

Habitat Suitability Modelling for Northern Goshawk in the Lillooet River Valley Using Geospatial Data



Prepared by:
Ryan Milia
for the Centre of Geographic Sciences

Abstract

This project developed a habitat suitability model for the Northern Goshawk (*Accipiter gentilis*) in the Tsetspa7 Forest Licence area, British Columbia, to support conservation efforts through advanced geospatial analysis. The study integrated Light Detection and Ranging (LiDAR), Vegetation Resource Inventory (VRI), and Geographic Information Systems (GIS) with Maximum Entropy (MaxEnt) modelling to identify critical habitat characteristics and predict high-suitability areas. Three MaxEnt models were developed: Model 1 (12 original occurrence points), Model 2 (212 points including 200 synthetic), and Model 3 (23 original points with comprehensive LiDAR coverage). Model 3 emerged as the preferred approach, providing the most detailed habitat insights by fully leveraging LiDAR-derived variables across the entire study area. Results classified habitat into low (55.37%), medium (28.95%), and high (15.68%) suitability zones, with high-suitability areas concentrated in mature coniferous forests. Deliverables include habitat suitability maps, R scripts, and this report detailing methodology, results, and recommendations for Northern Goshawk conservation in the Lillooet River Valley. This model provides actionable data to guide protective measures for this threatened species.

Table of Contents

Abstract	1
Background	3
Objectives	4
Methods.....	4
Study Area.....	4
Data Description	4
Data Processing.....	5
Variable Generation	5
MaxEnt Modelling.....	6
Model Validation.....	6
Results.....	6
Data Processing.....	6
Model Performance.....	6
Habitat Suitability Maps	7
Challenges.....	8
Recommendations.....	9
Conclusion	9
References:.....	9
Appendices:.....	10

Background

The Northern Goshawk (*Accipiter gentilis*) is a large forest raptor and apex predator, occupying a crucial role within temperate forest ecosystems (Wright et al., 2022). Its habitat requirements are highly specific, depending on factors such as forest composition, slope, canopy cover, and prey availability (Mahon, 2025). As an indicator species, the Northern Goshawk reflects the health of its habitat. The coastal subspecies (*Accipiter gentilis laingi*) is listed as threatened under the federal Species at Risk Act (SARA) and is also provincially red-listed in British Columbia (Wright et al., 2022). Its dependence on open, mature, and old-growth forests for foraging and nesting makes it particularly vulnerable to habitat loss due to logging and other forest management activities (Mahon, 2025). Due to these habitat requirements, the coastal subspecies faces a high risk of population decline, threatening its long-term survival in coastal British Columbia (Mahon, 2025).

The habitat of the coastal Northern Goshawk is influenced by various factors, including the availability of large, undisturbed forested areas that provide cover, foraging opportunities, and safe nesting sites (Wright et al., 2022). Unfortunately, these habitats are increasingly fragmented and degraded due to human activities such as logging, including salvage operations in response to mountain pine beetle (*Dendroctonus ponderosae*) infestations, which reduce suitable habitat (Mahon, 2025). Additionally, forest management practices that do not account for the specific habitat requirements of the Northern Goshawk contribute to the decline in habitat quality, increasing the risk to the species. Therefore, identifying and protecting critical habitats for the Northern Goshawk is essential for its conservation (Wright et al., 2022; Mahon, 2025).

Habitat suitability modelling is a valuable tool for guiding conservation and forest management strategies by predicting the potential distribution of species within an area. For species like the Northern Goshawk, whose habitat is in decline, these models can help inform conservation strategies by identifying areas that should be protected (Wright et al., 2022). One widely used and effective technique for habitat suitability analysis is MaxEnt modelling. This approach allows researchers to identify critical habitat features and assess the potential impact of land use changes on species conservation.

Remote sensing data, particularly LiDAR, has become an essential resource in habitat suitability studies. LiDAR provides highly accurate measurements of forest structure, including canopy height, vegetation density, and terrain characteristics, all key factors influencing goshawk habitat selection (Mahon, 2025). When combined with VRI data, which offers detailed information on forest composition and age classes, these datasets enable a comprehensive analysis of habitat conditions. Geographic Information Systems enhance the ability to analyze multiple spatial datasets, supporting evidence-based decisions for conservation and resource management (Wright et al., 2022). GIS tools facilitate the visualization of habitat distribution, assessment of habitat fragmentation, and monitoring of environmental changes over time.

This project utilized advanced geospatial techniques and MaxEnt modeling to develop habitat suitability models for the Northern Goshawk in the Tsetspa7 Forest Licence area. By integrating LiDAR, VRI, and species occurrence data, the study identified critical habitat features and produced suitability maps to guide forest management and conservation efforts.

Objectives

The goal of this project was to develop a habitat suitability model for the Northern Goshawk (*Accipiter gentilis*) in the Lillooet River Valley. The model was used to identify key habitat characteristics and predict areas where suitable habitat was most likely to occur. The objectives of the study were to: (i) assess the habitat suitability of the Northern Goshawk using environmental and ecological variables, (ii) evaluate the impact of forest structure, such as canopy cover and vegetation density, on habitat suitability, (iii) develop habitat suitability maps using MaxEnt modelling to classify areas as low, medium, or high suitability, and (iv) integrate LiDAR, VRI, and GIS to refine the model.

Methods

Study Area

The study area is the Tsetsa7 Forest Licence, located within the Sea to Sky Forest District. The project site is north of Harrison Lake, extending east of Golden Ears Provincial Park and southeast of Lillooet Lake, in the lower Lillooet River valley. Access to the area is via Duffey Lake Road (Highway 99) and the Lillooet River Forest Service Road (FSR). Figure 1 shows the geographic extent of the licence area.

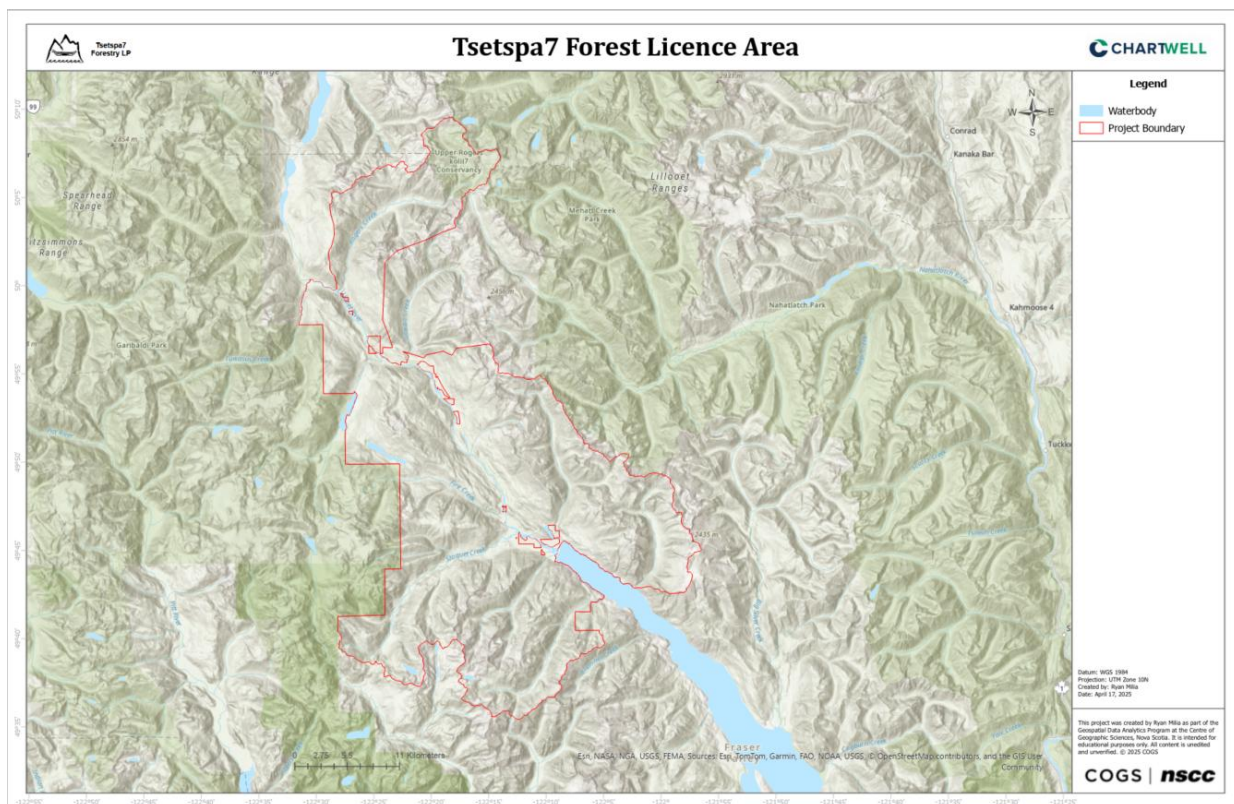


Figure 1. Extent of Tsetsa7 Project Area: Boundary is shown within the red polygon

Data Description

This study utilized multiple geospatial data sources to develop a comprehensive habitat suitability model. LiDAR data, provided by Tsetsa7 Forestry, supplied detailed topographic information including elevation, slope, and terrain features in LAZ and LAS formats. Vegetation

Resource Inventory (VRI) data, sourced from the British Columbia open data portal, provided essential information on vegetation cover, species composition, and forest age classes (Government of British Columbia, 2023). Species occurrence data was obtained from Chartwell Resource Group Ltd. To accurately represent the landscape, water body information was incorporated from Statistics Canada's "Water File - Lakes and Rivers" dataset, which allowed for the masking of water features from the analysis.

Data Processing

The processing workflow involved several strategic steps to prepare the data for habitat suitability modelling. First, 298 LAZ files were converted to LAS format using ArcGIS Pro's "Convert LAS" tool in a batch operation. Additional LAS files were unzipped using 7zip software, with all data stored on an external drive due to the significant storage requirements. The VRI dataset processing involved clipping three datasets to the project boundary, with only the VEG_COMP_LYR_L1_POLY dataset retained for analysis. Unnecessary fields were removed to streamline the data and improve processing efficiency. Water features were masked from both LiDAR and VRI datasets by converting Statistics Canada water data to raster format and applying as a mask. The species occurrence data required conversion from KML to a vector point layer, followed by removal of 12 duplicate points to ensure data integrity. Environmental variables were then extracted to these points using the "Extract Multi Values to Points" tool to enable model training.

Variable Generation

The generation of environmental variables focused on creating meaningful predictors for the habitat suitability model. From the LiDAR data, two Digital Elevation Models (DEMs) and Digital Surface Models (DSMs) were created using the "LAS Dataset to Raster" tool. These were used to develop two Canopy Height Models (CHMs) through the "Calculate Raster" tool. The separate DEMs and CHMs were subsequently merged using the "Mosaic to New Raster" tool and reprojected to NAD 1983 UTM Zone 10N to resolve pyramid display issues. All raster files were clipped to the project boundary, which had been converted from KML to SHP format using the "Clip Raster" tool.

From the VRI data, several classification variables were derived based on their relevance to Northern Goshawk habitat requirements. Forest Type was categorized as Coniferous (1), Deciduous (0), or Unknown (2). Crown Closure was classified as Low (0-30%, value 3), Medium (30-70%, value 2), or High (70-100%, value 1). Stand Age was divided into Young (0-40 years, value 3), Mature (40-100 years, value 2), and Old Growth (100+ years, value 1). BEC Zone was simplified to High Suitability (CWH/MH, value 1) or Low Suitability (all others, value 0). The Vertical Complexity variable was ultimately excluded from the analysis due to extensive missing data in the northern project area. All classification variables were rasterized to a 5m resolution using the "Polygon to Raster" tool to maintain consistency with the LiDAR-derived raster files.

MaxEnt Modelling

Three distinct MaxEnt models were developed in RStudio, each documented in separate markdown notebooks. Model 1 (*tsetspa7_model*) used the original 12 known occurrence points with a regularization multiplier (regmult) of 6 to reduce overfitting due to limited sample size. It incorporated three predictors: stand age, forest type, and BEC zone. Due to the limited species occurrence data, I decided to create a second model with additional synthetic occurrence points. Model 2 (*tsetspa7_synthetic_model*) expanded the dataset to 212 presence points by generating 200 synthetic points based on environmental similarity to the original occurrences. This model used a lower regmult of 2 to allow more model flexibility and included five predictors: stand age, forest type, BEC zone, DEM, and CHM. Model 3 (*tsetspa7_model_v2*) utilized 23 occurrence points and maintained the same five predictors as Model 2. This model incorporated the full extent of available LiDAR data across the project area, enabling a higher-resolution characterization of habitat conditions. A regmult of 4 was used to balance model complexity and generalizability. Each model was calibrated to optimize the trade-off between model fit and predictive power, with regularization multipliers adjusted to mitigate overfitting while preserving ecologically relevant patterns.

Model Validation

Model validation employed a combination of robust techniques to evaluate the performance and reliability of each MaxEnt model. Leave-One-Out Cross-Validation (LOOCV) was applied to all three models. For Models 2 and 3, which included larger datasets, K-Fold Cross-Validation was also conducted to provide more stable and generalizable performance metrics. A 70/30 train-test split was used across all models to assess sensitivity (true positive rate) and specificity (true negative rate). Area Under the Curve (AUC) values were calculated to measure each model's ability to discriminate between suitable and unsuitable habitat, with higher values indicating better performance. Finally, the spatial distribution of predicted habitat suitability - classified into low, medium, and high categories - was compared across models to evaluate their practical implications for conservation planning. These combined methods provided a thorough assessment of model accuracy and informed the selection of the most appropriate approach for supporting goshawk conservation efforts.

Results

Data Processing

The processing of geospatial data presented several challenges that were successfully addressed. LiDAR data conversion and management required significant computational resources but was completed effectively despite storage and processing limitations. All LiDAR data was clipped to match the project area. The VRI data was classified using python and generated rasterized variables for forest type, crown closure, stand age, and BEC subzones, with the vertical complexity variable excluded due to significant data gaps. Extent misalignment issues between datasets were resolved using the "Resample" tool with the "Snap Raster" function, ensuring that all raster files shared identical spatial extent, resolution (5m), and projection (NAD 1983 UTM Zone 10N). Crown closure was excluded from the modeling process due to its low predictive importance.

Model Performance

The three MaxEnt models demonstrated varying performance characteristics across validation metrics. Model 1, based on 12 original occurrence points, showed high specificity in the 70/30 train-test split with a presence prediction rate of 0.3985 and a background prediction rate of 0.0005. In Leave-One-Out Cross-Validation (LOOCV), it achieved an AUC of 0.9182, with a mean predicted suitability at presence points of 0.7090 and a maximum suitability of 0.7730. The area distribution was 32.95% low suitability, 28.47% medium suitability, and 38.58% high suitability. While Model 1 demonstrated strong precision for the original occurrences, its generalizability was limited by the small sample size.

Model 2, which incorporated 212 presence points (12 original and 200 synthetic), improved sensitivity in the 70/30 split with a presence prediction rate of 0.5122 and a background prediction rate of 0.0870. LOOCV produced a high AUC of 0.9567, with a mean suitability of 0.1763 and a maximum suitability of 0.9398. K-Fold Cross-Validation confirmed strong performance with an AUC of 0.9500, a mean suitability at presence points of 0.5052, and a background rate of 0.1867. Area classification results showed 47.94% low suitability, 33.61% medium suitability, and 18.44% high suitability. Model 2's performance was bolstered by the larger sample size and strong generalization capacity.

Model 3, using 23 original occurrence points and full LiDAR coverage, demonstrated moderate performance in the 70/30 validation with a presence prediction rate of 0.5498 and a background prediction rate of 0.2988. LOOCV yielded an AUC of 0.8268, with a mean predicted suitability of 0.288, a maximum suitability of 0.9076, and a mean suitability at presence points of 0.5323. K-Fold validation further supported this with a comparable AUC of 0.8275 and a background rate of 0.2891. The area distribution was the most conservative, with 55.37% low suitability, 28.95% medium suitability, and 15.68% high suitability. Although Model 3 had a lower AUC than the others, it provided the most spatially detailed habitat characterization by leveraging high-resolution LiDAR data across the full study area.

Habitat Suitability Maps

The habitat suitability maps generated from the three MaxEnt models revealed distinct spatial patterns in predicted habitat quality for the Northern Goshawk across the Tsetspa7 Forest Licence area. Model 1 predicted a higher proportion of high-suitability habitat (38.58% of the study area), with 28.47% medium and 32.95% low suitability, reflecting its reliance on a limited dataset of 12 occurrence points. Model 2, incorporating 212 points (including 200 synthetic), showed a more even distribution, with 47.94% low, 33.61% medium, and 18.44% high suitability, indicating a broader spread of predicted habitat quality. Model 3, utilizing 23 original occurrence points, produced the most conservative predictions, classifying 55.37% of the area as low suitability, 28.95% as medium, and 15.68% as high suitability. This conservative distribution highlights Model 3's detailed habitat characterization, driven by the integration of additional LiDAR data across the entire study area. Comparative analysis of the maps revealed strong correlations between habitat suitability and key predictor variables, notably BEC zone (CWH/MH), forest type (coniferous), and stand age (mature and old growth), underscoring the relevance of these factors in shaping goshawk habitat preferences. The final habitat suitability

map for Model 3, which offers the most spatially complete and precise delineation of potential habitat, is presented in Figure 2.

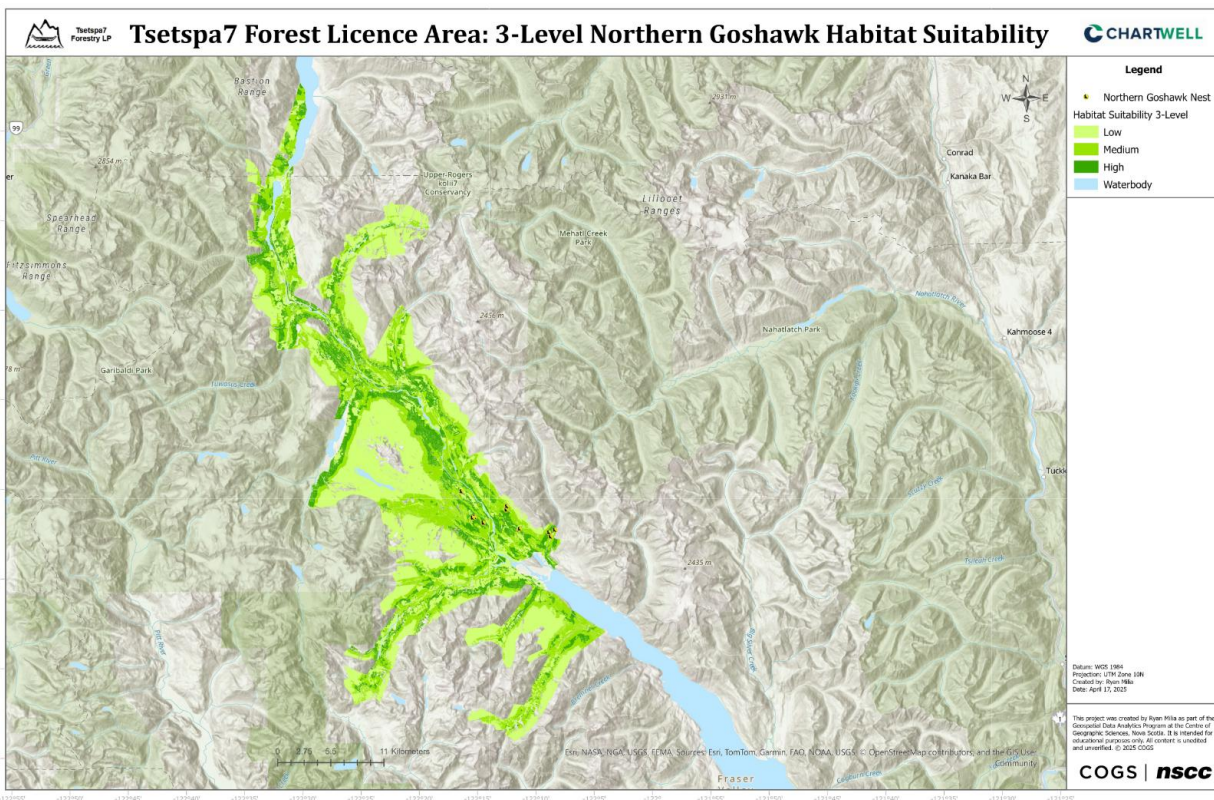


Figure 2. Final Habitat Suitability Map for Model 3.

Challenges

The development of the Northern Goshawk habitat suitability model presented several significant technical and methodological challenges throughout the project. Data processing, particularly the management of large LiDAR datasets, required considerable time investment and necessitated the use of external storage due to the substantial file sizes. Extent misalignment between different data sources presented a persistent challenge that required iterative application of "Extract by Mask" and "Resample" tools to ensure spatial consistency across all datasets. Missing data posed another obstacle, particularly for the vertical complexity variable which had to be excluded from analysis due to extensive NULL values in the northern project area. The limited occurrence dataset in Model 1 (with only 12 species occurrence points) constrained validation options and required careful interpretation of results; this limitation was addressed through the introduction of synthetic points in Model 2. The computational intensity of processing was particularly evident during DEM and CHM creation, which took several hours and required careful parameter tuning, especially for extent settings, to ensure accurate results. Despite these challenges, the project successfully developed three viable models with different strengths for conservation application.

Recommendations

Based on the habitat suitability analysis, conservation efforts should prioritize the high-suitability areas identified by Model 3, which covers 15.68% of the study area. Although Model 3 has a lower AUC compared to other models, its comprehensive LiDAR coverage offers the most spatially accurate representation of potential habitat across the project area. To improve model accuracy, additional field-verified occurrence data is needed to reduce reliance on synthetic points. Expanding this approach to other regions of British Columbia would support broader Northern Goshawk conservation, particularly for the threatened coastal subspecies. Using the final model to detect anticipated habitat enables proactive planning to balance conservation and sustainable forestry.

Conclusion

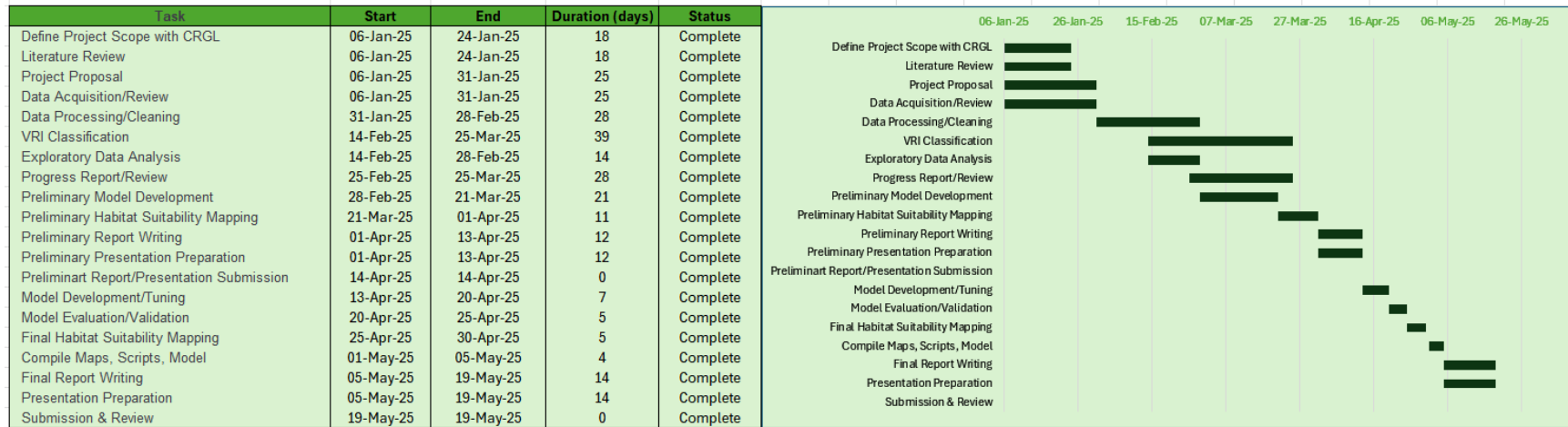
This project successfully developed habitat suitability models for the Northern Goshawk in the Tsetspa7 Forest Licence area through the integration of LiDAR, VRI, and MaxEnt modeling techniques. Among the three models developed, Model 3 emerged as the preferred approach despite its lower AUC (0.8275) compared to Models 1 and 2. Its preference stems from its utilization of comprehensive LiDAR coverage across the entire project area and its reliance on a larger set of original occurrence points (23), providing the most spatially complete and biologically relevant habitat characterization. The resulting suitability maps classify 15.68% of the study area as high suitability, 28.95% as medium suitability, and 55.37% as low suitability, offering a conservative but precise delineation of potential Northern Goshawk habitat. The methodology developed and insights gained aim to contribute to the long-term conservation of the Northern Goshawk in British Columbia, demonstrating the value of integrated geospatial analysis in species conservation efforts.

References:

- Mahon, T. (2025). *Northern goshawks in west-central British Columbia*. Retrieved from <https://www.sccp.ca/sites/default/files/species-habitat/documents/08-07NadinaGoshawkTechReport.pdf>
- Smith, A., et al. (2012). Using maximum entropy modeling to predict the potential distributions of large trees for conservation planning. *Ecosphere*, 3(6), 1–21. <https://doi.org/10.1890/ES12-00053.1>
- Wright, M. E., Anderson, R., & Thomas, J. (2022). Habitat suitability modeling and ecological forecasting of northern goshawk nesting habitat. *bioRxiv*. <https://doi.org/10.1101/2022.01.01.22282965>
- COSEWIC. (2013). *COSEWIC assessment and status report on the Northern Goshawk (Accipiter gentilis) in Canada*. Environment and Climate Change Canada.
- Stuart-Smith, K., Wells, R., & Harrower, W. (2011). *A comparison of historic, current, and future nesting habitat for the Northern Goshawk in the East Kootenay region of British Columbia (Final Report FSP # Y102102)*. Retrieved from https://www.hcwildlife.ca/uploads/8/4/9/1/84917238/stuart-smith_et_al_ek_nogo_current_historic_29mar2011.pdf
- Government of British Columbia. (2018). *Implementation plan for the recovery of Northern Goshawk, laingi subspecies*. Ministry of Forests, Lands, Natural Resource Operations and Rural Development. Retrieved from https://www2.gov.bc.ca/assets/gov/environment/plants-animals-and-ecosystems/species-ecosystems-at-risk/recovery-planning/implementation_plan_for_the_recovery_of_northern_goshawk.pdf
- Parkkinen, A. (2021). *Prediction of goshawk nesting sites in Central Finland with MaxEnt modeling and RandomForest classifier* (Master's thesis, University of Jyväskylä). Retrieved from <https://www.readkong.com/page/prediction-of-goshawk-nesting-sites-in-central-finland-with-3494263>
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Government of British Columbia. (2023). *Vegetation resource inventory (VRI)*. Retrieved from <https://catalogue.data.gov.bc.ca/dataset/vegetation-resource-inventory-vri>
- Government of British Columbia. (n.d.). *Spatial ready data for forest inventory management*. Retrieved from <https://www2.gov.bc.ca/gov/content/industry/forestry/managing-our-forest-resources/forest-inventory/data-management-and-access/spatial-ready-data>

Appendices:

Appendix A – Gantt Chart



Appendix B – Data Dictionary

Name	Description	Type	Source	Scale/Resolution	Coordinate System
VEG_COMP_LYR_LVL_POLY_2023.gdb	Vegetation composition layer for Level 1 polygons in 2023	GDB	British Columbia open data portal	1:20,000	EPSG:32610 (WGS 84/UTM Zone 10N)
Tsetspa7 FNvL v2.kml	First Nations Woodlot Licence Area	KML	CRGL LTD.	1:50,000	EPSG:32610 (WGS 84/UTM Zone 10N)
Tsetspa7 GH Nests v2.kml	Species Occurrence Data for Northern Goshawk	KML	CRGL LTD.	1:50,000	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.shx	Shape index format for Project Boundary	SHX	CRGL LTD.	1:10,000	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.shp.xml	Metadata for Project Boundary shapefile	XML	CRGL LTD.	1:10,000	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.shp	Shapefile format for Project Boundary	SHP	CRGL LTD.	1:10,000	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.sbx	Spatial index for Project Boundary shapefile	SBX	CRGL LTD.	N/A	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.sbn	Spatial index for Project Boundary shapefile	SBN	CRGL LTD.	N/A	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.prj	Projection format for Project Boundary shapefile	PRJ	CRGL LTD.	N/A	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.dbf	Attribute format for Project Boundary shapefile	DBF	CRGL LTD.	N/A	EPSG:32610 (WGS 84/UTM Zone 10N)
Project_Boundary_v1.cpg	Code page for Project Boundary shapefile	CPG	CRGL LTD.	N/A	N/A
FC_CHM_WM_1m.tif	LiDAR derived Canopy Height Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
FC_DEM_WM_1m.tif	LiDAR derived Digital Elevation Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
HED_CHM_1m.tif	LiDAR derived Canopy Height Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
HED_DEM_1m.tif	LiDAR derived Digital Elevation Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
LillooetTuvassus_CHM_WM.tif	LiDAR derived Canopy Height Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
LillooetTuvassus_DEM_WM.tif	LiDAR derived Digital Elevation Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
TSE_CHM_1m_WM.tif	LiDAR derived Canopy Height Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
TSE_DEM_1m_WM.tif	LiDAR derived Digital Elevation Model	TIF	CRGL LTD.	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_562_5512_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_546_5543_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_545_5530_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_543_5548_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_542_5538_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_541_5542_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)
WHS_540_5533_UTM10_SPO00-3461-02_APO00-3461-02_v1.laz	LiDAR data for specific area	LAZ	Tsetspa7 Forestry	1:10,000/1m	EPSG:32610 (WGS 84/UTM Zone 10N)