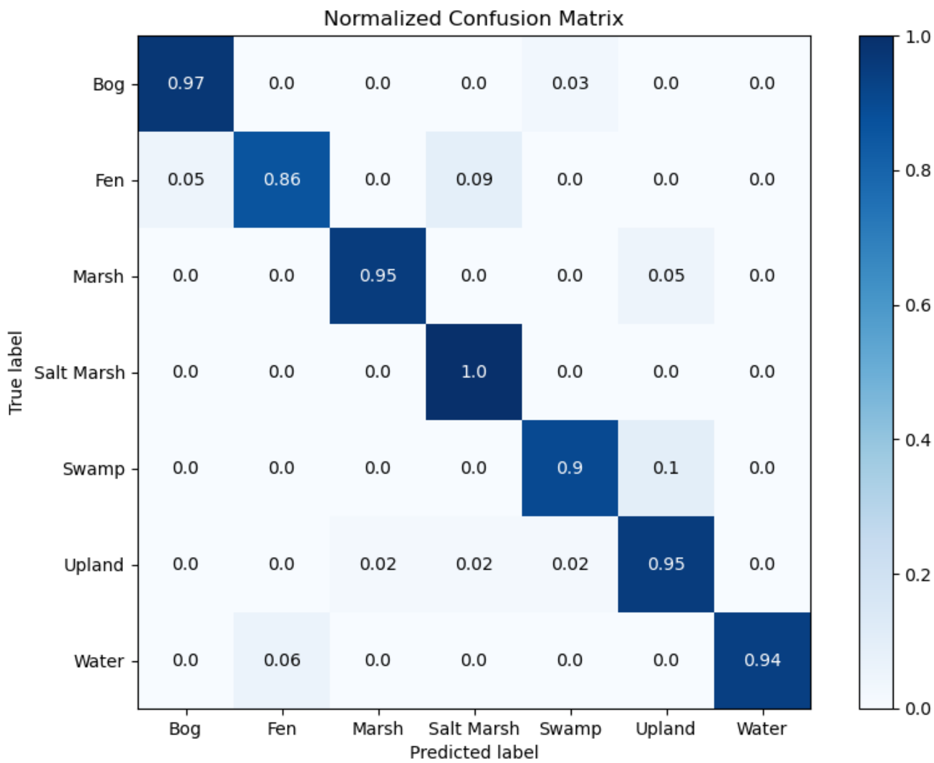



*Figure 34 shows the normalized confusion matrix of the wetland prediction model trained with ArcGIS AutoML on 2,041 training points with 10% withheld for validation.*

# 4 Discussion

This project provides a detailed methodology for significant improvement for identifying wetland areas in Nova Scotia compared to the existing Nova Scotia Wetland Inventory. A similar methodology that was tested using only Google Earth Engine (GEE) by NSECC also indicated significant improvements in wetland identification based on expert opinion at NSECC. That being said – success in both frameworks (GEE and ArcGIS, used here) are contingent on good feature layers and well-placed training data as well as validation including perhaps new ground truthing.

In either case, accurate wetland mapping in Nova Scotia is founded upon use of the high-resolution lidar inventory coupled with best available satellite data. With respect to lidar, we are lucky to have a remarkable province-wide lidar inventory of a very high quality. This enables much of the analysis to be conducted at minimal expense. Based on experimentation and previous studies, lidar plays a critical role in accurate wetland mapping. It may even be reasonable to use lidar only methods for wetland detection to compare results against methods such as the approach described in this project, which incorporates satellite information. In the satellite case, we utilize the best available data from Sentinel-2 provided by Google Earth Engine (GEE), which make use of their cutting-edge cloud filtering methods. While the European Space Agency produced Sentinel-2 produces freely, value added delivery methods, such as the seasonal composites and cloud masking, make proprietary systems such as GEE a valuable option for accessing such data. Though these analyses are typically free for research and educational use, there is a cost when deploying these systems for other uses and these costs should be considered when further exploring the methods described in this report.

It should be clearly noted that while this workflow been otherwise completed in ArcGIS, with satellite data provided from GEE – there are a series of alternative avenue to complete such analysis which may be considered including the free python tools available with scikit-learn for machine learning tools, and raster or vector processing with rasterio and GDAL. Many of the machine learning techniques utilized by ArcGIS in the AutoML tool, such as XGBOOST, are available as free and readily deployable tools though python libraries directly.

## 4.1 Interpretation of Results

Visual inspection of the output indicates that the methodology generally performs well at identifying wetlands in the province. Specifically, we observe better than expected predictions of saltmarshes along the coast. In earlier iterations of the model, the water class had the lowest reported single metric for accuracy in validation (0.76% recall) while it has visually performed very well – that is to say that water classifications conform well to lake and coast boundaries. This is to be expected with the inclusion of several high quality NDVI raster values averaged across the segments. In total, this should present a very strong signal to represent water per segment. Water however, when sampled as points for the model training, may validate less capably due to local variations such as glint and other artifacts. This should be a consideration for future training of similar models.

On the other hand, the observed deviation between the validation accuracy on points (94%) and variation of predicted confidence of segments (85.8 %, +/- 16%) can likely be attributed to an increased variance across the total set of segments. That is to say, unlike water polygons, other land segments shows more variation within them. This lower prediction confidence, if the model truly is performing well based on training points, could suggest that the segments may benefit from further segmentation or refinement. Perhaps, for example, a workflow could be developed to divide low prediction confidence segments into sub-segments programmatically.

Now that this functional workflow has been established, further analysis would likely include providing additional training data and refining the segmentation techniques based on more expert validation and inspection. It is particularly concerning given that key categories with relatively low representation (Bogs, Marshes, and Salt Marshes) show relatively low proportion of prediction confidence, suggesting that the segmentation used specifically may not adequately differentiate between neighboring wetland categories. Alternative segmentation techniques can be found in the scikit-image python library for example.

This segmentation issue can also be addressed by re-assigning the function of the ***Predict Using AutoML*** tool from predicting features (segments) to a raster output. This functionality was not successful during this project but is a sensible future direction, though it would also require all explanatory input to be in a raster format. This methodology may come at the expense of spatial resolution as some input data (REIP, NARI) are 20 m native resolution

| Eco district | Bog | Fen | Marsh | Salt Marsh | Swamp | Upland | Water | Predicted Wetland | Reported Wetland | Delta |
|---|---|---|---|---|---|---|---|---|---|---|
| Average | 1.3% | 1.0% | 0.0% | 1.0% | 19.5% | 67.4% | 8.1% | 23.0% | 7.3% | 15.7% |
| 100 | 10.7% | 0.8% | 0.0% | 0.0% | 9.0% | 76.0% | 3.5% | 20.5% | 33.1% | -12.6% |
| 210 | 2.1% | 0.2% | 0.0% | 0.0% | 9.2% | 86.8% | 1.8% | 11.4% | 12.6% | -1.2% |
| 220 | 0.1% | 0.1% | 0.0% | 0.7% | 10.6% | 72.2% | 16.3% | 11.5% | 3.3% | 8.2% |
| 310 | 0.3% | 0.1% | 0.0% | 0.0% | 6.4% | 92.4% | 0.7% | 6.9% | 1.9% | 5.0% |
| 320 | 0.4% | 0.5% | 0.4% | 0.9% | 20.5% | 62.6% | 14.7% | 22.7% | 6.4% | 16.3% |
| 330 | 0.1% | 0.2% | 0.0% | 0.0% | 13.5% | 85.5% | 0.7% | 13.9% | 2.8% | 11.1% |
| 340 | 0.1% | 0.2% | 0.0% | 0.0% | 10.5% | 88.5% | 0.7% | 10.8% | 1.9% | 8.9% |
| 350 | 0.1% | 0.1% | 0.1% | 0.0% | 11.7% | 87.4% | 0.6% | 12.0% | 1.7% | 10.3% |
| 360 | 0.9% | 0.8% | 0.0% | 0.0% | 21.6% | 71.5% | 5.1% | 23.3% | 5.9% | 17.4% |
| 370 | 1.0% | 0.8% | 0.2% | 0.0% | 22.3% | 71.0% | 4.6% | 24.4% | 7.8% | 16.6% |
| 380 | 0.8% | 0.4% | 0.0% | 0.0% | 18.1% | 79.8% | 1.0% | 19.3% | 4.3% | 15.0% |
| 410 | 0.2% | 0.2% | 0.1% | 0.0% | 12.7% | 86.4% | 0.5% | 13.1% | 1.1% | 12.0% |
| 430 | 0.4% | 0.6% | 0.0% | 0.0% | 10.1% | 77.4% | 11.4% | 11.2% | 5.0% | 6.2% |
| 440 | 2.7% | 1.3% | 0.1% | 0.0% | 21.4% | 66.3% | 8.1% | 25.6% | 8.7% | 16.9% |
| 450 | 2.1% | 0.9% | 0.0% | 0.0% | 23.0% | 67.8% | 6.1% | 26.1% | 7.4% | 18.7% |
| 510 | 1.4% | 0.7% | 0.2% | 0.4% | 23.8% | 64.4% | 9.1% | 26.5% | 8.8% | 17.7% |
| 520 | 0.1% | 0.3% | 0.4% | 0.6% | 16.4% | 80.0% | 2.2% | 17.8% | 3.0% | 14.8% |
| 530 | 0.4% | 1.0% | 0.6% | 1.0% | 33.8% | 59.1% | 4.2% | 36.7% | 5.5% | 31.2% |
| 540 | 0.2% | 0.5% | 0.4% | 0.0% | 18.9% | 79.4% | 0.6% | 20.0% | 3.3% | 16.7% |
| 550 | 4.3% | 3.6% | 0.4% | 22.2% | 20.5% | 38.0% | 11.0% | 51.0% | 27.9% | 23.1% |
| 560 | 1.1% | 1.2% | 0.2% | 0.2% | 31.6% | 64.6% | 1.1% | 34.3% | 8.4% | 25.9% |
| 610 | 0.1% | 0.6% | 1.4% | 1.7% | 21.5% | 72.3% | 2.4% | 25.3% | 4.8% | 20.5% |
| 620 | 0.1% | 0.7% | 1.2% | 1.6% | 21.5% | 60.7% | 14.1% | 25.1% | 7.3% | 17.8% |
| 630 | 0.6% | 1.1% | 1.3% | 0.4% | 25.4% | 69.2% | 2.0% | 28.9% | 6.2% | 22.7% |
| 710 | 0.0% | 0.4% | 0.0% | 0.1% | 11.9% | 86.9% | 0.6% | 12.5% | 1.4% | 11.1% |
| 720 | 0.9% | 1.5% | 0.0% | 0.0% | 28.4% | 61.4% | 7.8% | 30.8% | 6.1% | 24.7% |
| 730 | 0.6% | 1.9% | 0.3% | 0.3% | 32.6% | 53.2% | 11.1% | 35.7% | 5.6% | 30.1% |
| 740 | 0.4% | 1.3% | 0.1% | 0.0% | 28.9% | 59.1% | 10.2% | 30.8% | 4.5% | 26.3% |
| 750 | 0.5% | 1.3% | 0.0% | 0.0% | 16.2% | 17.1% | 7.8% | 18.0% | 6.1% | 11.9% |
| 760 | 3.3% | 4.5% | 0.0% | 0.0% | 46.7% | 39.5% | 6.0% | 54.5% | 14.7% | 39.8% |
| 770 | 1.5% | 2.9% | 0.0% | 0.0% | 40.4% | 49.8% | 5.4% | 44.8% | 8.6% | 36.2% |
| 780 | 0.4% | 1.0% | 0.0% | 0.0% | 18.2% | 72.5% | 7.8% | 19.7% | 4.1% | 15.6% |
| 810 | 4.3% | 0.8% | 0.1% | 0.4% | 19.4% | 63.9% | 11.1% | 25.0% | 14.6% | 10.4% |
| 820 | 4.6% | 0.8% | 0.0% | 0.4% | 16.0% | 70.3% | 7.8% | 21.9% | 9.4% | 12.5% |
| 830 | 2.5% | 3.3% | 0.1% | 1.4% | 34.6% | 53.5% | 4.6% | 41.9% | 15.9% | 26.0% |
| 840 | 0.8% | 2.3% | 0.3% | 6.2% | 32.7% | 48.5% | 9.2% | 42.3% | 16.3% | 26.0% |
| 850 | 0.0% | 0.2% | 0.0% | 0.6% | 0.2% | 86.6% | 12.4% | 1.0% | 0.0% | 1.0% |
| 910 | 0.5% | 0.4% | 0.5% | 0.6% | 9.5% | 87.9% | 0.7% | 11.4% | 3.6% | 7.8% |
| 920 | 0.0% | 0.4% | 0.1% | 0.1% | 10.8% | 86.7% | 2.0% | 11.3% | 1.6% | 9.7% |
| NONE | 0.0% | 0.0% | 0.0% | 0.4% | 0.1% | 1.3% | 98.1% | 0.6% | 0.0% | 0.6% |

*The above table indicates the total area percentage of each wetland designation predicted based on our methods broken down per Ecodistrict as determined by the NS ELC (see figure below). Each total wetland areas is further compared with the wetland percent reported by the ELC (Neily et al. 2017)*

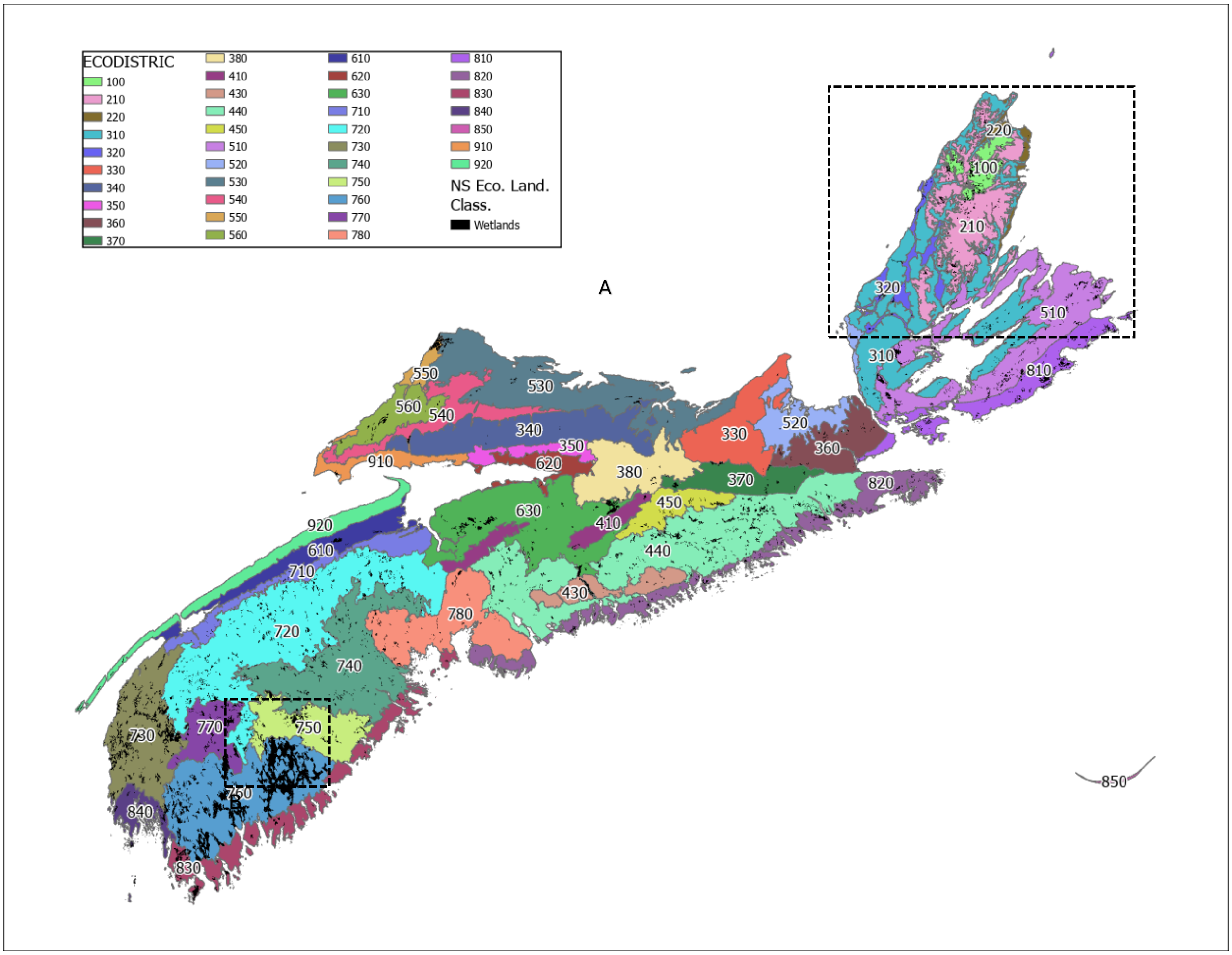



Figure 35 shows the numerical Ecodistricts as depicted in the Nova Scotia ecological land classification (ELC). Also indicated are the wetland areas designated in that analysis (Neily et al. 2017).

A1 A2 A3

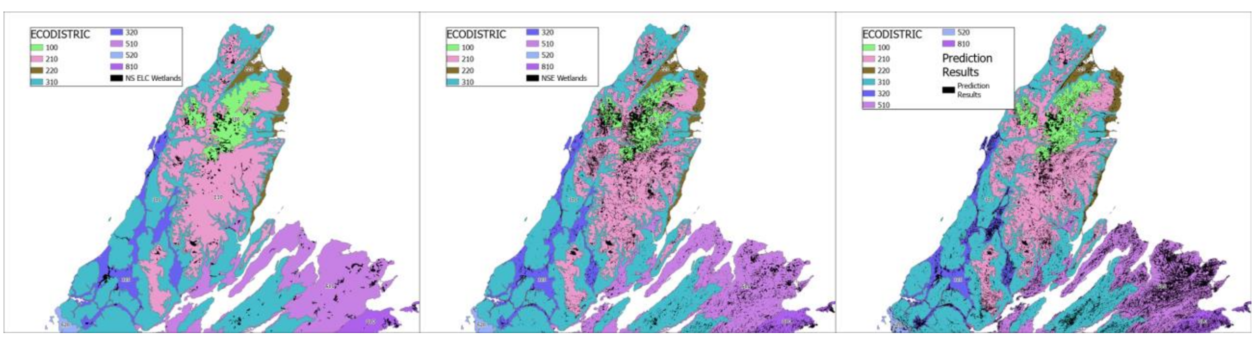



Figure 36 highlights the deviation between, ***A1***; the coarse wetland mapping product presents in the NS ELC (Neily and 2017), ***A2***; all wetland types as depicted by the by existing wetland inventory, and ***A3***; Our results. As shown in the above table, the cape Breton highland is the only Ecodistrict where we detect an overall lower abundance of wetlands. This is likely due to the unique characteristics of the highlands which are not adequately captured in the model. Perhaps additional training or values may help correct this perceived discrepancy.

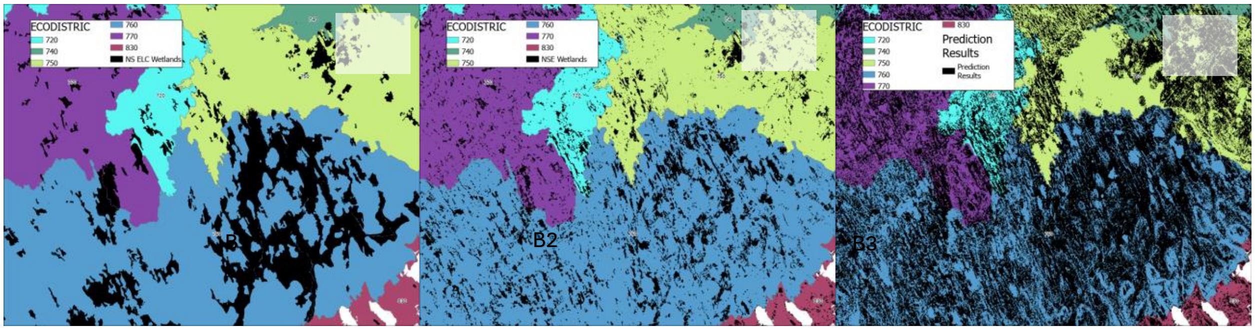



*Figure 37 shows Ecodistrict 760 (Sable), directly to the south of Lake Rossignol, NS. This area presents the largest systematic discrepancy in terms of our greater abundance of swamp type wetlands. Considering all wetlands, we see that the wetland extents presented in the NS ELC (**B1**) are in support of our results (**B3**) in that there are similarly more overall wetlands depicted than currently in the Nova Scotia wetland Inventory (**B2**).*

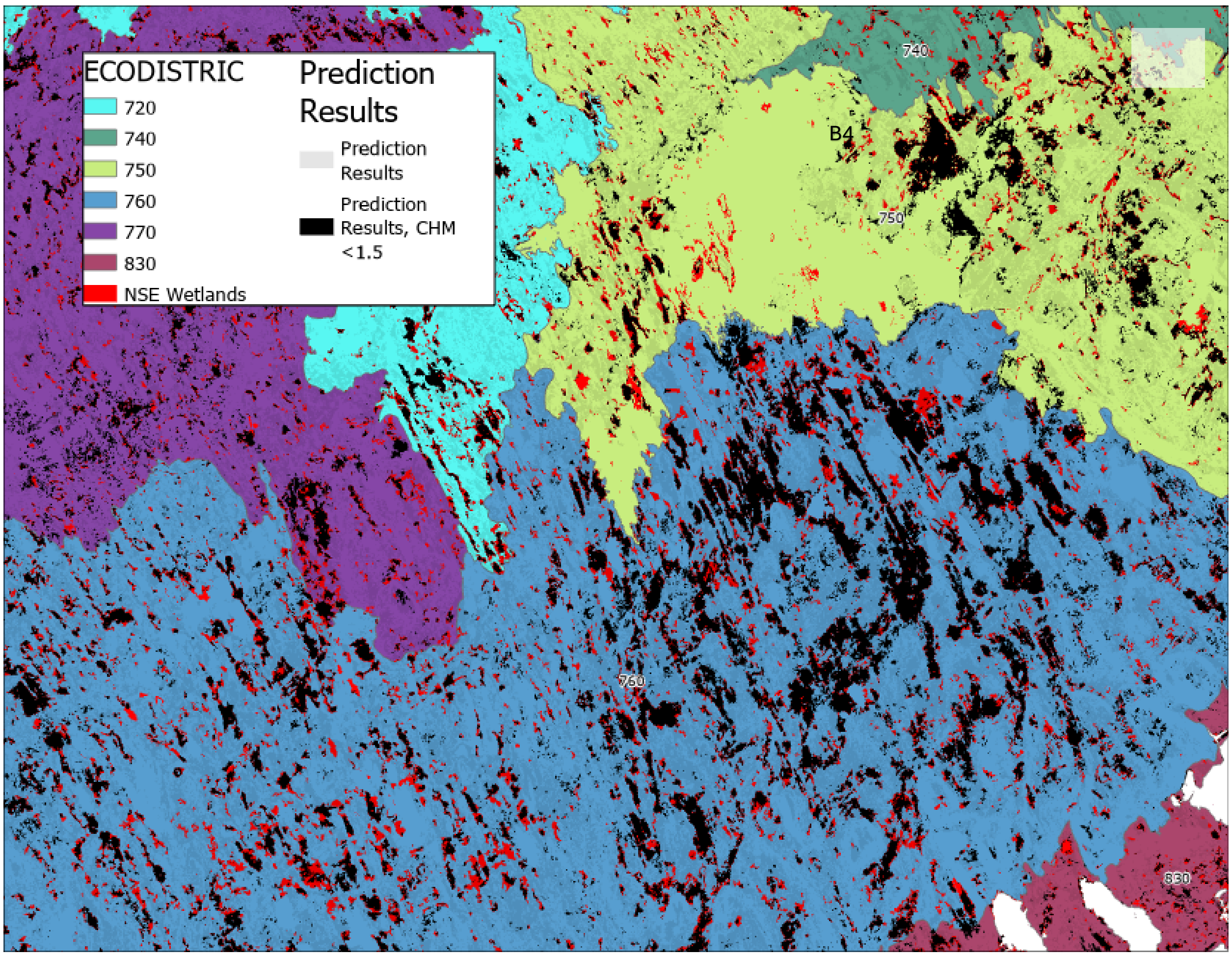



*Figure 38 shows the same area Ecodistrict 760 (Sable) south of Lake Rossignol. Here our results are filtered to segments where the mean canopy height value (CHM) is less than 1.5 meters. This indicates wetlands which do not have significant vegetation. Note that with this restriction, our then results (black) are more in line with what is presented by the current Nova Scotia wetland Inventory in terms of all wetland types (red).*

### 4.1.1 Implications of Forest Canopy

By restricting our predicted results to segments where the canopy height average is less than some threshold, we can see our results are more in line with the existing Nova Scotia wetland inventory. This is indicated in the figure above.

*Table 9 shows area and percent area of different wetland types using different CHM thresholds.*

| | *All segments* | | *Segments where CHM < 1.5 m* | | *Segments where CHM < 4.0 m* | |
|---|---|---|---|---|---|---|
| *type* | Area (Sq. Km) | Area (%) | Area (Sq. Km) | Area (%) | Area (Sq. Km) | Area (%) |
| *Bog* | 723.94 | 1.16% | 622.40 | 1.00% | 718.32 | 1.15% |
| *Fen* | 666.87 | 1.07% | 445.54 | 0.72% | 617.92 | 0.99% |
| *Marsh* | 114.24 | 0.18% | 84.34 | 0.14% | 108.78 | 0.17% |
| *Salt Marsh* | 231.86 | 0.37% | 208.50 | 0.34% | 227.15 | 0.37% |
| *Swamp* | 12,767.99 | 20.52% | 261.51 | 0.42% | 2,784.90 | 4.48% |
| *Upland* | 38,156.96 | 61.32% | **total** | **2.61%** | **total** | **7.16%** |
| *Water* | 9,566.54 | 15.37% | | | | |
| ***Total*** | **62,228.40** | | | | | |

As indicated by the table above, when restricting our predicted wetlands to segments where the canopy height (average) is less than 4.0m, we see an overall percentage land cover of wetlands of 7.16% whereas the current Nova Scotia wetland inventory indicates 6.8% wetland cover.

## 4.2 Comparison of ArcGIS tools

Several inconsistencies and issues in the ArcGIS workflow resulted in many stumbling blocks throughout this project.

After using the Train Using AutoML tool on a suite of explanatory rasters, one would expect to then generate an output prediction raster using the accompanying Predict Using AutoML tool directly on the same set of rasters. However, at the time of writing this report, this workflow was not well documented and did not appear to function as expected. It was noted, however, that running the Predict Using AutoML tool using a set of features attributed with the same set of explanatory rasters would work as expected.

When issues were encountered with the standard ArcGIS workflow to predict raster outputs from the trained model, an alternative method was discovered which ultimately resulted in a more robust workflow. Using segmentation from a separate set of rasters to generate features, which were attributed with the full list of explanatory rasters, has the advantage of decoupling the polygonal boundary of the wetlands from the full list of feature rasters. In effect, this means that the resolution of the resulting wetland layer can be determined by high resolution layers such as the lidar elevation model (1m) while still incorporating low resolution data (such as REIP, derived from 20m sentinel-2 data, as mentioned).

Additionally, significant time was allotted to debugging alternative random forest tools in ArcGIS. Specifically, a significant amount of time was spent attempting to get the ***Forest Based Classification and Regression*** tool (FBCR) to report reasonable validation results. Despite best efforts, and with the same training data as utilized for the AutoML results of this report, an accuracy of only ~42% was observed. This is despite no small effort to normalize the feature layers, training distribution, and even to exclude classes. To test the tools results, it was compared against the industry standard python forest-based classifier XGBOOST in a pure python environment. When this tool successfully yielded the expected ~92% accuracy, the FBCR tool was determined to be the issue and development focused on alternatives, ultimately settling on the AutoML approach.

Below illustrates the variance in achieved accuracy/validation using ArcGIS FBCR and python XGBOOST on the same training data:

*Table 10 Python 3.10.12 xgBoost version 2.0.3 class codes listed as ID numbers (but perform generally well)*

| *Category* | *F1-Score* | *MCC* | *Sensitivity* | *Accuracy* |
|---|---|---|---|---|
| *Bog* | 0.89 | 0.87 | 0.93 | 0.97 |
| *Fen* | 0.93 | 0.91 | 0.96 | 0.97 |
| *Marsh* | 0.85 | 0.83 | 0.78 | 0.97 |
| *Salt Marsh* | 0.96 | 0.96 | 0.95 | 0.99 |
| *Swamp* | 0.93 | 0.91 | 0.94 | 0.96 |
| *Upland* | 0.91 | 0.90 | 0.83 | 0.98 |
| *Water* | 0.96 | 0.96 | 1.00 | 0.99 |

*Table 11 ArcGIS v 3.1.1, Forest Based Classification & Regression Tool, very poor performance across multiple classes.*

| *Category* | *F1-Score* | *MCC* | *Sensitivity* | *Accuracy* |
|---|---|---|---|---|
| *Bog* | 0.29 | 0.36 | 0.18 | 0.86 |
| ***Fen*** | **0.29** | **0.25** | **0.94** | **0.51** |
| *Marsh* | 0.11 | 0.14 | 0.06 | 0.90 |
| *Salt Marsh* | 0.62 | 0.61 | 0.50 | 0.95 |
| *Swamp* | 0.41 | 0.43 | 0.27 | 0.88 |
| *Upland* | 0.71 | 0.67 | 0.57 | 0.86 |
| *Water* | 0.76 | 0.74 | 0.75 | 0.96 |

While the ArcGIS processing issues described above do continue to be major stumbling blocks, it should be reiterated that this work was conducted using ArcGIS Pro version 3.1.1, which may now be updated to version 3.3. There is, however, no mention of specific fixes to the relevant tools found mentioned in the release notes for either version 3.3 or 3.2. Interestingly, new features added to the Forest Based Classification and regression suggesting this tool has continued support despite its poor classification performance noted above. Perhaps critically though, ArcGIS version 3.3 does include an update from python version 3.9.18 (to 3.11.8) which may have substantial impact on the performance and stability of various arcpy tools. Additionally, version 3.3 includes modifications to the geoprocessing framework including new techniques for data caching. Ultimately, a workflow wherein the input rasters may be used in the training and prediction steps reliably and in memory would facilitate a much simpler and lightweight approach. Ideally this updated approach may include a viable option to predict directly to a raster using the Predict Using AutoML tool. This would allow for more flexibility considering complications related to raster segments which we are relying on currently.

## 4.3 Feature Layers

While not adequately covered in this report, a typical benefit of most forest-based classifiers is the inherent quantization of the importance of input feature layers. In general, this is good practice to report which layers had the highest utility in the model such that others can prioritize similar efforts accordingly.

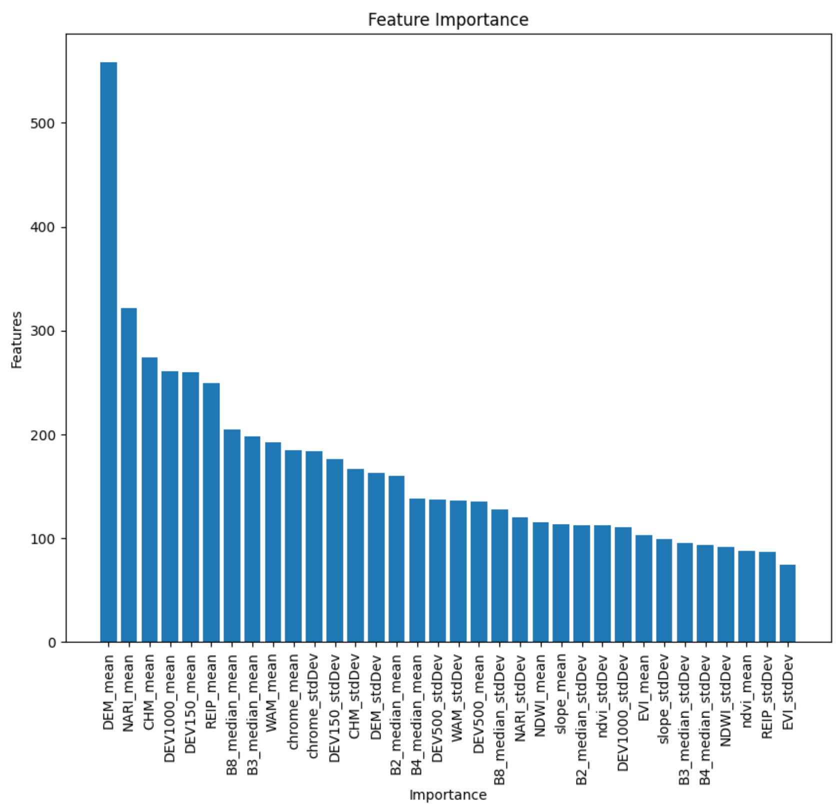



*Figure 39 Python 3.10.12 xgBoost version 2.0.3 feature importance.*

As an aspect of the ensemble model process of AutoML, the final prioritisation of all input feature layers has not yet been discovered, though earlier iterations of the process (i.e., when predicting a 'basic' level model) do provide information on relative feature importance. For the purposes of demonstrating the approximate level of importance for features used in this report, we have included a figure demonstrating the relative importance computed with XGBOOST. Note that the highest performing layers include the DEM, NARI, CHM, and DEV1000(m), and REIP.

### 4.3.1 Relative variable importance

The AutoML tool in ArcGIS has a particular limitation with regard to reporting the relative variable importance, specifically when utilizing the ADVANCED training mode. As such, it is yet possible to ascertain the relative variable importance only when performing the BASIC training. While this does not exactly relate to the specific model and results we achieved using the ADVANCED mode, this does provide good insight into the general utility of each input layer. From conducting a BASIC model training with the same training and input data as our ADVANCED model, we can infer that the DEM was the most relevant input layer and was specifically important in detecting marsh and salt marsh type features. Second, the spring near infrared layer was identified as most valuable for differentiating water features. The Anthocyanin sensitive index (NARI) proved useful across all classes, specifically uplands and fens. Various lidar metrics including our novel CHROME index, the provided depth to water layer (DTW), and the computed topographic water index (TWI), appeared most useful in detecting uplands, bogs, and fens respectively – although all were generally highly useful. Notably, NDVI and canopy height (CHM) performed well in determining swamp features. Additional input variables provided varied and diminishing utility in the prediction. The SHAP score importance for the synonymous BASIC model of the top performing input layers is presented to the right:

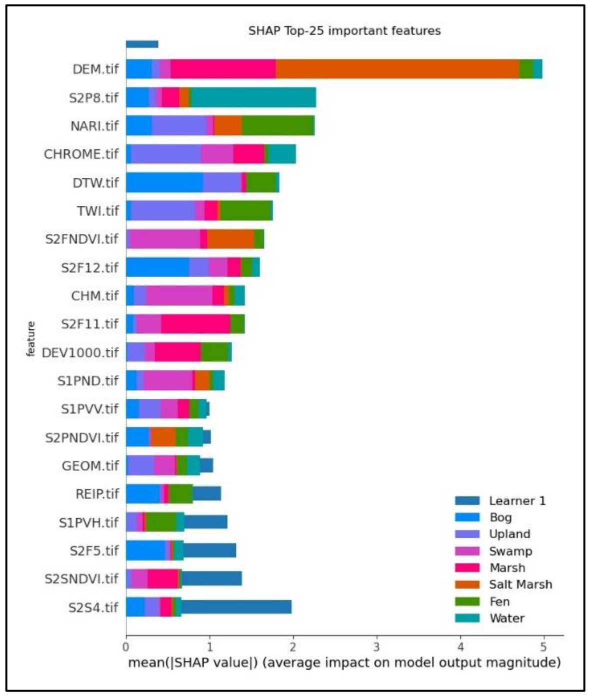



*Figure 40 shows the SHAP feature importance score for the AutoML BASIC model, which was run with the same inputs as the ADVANCED model used to generate out results.*

## 4.4 Implications for Flood Mitigation and Ecosystem Services

This report provides a framework for a method to generate wetland classification polygons using ArcGIS. The results showed that the current Nova Scotia wetland inventory is underrepresenting both the number and total area of wetlands throughout Nova Scotia significantly, especially swamps. This finding presents both an opportunity and a challenge, as wetlands have an important role in reducing the risk of climate hazards, such as flooding and ecosystem impacts, but also require special management efforts.

# 5 Recommendations

## 5.1 Strategic Recommendations for Wetland Management

This report has been conducted with technical advice/assistance provided by NSECC. Remote sensing-based methods rely heavily on the high-resolution lidar data provided by the province as well as satellite data, including Sentinel-2. A provincial priority should be placed on continuing and initiating acquisition of such data to monitor changes into the future including both lidar, aerial photography, and high-resolution satellite data. None such high-quality analysis can be performed regardless of platform or technique in the absence of adequate, up to date remote sensing data.

In terms of strategic directions, a valuable next step to any wetland predictive model would be to incorporate additional explanatory variables directly tied to climate, such as rainfall, heat stress, or sea-level. With the inclusion of these layers, predictive models can be adjusted to match future climatic scenarios and extrapolate possible effects to wetland distribution directly. The inventory and approach generated by this project may be useful for tracking change in wetland extent over time due to stressors such as urbanization and climate change.

Moreover, integrating land use data with predicted wetland distributions can identify potential restoration sites and improve model accuracy. Assessing the functional importance of wetlands, such as their roles in flood regulation near urban centers and carbon sequestration, will provide valuable insights for management. Incorporating these elements will enhance conservation strategies and align with broader environmental goals.

## 5.2 Suggestions for Methodology Refinement

This model as described would benefit from refinement to the segmentation methodology. In general, an overall comparison between this method based in ArcGIS and those conducted using Google Earth Engine could yield a single best approach which could be adopted by the province. Results from this or other similar modelled outputs depicting wetland distribution would benefit from a significant level of field validation. This could be conducted semi-efficiently with the aid of low altitude UAS/drone equipped with multi-spectral imaging sensors. The Applied Geomatics Research Group has such equipment available for use in future iterations of this project.

It is important to emphasize that this workflow is the result of various compromises due to limitations encountered during its development. For instance, segmentation of the input rasters continues to be a challenging component to the workflow. The segmentation technique as it is currently implemented certainly negatively impacts both the processing efficiency and quality of the results. While continued experimentation with regard to selecting and scaling various raster data to perform the segmentation is likely worthwhile, predicting wetland classes directly to a raster would have certain advantages. This unfortunately was never made possible due to apparent issues with the Predict Using AutoML tool during workflow development.

Moreover, the approach used to classify all wetland types alongside upland and water features does impact the interpretability of the model confidence. Specifically, for example, a 'low' confidence bog segment does not clearly indicate this model is struggling to differentiate this feature as a bog vs. another wetland type, of if it is not confident it is a wetland at all. This confusion could be resolved by performing the overall classification in stages whereby first wetlands are differentiated from non-wetlands and water, and then proceeding to distinguish individual wetland types. If in this method, for example, a presence-absence model such as Maximum Entropy were used to detect wetlands – this could constitute a mask to in turn restrict and improve various steps including band scaling and segmentation which may in turn improve the differentiation of wetland types further. An approach to facilitate this technique was experimented with utilizing a simple slope-based threshold to determine the scaling for segmentation rasters in attempt to better differentiate wetland types, but this was not integrated into the methods proposed here.

Additionally, further subdividing the analysis over the 39 recognized Ecodistricts of Nova Scotia (Neily 2003 and 2017) may yield better results, especially when discerning one wetland classification from another. This could be done either by separate rounds of classification, or perhaps simply by including the Ecodistrict as categorical variable in the decision tree classifier.

Ideally, and to make better use of an automated wetland classification, future improvements to these methods should attempt to include temporal and environmental attributes. This could be done, for example, by compiling Sentinel-2 data and comparing the impact across multiple model prediction runs. The resulting changes in wetland coverage could therefore be linked to a specific time frame. Similarity, by including in the model, specific environmental attributes perhaps relevant to wetland distribution, including temperature – model predictions could then suggest potential changes to wetland distributions in relation to various climate projections. This would be particularly relevant for salt marshes with respect to sea level rise, which would require we adjust the digital elevation model to tide for training, which could similarly be adjusted for prediction based on specific sea level rise scenarios.

Perhaps the most critical way to improve the results of this project would be to thoroughly investigate the proposed wetland distributions both in map form and in the field where necessary. The data generated from this report represents a significant challenge for comprehensive quality assurance and was not a major consideration of this project due to that scope. The results from this work do propose major modifications to the existing Nova Scotia wetland inventory, for example, overestimation of swamp areas in some cases. However, there may be wetland areas which are yet generally unmapped that this project has detected which may warrant further investigation.

In general, this report presents a somewhat robust technique for automatic province wide wetland classification, though most aspects of the technique used can be improved in some way or another. Generally, limiting the analysis to tools found in ArcGIS when possible and data available in Google Earth Engine may have negatively impacted the overall simplicity and efficiency. Many of the specific techniques used may benefit from exploring alternative options in future iterations. Importantly though, many of the challenges overcome here may have benefit to the province outside of wetland mapping. Mapping other important features, such as forest inventory, fire risk, impervious surface mapping and more may be more easily undertaken with this overall framework in place.

# 6 CONCLUSIONS

In this report we detail a concise methodology using ArcGIS tools (AutoML) to develop an updated wetland inventory for the Province of Nova Scotia. Though the focus of this work should be the method as a framework and not the results – we do present a new province wide wetland inventory with a reported accuracy of ~94%. This model is based on 2041 training points (where 10% is validation) and predict wetlands using the Canadian Wetland Classification System (marshes, swamps, bogs, fens, uplands, and water – shallow water not included). Predictions were computed based on 38 explanatory variables derived from lidar and satellite derived indexes across 10 million segments spanning the province of Nova Scotia with a functional resolution of approximately 10m. The overall confidence of the predictions is variable (85.8% +/- 16.0%), possibly due to the coarse nature of the geographic segments used. The classification revealed 72% of the area as 'Upland' (non-wetland), and 27.5% as wetland. We remove inland water from this percentage as our methods do not differentiate ocean from lake water. Among wetlands, swamps were predominant, accounting for 88.0% of total wetland area (24.3 % of the province land), with bogs accounting for 5% of wetlands (1.4 % of the province), fens 4.6% of wetlands (1.3% of province), salt marshes 1.6% (0.4% of the province), and marshes 0.8% of wetlands (0.2% of the province).

By area, the existing provincial wetland inventory indicates a total bog and fen area of 1600 sq. km, while we detect 1390 sq. km. (a factor of 0.87). Similarly, the inventory contains 430 sq. km. of combined salt marshes and marshes, while we detect 346 sq. km. (a factor of 0.81). While these other wetland results underpredict current published values slightly, our predicted swamp area shows an increase by a factor of 7.3 (12,769 sq. km) compared to the published total area (1,761 sq. km.). Note that 215 sq. km. of wetland area in the provincial inventory are of an unknown wetland classification.

Our results do indicate a significant overall increase of wetland area when compared to the existing Nova Scotia Wetland Inventory (which is approximately 6.8%). Our results, however, when restricted by canopy height of areas less than <4.0 m, indicate an overall wetland coverage of approximately 7.2% which is much more in line with the published extents. This factor may indicate that our approach, vs. previous work, could be better detecting treed swamps, likely due to the use of lidar based indexes, though these findings are highly conditional to future ground validation.

## 6.1 SUMMARY OF KEY FINDINGS

ArcGIS tools are at variable states of functionality and documentation. The best forest-based classifier at the time of writing this report was determined to be AutoML. Unfortunately, this tool was not seemingly paired with a functional method to predict directly to raster output. A positive compromise was to predict attributed segments – which added a certain level of spatial flexibility to the analysis.

## 6.2 FINAL THOUGHTS

The provincial lidar was pivotal in the success of this project, including complex lidar derivatives such as the canopy height model. These data sets Nova Scotia apart from some other jurisdictions wherein we should perhaps become leaders in the adoption and utilization of these data. Additional earth observation tools such as Google Earth Engine, when paired with high precision lidar, can unlock an unprecedented ability to map and understand changes. Machine learning tools empower expert identification to be extrapolated across wide areas of high-resolution data. While Nova Scotia is a complex geography of highly variable ecological conditions - the power and detail offered by these technologies combined enable us to better understand, manage, and protect our diverse landscapes, showcasing a model for effective environmental stewardship.

# 7 ACKNOWLEDGMENTS

## 7.1 CONTRIBUTIONS

This work has been completed by research specialists with the Applied Geomatics Research Group (AGRG) with significant contributions from Wetland Specialists with the Nova Scotia Department of Environment and Climate Change.

## 7.2 SUPPORT

This report has been conducted by the Applied Geomatics Research Group, as part of Nova Scotia Community College Applied Research. This study is funded by the Province of Nova Scotia with additional in-kind support from GeoNOVA. We would like to thank the Province of Nova Scotia for their continued support for AGRG's research in the region.

## 7.3 REFERENCES

Aguilar, Fernando J., et al. "Modelling vertical error in LiDAR-derived digital elevation models." *ISPRS Journal of Photogrammetry and Remote Sensing* 65.1 (2010): 103-110.

Amani, Meisam, et al. "Canadian wetland inventory using Google Earth Engine: The first map and preliminary results." *Remote Sensing* 11.7 (2019): 842.

Ballanti, Laurel, et al. "Remote sensing for wetland mapping and historical change detection at the Nisqually River Delta." *Sustainability* 9.11 (2017): 1919.

Bayle, Arthur, et al. "Improved mapping of mountain shrublands using the sentinel-2 red-edge band." *Remote Sensing* 11.23 (2019): 2807.

Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32.

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.

Convention on Wetlands Secretariat. "The List of Wetlands of International Importance." 2024. https://www.ramsar.org/our-work/wetlands-international-importance/ramsar-list.

De Reu, Jeroen, et al. "Application of the topographic position index to heterogeneous landscapes." *Geomorphology* 186 (2013): 39-49.

Deering, Donald Wayne (1978). Rangeland reflectance characteristics measured by aircraft and spacecraft sensors. Texas A&M University. Texas A&M University. Libraries. Available electronically from https://hdl.handle.net/1969.1/DISSERTATIONS-253780.

DeLancey, Evan. "ABMI Wetland Inventory-Technical documentation ABMI Geospatial Centre March, 2021." (2021).

Nova Scotia Department of Environment and Climate Change (NSECC). “Wetlands”. Government of Nova Scotia. https://novascotia.ca/nse/wetland/

Environment and Climate Change Canada (2016). Canadian Environmental Sustainability Indicators: Extent of Canada's Wetlands. Consulted on March 14 2024. Available at: www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=69E2D25B-1.

Fluet-Chouinard, Etienne, et al. "Extensive global wetland loss over the past three centuries." *Nature* 614.7947 (2023): 281-286.

Gallant, John C., and Trevor I. Dowling. "A multiresolution index of valley bottom flatness for mapping depositional areas." *Water resources research* 39.12 (2003).

Google Developers. (n.d.). Sentinel-2 MSI: MultiSpectral Instrument, Level-2A. Retrieved May 15, 2024, from https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED#bands

Gorelick, Noel, et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." Remote sensing of Environment 202 (2017): 18-27.

Halabisky, Meghan, et al. "The Wetland Intrinsic Potential tool: Mapping wetland intrinsic potential through machine learning of multi-scale remote sensing proxies of wetland indicators." EGUsphere 2022 (2022): 1-19.

Herrmann, I., et al. "Assessment of leaf area index by the red-edge inflection point derived from VENμS bands." *Proceedings of the ESA hyperspectral workshop*, Frascati, Italy. Vol. 683. 2010.

Huete, A. R., et al. "A comparison of vegetation indices over a global set of TM images for EOS-MODIS." *Remote sensing of environment* 59.3 (1997): 440-451.

Hunt Jr, E. Raymond, and Barrett N. Rock. "Detection of changes in leaf water content using near- and middle-infrared reflectances." *Remote sensing of environment* 30.1 (1989): 43-54.

Jahncke, Raymond, et al. "Mapping wetlands in Nova Scotia with multi-beam RADARSAT-2 Polarimetric SAR, optical satellite imagery, and Lidar data." *International Journal of Applied Earth Observation and Geoinformation* 68 (2018): 139-156.

Kaplan, Gordana, and Ugur Avdan. "Object-based water body extraction model using Sentinel-2 satellite imagery." *European Journal of Remote Sensing* 50.1 (2017): 137-143.

Khosravipour, Anahita, et al. "Effect of slope on treetop detection using a LiDAR Canopy Height Model." *ISPRS journal of photogrammetry and remote sensing* 104 (2015): 44-52.

Kopecký, Martin, Martin Macek, and Jan Wild. "Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition." *Science of the Total Environment* 757 (2021): 143785.

LaRocque, Armand, et al. "Wetland mapping with landsat 8 OLI, sentinel-1, ALOS-1 PALSAR, and LiDAR data in Southern New Brunswick, Canada." *Remote Sensing* 12.13 (2020): 2095.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

Li, Xiuzhen, et al. "Coastal wetland loss, consequences, and challenges for restoration." *Anthropocene Coasts* 1.1 (2018): 1-15.

Lindsay, John B., J. M. H. Cockburn, and H. A. J. Russell. "An integral image approach to performing multi-scale topographic position analysis." *Geomorphology* 245 (2015): 51-61.

Ma, Chunfeng, Xin Li, and Matthew F. McCabe. "Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data." *Remote Sensing* 12.14 (2020): 2303.

Mahdavi, Sahel, et al. "Remote sensing for wetland classification: A comprehensive review." *GIScience & remote sensing* 55.5 (2018): 623-658.

Mainali, Kumar, et al. "Convolutional neural network for high-resolution wetland mapping with open data: Variable selection and the challenges of a generalizable model." *Science of The Total Environment* 861 (2023): 160622.

Mattivi, Pietro, et al. "TWI computation: a comparison of different open source GISs." Open Geospatial Data, Software and Standards 4.1 (2019): 1-12.

McFeeters, Stuart K. "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." *International journal of remote sensing* 17.7 (1996): 1425-1432.

Murphy, Paul NC, et al. "Mapping wetlands: a comparison of two different approaches for New Brunswick, Canada." *Wetlands* 27.4 (2007): 846-854.

National Wetlands Working Group. 1997. The Canadian Wetland Classification System, 2nd Edition. Warner, B.G. and C.D.A. Rubec (eds.), Wetlands Research Centre, University of Waterloo, Waterloo, ON, Canada. 68 p.

Neily, Peter D., et al. Ecological land classification for Nova Scotia. Nova Scotia Department of Natural Resources, Renewable Resources Branch, 2003.

Neily, P., Basquill, S., Quigley, E., & Keys, K. (2017). Ecological Land Classification for Nova Scotia. Nova Scotia Department of Natural Resources, Renewable Resources Branch.

Pang, Y., et al. "Identifying spectral features of characteristics of Sphagnum to assess the remote sensing potential of peatlands: A case study in China." *Mires and Peat* 26.25 (2020).

Pasquarella, Valerie J., et al. "Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

Reis, Vanessa, Virgilio Hermoso, Stephen K. Hamilton, Douglas Ward, Etienne Fluet-Chouinard, Bernhard Lehner, and Simon Linke. "A global assessment of inland wetland conservation status." *Bioscience* 67, no. 6 (2017): 523-533.

Stepinski, Tomasz F., and Jaroslaw Jasiewicz. "Geomorphons-a new approach to classification of landforms." Proceedings of geomorphometry 2011 (2011): 109-112.

Tesfamariam, Birhane Gebrehiwot, Berhan Gessesse, and Farid Melgani. "MaxEnt-based modeling of suitable habitat for rehabilitation of Podocarpus forest at landscape-scale." Environmental Systems Research 11.1 (2022): 4.

Tozer, Douglas C., Owen Steele, and Mark Gloutney. "Multispecies benefits of wetland conservation for marsh birds, frogs, and species at risk." *Journal of environmental management* 212 (2018): 160-168.

Uuemaa, Evelyn, Andrew O. Hughes, and Chris C. Tanner. "Identifying feasible locations for wetland creation or restoration in catchments by suitability modelling using light detection and ranging (LiDAR) digital elevation model (DEM)." *Water* 10.4 (2018): 464.

Vinod, P. G. "Development of topographic position index based on Jenness algorithm for precision agriculture at Kerala, India." *Spatial Information Research* 25.3 (2017): 381-388.

Wang, Jing, et al. "Mapping Spartina alterniflora biomass using LiDAR and hyperspectral data." *Remote Sensing* 9.6 (2017): 589.

Wardlow, Brian D., Stephen L. Egbert, and Jude H. Kastens. "Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains." *Remote sensing of environment* 108.3 (2007): 290-310.

White, Barry, et al. "Using the cartographic depth-to-water index to locate small streams and associated wet areas across landscapes." Canadian Water Resources Journal/Revue canadienne des ressources hydriques 37.4 (2012): 333-347.

# 8 APPENDICES

## 8.1 PROCESSING TOOLS

**Note this process was conducted on a 12th Gen Intel(R) Core(TM) i9-12900K 3.20 GHz with 128 GB of usable ram**

Processing steps used for all data presented as results in this report are provided in accompaniment to this document as:

\scripts\workflow_v2:

_run.py

This python script provides an example of the full model run including pathing to the original input data locations (ie, run2). For the purposes of testing these tools externally, a full copy of all 38 explanatory rasters has been provided has been provided at **\Outputs\run2\1_rasters**.

***Note these rasters do include direct mosaiced layers as described from google earth engine (GEE) and are intended for explicitly for the validation of the methods of this report and not available to the redistributed.***

Accompanying script files are required to be located with _run.py for the process to function. These include:

- _0_utils.py
  - Contains a set of logging and other functions
- _1_generate.py
  - Performs the raster input layer resampling and raster index computations producing 38 explanatory rasters (10 m resolution)
- _2_train.py
  - Performs the AutoML training using selected training points and all rasters.
- _3_segment.py
  - computes segmentation based on 3 select raster producing vector output
- _4_predict.py
  - Iterates over batches of segment polygons to compute wetland prediction and model confidence per segment. Provides a final merged wetland classification segment feature class polygon.

For the purposes of testing this script, it is recommended to skip the generate step and utilize the provided example rasters as indicated. This can be done with the example provided in _run.py:

```
run_test(r"INSERT DIRECTORY LOCATION FOR 20240709_AGRG_NS_Wetlands")
```

to run this, modify the section in _run.py to the location where the provided folder "20240709_AGRG_NS_Wetlands" was copied locally. Note that this full list of is 257 GB.

For example, this can be modified as (if placed on the D drive root):

```
run_test(r"D:\20240709_AGRG_NS_Wetlands")
```

this will be default attempt to process the entire province.

To set a processing extent, the following can be attempted with a shapefile of a test extent

```
run_test(r"INSERT DIRECTORY LOCATION FOR 20240709_AGRG_NS_Wetlands", extent =
r"PATH TO SHAPEFILE")
```

ie:

```
run_test(r"D:\20240709_AGRG_NS_Wetlands", extent = r"D:\test.shp")
```

Note: It may also be wise to skip the segmentation step as this an be ram intensive. In such case, the user may want to attempt the train and autoML directly using the ArcGIS interface. Note that with the provided training data, rasters, and segments, new models and predictions can be computed directly but some complications may exist in ensuring explanatory raster names are kept consistent between process steps when performing the analysis step by step.

For continued support and updates to this workflow please reach out to kevin.mcguian@nscc.ca for access to the GitHub repository.

## 8.2 Detailed Methods

### 8.2.1 Verbose Methods

This appendix delineates the comprehensive methodology employed for applying an ensemble classification model, trained using AutoML, to raster data for enhancing wetland mapping within Areas of Interest (AOIs) in Nova Scotia. The workflow is articulated to support replication by researchers or GIS professionals, detailing specific GIS tools, input datasets, and procedural steps utilized.

#### *8.2.1.1 Data Inputs*

**Training Dataset:** A shapefile comprising pre-classified wetland and non-wetland samples for model training.

**Environmental and Topographical Rasters:** A collection of rasters, including the Topographic Wetness Index (TWI), Chrome DEM, Digital Elevation Model (DEM), and others, employed as explanatory variables for model training and prediction.

**Composite Raster:** A merged raster combining all environmental and topographical rasters post-scaling and preprocessing for use in the model application phase.

#### *8.2.1.2 Thresholds and Parameters*

**AutoML Configuration:** Parameters such as total time limit for the AutoML run and the percentage of data designated for validation.

**Scaling Factors:** Specific minimum and maximum values used for normalizing rasters like DEM and NDVI to a 0-255 range, facilitating better model performance.

#### *8.2.1.3 Analysis Steps*

**Model Training with AutoML:** Utilize the TrainUsingAutoML function within ArcGIS's GeoAI Tools toolbox, specifying the training dataset, the path for the output model, target prediction variable, and explanatory rasters.

Output consists of the trained model and a comprehensive report detailing the model's performance.

**Raster Data Preprocessing:** Create a composite raster by merging and scaling environmental and topographical rasters according to predefined thresholds, preparing the dataset for segmentation and prediction.

**Segmentation:** Apply the SegmentMeanShift tool on the preprocessed composite raster to delineate the landscape into meaningful units (segments) for subsequent prediction.

#### *8.2.1.4 Application of the Trained Model:*

Execute the trained AutoML model on the segmented raster using the PredictUsingAutoML function, matching the explanatory variables defined during training to the segments.

#### *8.2.1.5 Post-processing:*

Convert the prediction output raster to a polygon feature class for easier analysis and visualization.

Join attributes from the prediction result to the polygon feature class, enriching it with the model's predictions and associated confidence scores.

#### *8.2.1.6 Outputs*

**Wetland Prediction Polygons:** Polygons derived from the model's predictions, indicating the predicted presence of wetlands within the AOIs.

These polygons are subsequently subjected to a quality assurance process to ensure their accuracy and reliability.

#### *8.2.1.7 Tools and Functions Utilized*

**ArcGIS Tools:** Tools from ArcGIS and the GeoAI toolbox, including TrainUsingAutoML, SegmentMeanShift, Raster to Polygon Conversion, and Join Field.

**Custom Scripting:** Python scripts within the ArcGIS environment to automate the application of the ensemble classification model across various AOIs, ensuring efficient and scalable analysis.

**For more information and access to the script in a GitHub repository, contact kevin.mcguigan@nscc.ca**