**Predicting Presence of Sharp-Tailed Grouse in Lac du Bois Grasslands Protected Area Based on Habitat Availability**

NRSC 4480 / ENVS 5480: Ecological Modelling

Tay Powrie, Calen Wong, Clara Boisclair Boussard, Cooper Butchart

Thompson Rivers University

****

Table of Contents

[Table of Figures 3](#_Toc163650726)

[Table of tables 3](#_Toc163650727)

[Introduction 4](#_Toc163650728)

[Methods 5](#_Toc163650729)

[Data Collection 5](#_Toc163650730)

[Model Creation 5](#_Toc163650731)

[Results 6](#_Toc163650732)

[Model Selection 6](#_Toc163650733)

[Model Assessment 7](#_Toc163650734)

[Discussion 10](#_Toc163650735)

[Appendix A - R Script 16](#_Toc163650736)

# Table of Figures

[Figure 1. Map of Lac du Bois Grasslands Protected Park in Relation to Kamloops. Data obtained from BC Data Catalogue (Government of British Columbia, 2019). 4](#_Toc163650632)

[Figure 2. Variable importance table ranking the variables used within a ranger probability model to determine the presence or absence of Columbia Sharp-tailed Grouse in the Lac du Bois grasslands. 8](#_Toc163650633)

[Figure 3. Prediction map of Sharp-tailed Grouse in the Lac du Bois grasslands 9](#_Toc163650634)

# Table of tables

[Table 1. Model parameters used. 6](#_Toc163650650)

[Table 2. Aggregated results of each resampling iteration based on outer tuning for spatial cross validation. The highlighted value with the asterisk is the model we selected as best fit for the study. 7](#_Toc163650651)

[Table 3. Confusion matrix for a ranger probability model used to determine the presence or absence of Sharp-tailed Grouse in the Lac du Bois grasslands. 8](#_Toc163650652)

# Introduction

Grasslands in British Columbia (BC), and worldwide are extremely important for biodiversity and face a threat from tree encroachment from lack of wildfire, and human encroachment from urbanization (Leupin & Chutter, 2007). There are many species in grasslands that have historically low numbers and with a habitat that is continually shrinking, it is important to employ as many tactics as possible to create an understanding of what the future may hold for them, and if our management tactics are sufficient or need altering. Lac du Bois Grasslands Protected Area (Lac du Bois) is the second largest area of protected grassland in BC (BC Parks, 2023). It is located Northwest of Kamloops, spanning an area of 15,207 hectares, and located within the traditional Territory of the Tk’emlúps te Secwe̓pemc peoples (BC Parks, 2000).

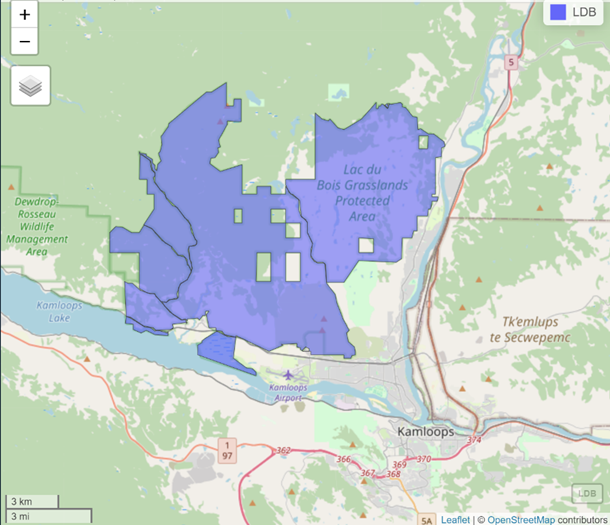


Figure . Map of Lac du Bois Grasslands Protected Park in Relation to Kamloops. Data obtained from BC Data Catalogue (Government of British Columbia, 2019).

Columbian Sharp-Tailed Grouse (CSTG) is one of six subspecies of Sharp-Tailed Grouse, but it is the only subspecies found in Lac du Bois (Leupin & Chutter, 2007). CSTG is a blue-listed species in BC likely due to the loss of grassland habitat (Leupin & Chutter, 2007). Grasslands are an important habitat for this subspecies during spring, summer, and fall (BC Parks, n.d.). CSTG are an indicator of a healthy grassland ecosystem and are sensitive to human disturbances (BC Parks, 2023).

The intent of this project is to determine habitat suitability for the Columbian Sharp-Tailed Grouse (*Tympanuchus phasianellus columbianus*) in Lac du Bois near Kamloops, BC. Coupling Lidar imagery data and previous bird survey data, we employed machine learning algorithms to create an ecological model which will help predict habitat suitability. Ecological modeling can serve as an essential tool for determining habitat suitability for species which could help with determining species abundance and preferential locations, giving land managers important information on conservation recommendations.

# Methods

## Data Collection

We modeled presence probability for CSTG (Figure 3) using three historical datasets, obtained from the publicly available BC Wildlife Species Inventory (WSI) (Government of British Columbia, 2024). Through a combination of multi-year telemetry observations, survey observations, and incidental observations spatial presence absence data for CSTG was recorded (n = 199).

LIDAR data for the Lac du Bois area was retrieved from Brian Wallace of the Ministry of Forests.

## Model Creation

All analysis for this presence probability model was conducted using R version 4.3.2 and a copy of all reproducible code can be found in Appendix A (R Core Team, 2023).

LIDAR data was initially retiled into 500m x 500m grid cells for improved processing rate. Data was then ground classified and normalized to create a digital elevation model (DEM) (Roussel et al., 2020). Using the R package Rsagacmd, various model layers were created using SAGA-GIS processing tools (Pawley, 2023). The model layers that were used in this analysis were: aspect, canopy height, elevation, overland flow distance, multi-resolution ridge top flatness (MRRTF), multi-resolution valley bottom flatness (MRVBF), negative topographic openness, positive topographic openness, slope, topographic wetness index, and topographic ruggedness index. All layers other than the canopy height model were derived from the DEM above. Using the R package terra, we then extracted the value of each spatial layer to the CSTG presence data frame (Hijmans, 2023). Using a random forest algorithm, we then created a tuned ranger probability model and employed spatial cross-validation to account for spatial autocorrelation (Becker & Schratz 2024; Brieman, 2001; Lang et al. 2019; Schratz & Becker, 2024). Using this best model, a presence probability map was created for CSTG in Lac-du-Bois. We further classified and identified the highly suitable habitats within our study area. Four potential habitat classes were grouped as follows: unsuitable (≤0.15), low quality (0.16 - 0.40), moderate quality (0.41 - 0.70) and high quality (≥0.71) (Figure 3).

Table . Model parameters used.

|  |  |
| --- | --- |
| **Parameter** | **Value(s)** |
| Number of Trees | 100, 250, 500, 750, 1000, 1500, 2000 |
| Mtry | 1:10 |
| Cross validation - inner folds | 10 |
| Cross validation - outer folds | 4 |

# Results

## Model Selection

After evaluating 700 models, nested within four outer spatial cross validation iterations, the best model was determined using three different metrics including classif.auc, classif.prauc, and classif.bbrier (Lang et al. 2019) (Table 2). These metrics are appropriate for two-class classification tasks with probability predictions. The model with the lowest classif.bbrier score was chosen as the best model, as this metric was best suited for our intended purpose of identifying the accuracy of probabilistic predictions (i.e., predicting the probability of CSTG presence on the landscape) (Brier 1950; Yang et al. 2022). The selected best-fit model had the following parameters: num.threads=1, importance=impurity, num.trees=100, and mtry=2.

Table . Aggregated results of each resampling iteration based on outer tuning for spatial cross validation. The highlighted value with the asterisk is the model we selected as best fit for the study.

|  |  |  |  |
| --- | --- | --- | --- |
| iteration | classif.auc | classif.bbrier | classif.prauc |
| 1 | 0.676955 | 0.278837 | 0.297317 |
| 2 | 0.663889 | 0.217228 \* | 0.508587 |
| 3 | 0.875 | 0.379851 | 0.992177 |
| 4 | 0.448864 | 0.395835 | 0.747301 |

## Model Assessment

To assess the best fit model, a variable importance table and a confusion matrix was produced (Table 3). The variable importance table depicts that topographic wetness index, multi-resolution ridge top flatness (MRRTF), negative topographic openness, and slope were the four main variables that were important in assessing the presence and absence of CSTG throughout the landscape. Overland flow distance, or distance from a stream channel, was the least important variable in predicting CSTG presence or absence.

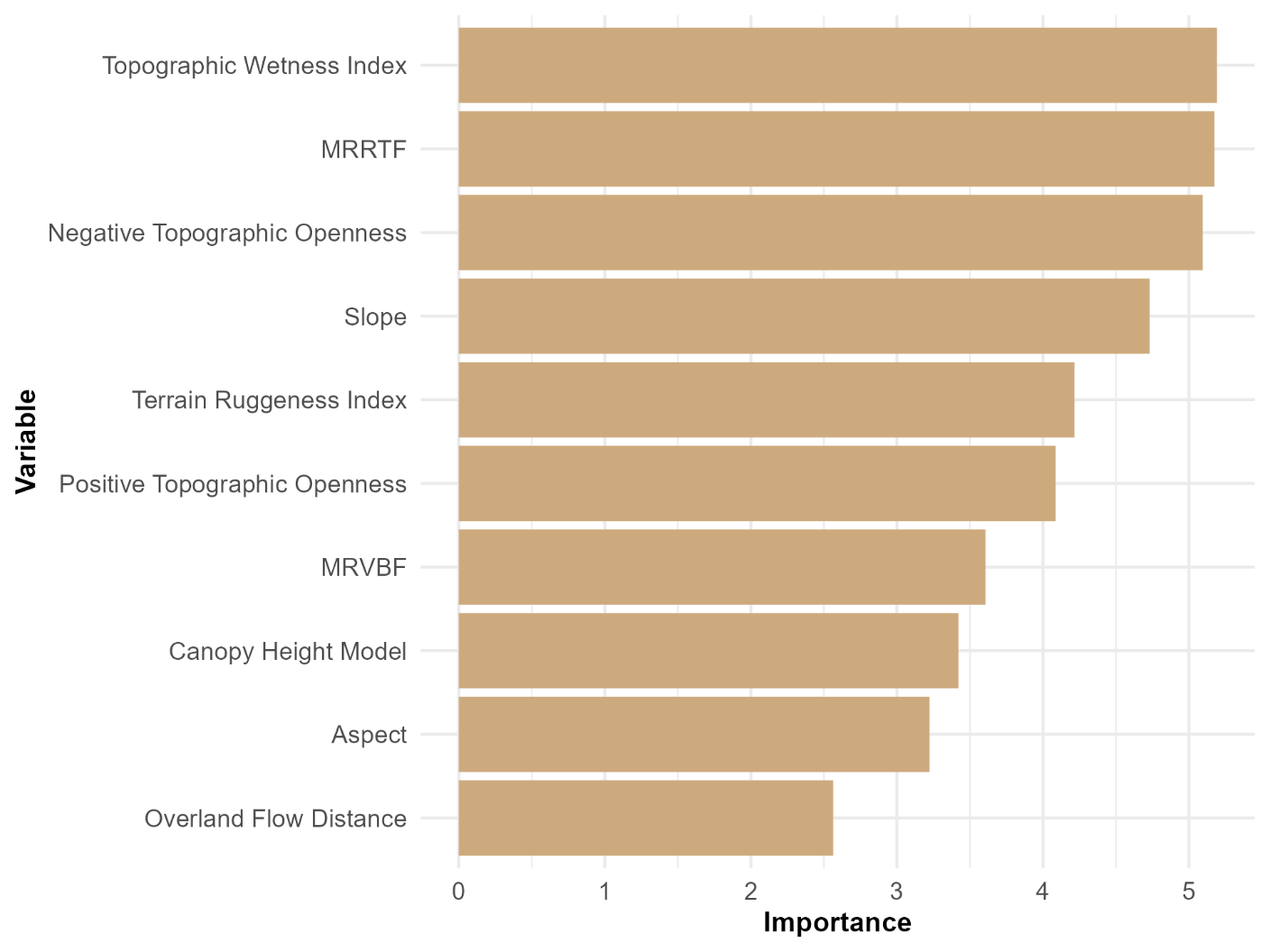


Figure . Variable importance table ranking the variables used within a ranger probability model to determine the presence or absence of Columbia Sharp-tailed Grouse in the Lac du Bois grasslands.

The confusion matrix depicts that the model had low accuracy in predicting CSTG presence within the Lac du Bois grasslands (Table 3). The model favoured designating areas as suitable CSTG habitat even if data showed no CSTG presence; however, there was still a considerable number of false predictions (i.e., not suitable CSTG habitat) where data suggests CSTG were present. This outcome is evident when evaluating the confusion matrix sensitivity, specificity, precision, and accuracy. The true positive rate was 55%, the true negative rate was 36%, the positive predictive value was 57%, and the overall accuracy of the model was 47%, respectively (Table 3).

Table . Confusion matrix for a ranger probability model used to determine the presence or absence of Sharp-tailed Grouse in the Lac du Bois grasslands.

|  |  |  |  |
| --- | --- | --- | --- |
|  | truth | FALSE | TRUE |
| response |  |  |  |
| FALSE |  | 21 | 42 |
| TRUE |  | 38 | 51 |

The following map was generated by applying the best fit model to each cell of a raster containing all the model layers (Figure 3). The probability of finding CSTG is categorized into four distinct bins that may translate to: unsuitable habitat (≤0.15), low quality habitat (0.16 - 0.40), moderate quality habitat (0.41 - 0.70) and high-quality habitat (≥0.71). Most of the unsuitable habitat depicted in the map encompasses areas with dense tree cover. The areas with high quality habitat are characterized by open grasslands and ridgetops. All of the high quality and moderate quality habitat can be found within the Lower Grasslands, dominated by Bluebunch wheatgrass and Big sage, and the Middle grasslands, dominated by Bluebunch wheatgrass and Rough fescue, both being within the Bunchgrass biogeoclimatic zone (Delesalle et al. 2009).

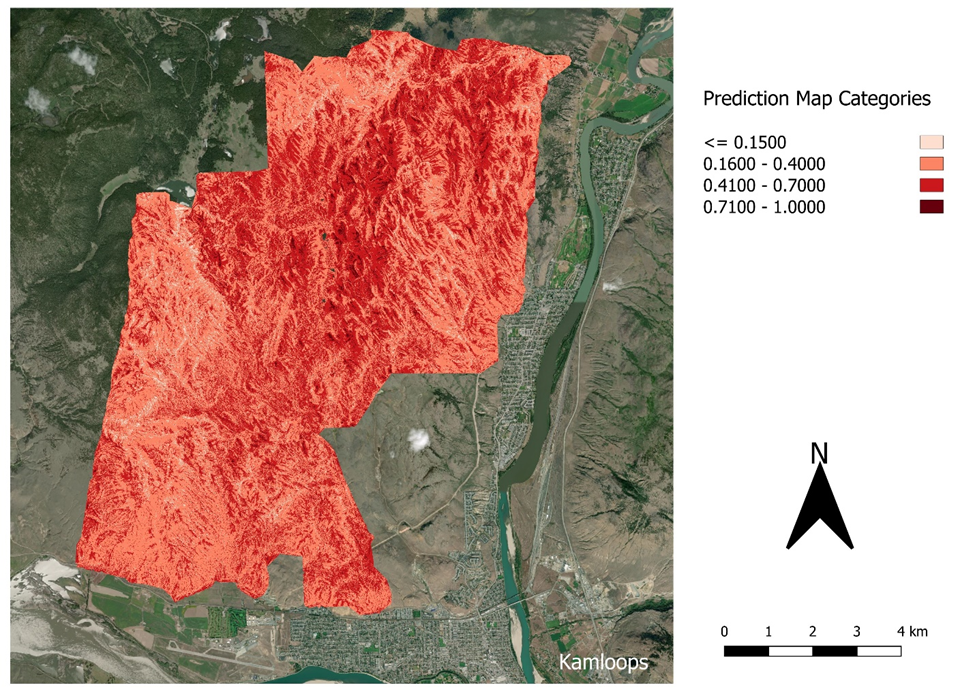


Figure . Prediction map of Sharp-tailed Grouse in the Lac du Bois grasslands

# Discussion

The predictive map suggests that the most suitable habitat for CSTG is on ridgetops and within open grasslands in Lac du Bois. This aligns well with the literature describing their breeding sites, termed leks.

“Openness is an important requirement of a dancing ground (lek) because it enables the detection of predators and in attracting grouse to the lek by seeing and/or hearing displaying males. Leks are often located on ridge tops or elevated ground but not necessarily the highest ground available.” (Ritcey & Jury 2004).

“Lek sites are typically located in small openings in shrubland habitats that have greater topographic relief relative to the local surrounding area” (Smith et al. 2016)

The high importance of the topographic wetness index (TWI), along with the MRRTF, may describe elevated sites and ridgetops, where water drains easily, thus making dry topography a significant factor in predicting CSTG habitat. TWI characterizes the relative wetness or moisture availability of a landscape based on its topography, with higher TWI values indicating greater potential for water accumulation and lower TWI values indicating drier areas with less water availability (Beven & Kirby 1979). Therefore, TWI can effectively identify both wet and dry areas across the landscape. In this case, the TWI can be used as a tool for identifying and delineating arid or drought-prone areas such as ridgetops. However, in contradiction to the model being a good predictor of arid ridgetops, the high importance of negative topographic openness (NTO) would describe concave features in the landscape. This contradiction may be one of the reasons for the poor accuracy of the model.

Specifically, the poor model accuracy is likely a result of poor precision or relevance in the model variables being used. The primary concern with the model is its tendency to produce a high rate of false positives, wherein it predicts CSTG presence despite no CSTG being observed in the data for these areas. This may be due to the large amount of rolling open grasslands within the study area that could be deemed as suitable habitat when the model predictions are mainly dictated by factors like TWI, MRRTF, and slope. There may be more specific requirements, such as distance to shrub cover or other features that provide protection from predators, that should be accounted for when building this model (Ritcey & Jury 2004). Furthermore, as there was still a considerable number of false negative predictions where CSTG were observed in the data, it is important to note that the variables used may describe lek sites well but may not describe CSTG habitat throughout the entire year (Ritcey 1990; Leupin & Murphy 2000). If year-round presence is to be assessed, and surveys were conducted at various times of the year, it would be necessary to incorporate more variables that pertain to non-breeding habitat as well.

The low accuracy of the model could also be attributed to the fact that Columbian sharp-tailed grouse depend on a variety of habitats throughout the year, and the surveys and observations, conducted at various times of year, were all pooled together. As noted above, during the breeding season, openness is an important characteristic, however during the nesting season, herbaceous cover of 15-30 cm is essential to conceal the nests from predators. Additionally, nests have been recorded at location as far as 1600m away from the lek sites. Regarding brood habitat preference over the summer months, ground insects seem to be an important food source for chicks; some studies in the US have shown a preference for shrub habitat, while others for grass and forb habitat. Data on brood habitat preference in British Columbia is limited. Finally riparian areas are important winter-feeding habitat (Ritcey & Jury, 2004). With such a range and complexity in habitat requirements, we recommend creating 5 different prediction models of habitat suitability to account for seasonal variations, divided into the 5 following seasons: breeding, nesting, brooding, fall, and winter.

Lastly, the Columbian sharp-tailed grouse is on the Blue List (species of special concern) in British Columbia (Leupin et al., 2003). This grouse has significantly declined since agriculture shifted to large-scale practices (Buss & Dziedzic, 1955). “In British Columbia, the Columbian sub-population might have declined as much as 70% since the early 1900s”(Leupin et al., 2003). Threats to the sharp-tailed grouse species include the fragmentation of their habitat through urbanization, agriculture development, heavy cattle grazing, and fire suppression (Ritcey & Jury, 2004). For these reasons, extensive and consistent monitoring of Columbian sharp-tailed grouse should be established in Lac Du Bois as well as throughout the province. Such approach would provide detection of further changes on their abundance and distribution, a better understanding of the subspecies, and facilitate the development of adaptive management strategies. Additionally, since habitat fragmentation seems to be a major threat, we recommend including a habitat connectivity layer to prediction models in future research.

In summary, our model predicted that the main drivers influencing the presence of Columbian sharp-tailed grouses in Lac du Bois were: topographic wetness index, multi-resolution ridge top flatness (MRRTF), negative topographic openness, and slope gradient. The overall accuracy of our model nearly reached 50%. This fairly low accuracy could be due to different factors such as the choice of predictor variables in the model as well as the range and complexity of the grouse’s habitat requirements throughout the year. Extensive long-term monitoring as well as more research and habitat modeling is essential to further investigate the ecology of Columbian sharp-tailed grouse and ensure adequate conservation and management of its habitat.

**Work Cited**

Allaire, JJ, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2023. *rmarkdown: Dynamic Documents for r*.<https://github.com/rstudio/rmarkdown>.

Appelhans, Tim, Florian Detsch, Christoph Reudenbach, and Stefan Woellauer. 2023. *mapview: Interactive Viewing of Spatial Data in r*. <https://CRAN.R-project.org/package=mapview>.

Beven, K.J., Kirkby, M.J. (1979). A physically-based variable contributing area model of basin hydrology. Hydrology Science Bulletin 24(1), p.43-69

BC Parks. (n.d.). *A Bird at Risk...* British Columbia Parks. Retrieved April 4, 2024, from <https://nrs.objectstore.gov.bc.ca/kuwyyf/sharp_tailed_grouse_sign_e58f4b2e94.pdf>

BC Parks. (2000). *Lac du Bois Grasslands Protected Area Management Plan*. Government of British Columbia. BC Parks. 2000. Lac du Bois Grasslands Park Management Plan Background Document

BC Parks. (2023). *Lac du Bois Grasslands Protected Area Management Plan*. BC Gov. <https://nrs.objectstore.gov.bc.ca/kuwyyf/lac_du_bois_grasslands_protected_area_draft_management_plan_16ffe34c93.pdf>

Becker, Marc, and Patrick Schratz. 2024. *Mlr3spatial: Support for Spatial Objects Within the “mlr3” Ecosystem*. <https://CRAN.R-project.org/package=mlr3spatial>.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>

Government of British Columbia. (2019). *BC Parks Map*. BC Gov. <https://catalogue.data.gov.bc.ca/dataset/bc-parks-map>

Government of British Columbia. (2024). *Wildlife Data & Information*. <https://www2.gov.bc.ca/gov/content/environment/plants-animals-ecosystems/wildlife/wildlife-data-information>

Brier GW, Brier, W. G. 1950. Verification of Forecasts Expressed in Terms of Probability. MWRv. 78(1):1. doi:10.1175/1520-0493(1950)078. <https://ui.adsabs.harvard.edu/abs/1950MWRv...78....1B/abstract>.

Buss, I. O., & Dziedzic, E. S. (1955). Relation of Cultivation to the Disappearance of the Columbian Sharp-Tailed Grouse from Southeastern Washington. *The Condor*, *57*(3), 185–187. <https://doi.org/10.2307/1364866>

Delesalle BP, Coupe BJ, Wikeem BM, Wikeem SJ. 2009. Grasslands Monitoring Manual for British Columbia: Tool for Ranchers. Grasslands Conservation Council of British Columbia.

Hijmans, Robert J. 2023. *terra: Spatial Data Analysis*. [https://CRAN.R-project.org/package=terra](https://cran.r-project.org/package=terra).

Lang, Michel, Martin Binder, Jakob Richter, Patrick Schratz, Florian Pfisterer, Stefan Coors, Quay Au, Giuseppe Casalicchio, Lars Kotthoff, and Bernd Bischl. 2019. “mlr3: A Modern Object-Oriented Machine Learning Framework in R.” *Journal of Open Source Software*, December. <https://doi.org/10.21105/joss.01903>.

Leupin, E. E., & Chutter, M. J. (2007). *Status of Sharp-tailed Grouse, columbianus subspecies (Tympanuchus phasianellus columbianus) in British Columbia*. B.C. Ministry of Environment, Victoria, BC. <https://a100.gov.bc.ca/pub/eirs/finishDownloadDocument.do;jsessionid=DA5B9402084204A93BB0B62E9D421365?subdocumentId=7723>

Leupin, E. and M. Murphy. 2000a. 1998–2000 Sharp-tailed Grouse lek inventory project. Unpubl. Rep. prepared for Min. Environ., Lands and Parks, Kamloops, B.C.

Leupin, E. E. (Ernest E., British Columbia. Biodiversity Branch., & British Columbia. Conservation Data Centre. (2003). *Status of the sharp-tailed grouse (Tympanuchus phasianellus) in British Columbia*. B.C. Ministry of Water, Land, and Air Protection, Biodiversity Branch.

Pawley, Steven. 2023. *Rsagacmd: Linking r with the Open-Source “SAGA-GIS” Software*. [https://CRAN.R-project.org/package=Rsagacmd](https://cran.r-project.org/package=Rsagacmd).

Pebesma, Edzer. 2018. “Simple Features for R: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46.<https://doi.org/10.32614/RJ-2018-009>.

Pebesma, Edzer, and Roger Bivand. 2023. *Spatial Data Science: With applications in R*. Chapman and Hall/CRC.<https://doi.org/10.1201/9780429459016>.

Posit team. 2024. *RStudio: Integrated Development Environment for r*. Boston, MA: Posit Software, PBC.<http://www.posit.co/>.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Ritcey, R.W. 1990. Report of a survey of wintering Sharp-tailed Grouse in forested habitats near Williams Lake, B.C. Feb./Mar. 90. Report to Oreg. Nature Conservancy. Unpubl. 10 p.

Ritcey RW, Jury D. 2004. Columbian Sharp-Tailed Grouse: Tympanuchus phasianellus columbianus Original prepared by Species Information Taxonomy. Accounts Meas Manag Identified Wildl.:1–9.

Roussel, Jean-Romain, and David Auty. 2024. *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*.<https://cran.r-project.org/package=lidR>.

Roussel, Jean-Romain, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R. H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian de Boissieu, and Alexis Achim. 2020. “lidR: An r Package for Analysis of Airborne Laser Scanning (ALS) Data.” *Remote Sensing of Environment* 251: 112061.<https://doi.org/10.1016/j.rse.2020.112061>.

Schratz, Patrick, and Marc Becker. 2024. *Mlr3spatiotempcv: Spatiotemporal Resampling Methods for “mlr3”*. <https://CRAN.R-project.org/package=mlr3spatiotempcv>.

Smith KT, Beck JL, Mong TW, Blomquist FC. 2016. Identification of Columbian Sharp-tailed Grouse lek sites in south central Wyoming. West North Am Nat. 76(1):135–141.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686.<https://doi.org/10.21105/joss.01686>.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC.<https://bookdown.org/yihui/rmarkdown>.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>.

Yang W, Jiang J, Schnellinger EM, Kimmel SE, Guo W. 2022. Modified Brier score for evaluating prediction accuracy for binary outcomes. Stat Methods Med Res. 31(12):2287. doi:10.1177/09622802221122391.

# Appendix A - R Script

---

title: "Lac du Bois STG Habitat Analysis"

format: html

editor: visual

---

### Tile LDB LiDAR data

```{r}

library(lidR)

library(future)

library(Rsagacmd)

library(sf)

library(tidyverse)

library(terra)

library(mapview)

library(mlr3)

library(mlr3verse)

library(mlr3spatial)

library(mlr3spatiotempcv)

library(sf)

library(terra, exclude = "resample")

library(future)

library(dplyr)

ldb\_liDATA <- readLAScatalog("LacDuBois.las")

plot(ldb\_liDATA, mapview = TRUE)

opt\_output\_files(ldb\_liDATA) <- "01\_retile/{XLEFT}\_{YBOTTOM}"

opt\_chunk\_size(ldb\_liDATA) <- 500

#should not use a buffer when tiling, changed buffer to 0 (was 15)

opt\_chunk\_buffer(ldb\_liDATA) <- 0

opt\_chunk\_alignment(ldb\_liDATA) <- c(500, 500)

plot(ldb\_liDATA, chunk\_pattern = TRUE)

ldb\_tiled <- catalog\_retile(ldb\_liDATA)

plot(ldb\_tiled, mapview = TRUE)

```

### Perform ground point classification

```{r}

n\_cores <- availableCores() / 2

plan(multisession, workers = n\_cores)

opt\_output\_files(ldb\_tiled) <- "02\_ground/{XLEFT}\_{YBOTTOM}"

# here is where we want to add the buffer (only once we want to run other functions than tiling)

opt\_chunk\_buffer(ldb\_tiled)<-15

ldb\_ground <- classify\_ground(ldb\_tiled, algorithm = csf(sloop\_smooth = TRUE))

```

### Create DEM for Lac du Bois from LiDAR data

```{r}

ldb\_ground <- readLAScatalog("02\_ground", chunk\_buffer = 15)

ldb\_norm

opt\_output\_files(ldb\_ground) <- ""

dem <- rasterize\_terrain(ldb\_ground, res = 1, algorithm = tin())

dir.create("ta", showWarnings = FALSE)

dem <- writeRaster(dem, "ta/dem.tif", overwrite = TRUE)

```

### Produce Various Terrain-Based Layers from the DEM

```{r}

#locate where our saga program and provide proper path

saga\_path <- "C:/SAGA-GIS/saga-9.3.2\_x64/saga\_cmd.exe"

# Create an object (saga) which is a list-like object that contains all of the SAGA GIS functions.

saga <- saga\_gis(saga\_path, raster\_format = "GeoTIFF")

# 1- sink filled DEM:

dem\_preproc <- saga$ta\_preprocessor$sink\_removal(

dem = dem, dem\_preproc = "ta/dem\_preproc.tif")

sources(dem\_preproc)

# 2- Produce DSM

# I added a DSM, but we don't have to include it in model

# just use CHM in model instead

opt\_output\_files(ldb\_ground) <- ""

dsm <- rasterize\_canopy(ldb\_ground, res = 1, algorithm = dsmtin())

dsm <- writeRaster(dsm, "ta/dsm.tif", overwrite = TRUE)

plot(dsm)

plot\_dtm3d(dsm) # to see it in 3D

# 3- Produce CHM

# Normalize heights:

opt\_chunk\_size(ldb\_ground) <- 0

opt\_output\_files(ldb\_ground) <- "03\_norm/{\*}"

ldb\_norm <- normalize\_height(ldb\_ground, algorithm = tin())

# Create CHM

ldb\_norm <- readLAScatalog("03\_norm", chunk\_buffer = 15)

opt\_output\_files(ldb\_norm) <- ""

chm <- rasterize\_canopy(ldb\_norm, res = 1, algorithm = dsmtin())

chm <- writeRaster(chm, "ta/chm.tif", overwrite = TRUE)

plot(chm)

plot\_dtm3d(chm) # to see it in 3D

####### Morphometry layers

# 4- produce slope and aspect layer:

# If want to take a look at all pieces of this function:

View(tidy(saga$ta\_morphometry$slope\_aspect\_curvature))

slope\_aspect <- saga$ta\_morphometry$slope\_aspect\_curvature(

elevation = dem\_preproc, slope = "ta/slope.tif", aspect = "ta/aspect.tif",

method = 6, unit\_slope = "radians", unit\_aspect = "radians",

.all\_outputs = FALSE)

# 5- MRVBF/MRRTF

# (Multi-Resolution Valley Bottom Flatness/Multi-Resolution Ridge Top Flatness)

mrvbf\_thresh <- mrvbf\_threshold(res = res(dem)[1])

mrvbf <- saga$ta\_morphometry$multiresolution\_index\_of\_valley\_bottom\_flatness\_mrvbf(

dem = dem\_preproc, mrvbf = "ta/mrvbf.tif", mrrtf = "ta/mrrtf.tif",

t\_slope = mrvbf\_thresh)

# 6- Terrain ruggedness index:

tri <- saga$ta\_morphometry$terrain\_ruggedness\_index\_tri(

dem = dem\_preproc, tri = "ta/tri.tif")

####### Hydrology layers

### Following 4 lines (tca, sca, so, and cn) needed to create hydrology layers

### but not needed for modeling

# Total catchment area

tca <- saga$ta\_hydrology$flow\_accumulation\_top\_down(

elevation = dem\_preproc, flow = "ta/tca\_TEMP.tif", .all\_outputs = FALSE)

# Specific catchment area

sca <- saga$ta\_hydrology$flow\_width\_and\_specific\_catchment\_area(

dem = dem\_preproc, tca = tca, sca = "ta/sca\_TEMP.tif", .all\_outputs = FALSE)

# Strahler stream order

so <- saga$ta\_channels$strahler\_order(

dem = dem\_preproc, strahler = "ta/strahler\_TEMP.tif", .all\_outputs = FALSE)

# Channel network

cn <- saga$ta\_channels$channel\_network(

elevation = dem\_preproc, init\_grid = so, init\_value = 5,

chnlntwrk = "ta/cn\_TEMP.tif", .all\_outputs = FALSE)

# 7- Topographic wetness index

twi <- saga$ta\_hydrology$topographic\_wetness\_index(

slope = slope\_aspect$slope, area = sca, twi = "ta/twi.tif")

# 8 - Overland flow distanc

overland\_flow <- saga$ta\_channels$overland\_flow\_distance\_to\_channel\_network(

elevation = dem\_preproc, channels = cn, distance = "ta/o\_flow.tif",

disthorz = "ta/h\_flow.tif", distvert = "ta/v\_flow.tif", boundary = FALSE,

.all\_outputs = FALSE)

####### Lighting/visibility layers

# 9-Topographic openness

openness <- saga$ta\_lighting$topographic\_openness(

dem = dem\_preproc, pos = "ta/openness\_pos.tif", neg = "ta/openness\_neg.tif")

openness

```

### Extract the raster data where the points are located

```{r}

aspect <- rast("aspect.tif")

slope <- rast("slope.tif")

mrvbf <- rast("mrvbf.tif")

mrrtf <- rast ("mrrtf.tif")

tri <- rast("tri.tif")

twi <- rast ("twi.tif")

overlandflow\_distance <- rast ("o\_flow.tif")

openness\_p <- rast("openness\_pos.tif")

openness\_n <- rast("openness\_neg.tif")

chm <- rast("chm.tif")

#Reran Coopers part here just with new variables

model\_layers <- c(slope, aspect, mrvbf, mrrtf, tri, twi, overlandflow\_distance, openness\_p, openness\_n, chm)

# Convert .gpkg to a SpatVector object. Lecture 5.

sf\_grouse <- st\_read("Sharp-tailed grouse.gpkg", stringsAsFactors = TRUE)

grouse <- vect(sf\_grouse)

mapview(grouse)

# Perform data extraction at each point in each raster layer

ldb\_data <- terra::extract(model\_layers, grouse, bind = TRUE, na.rm = TRUE) %>%

st\_as\_sf()

mapview(ldb\_data, zcol = "presence")

ldb\_data <- na.omit(ldb\_data)

```

### Create a tuned ranger classification probability model using the mlr3 package, employing spatial cross validation in order to eliminate the effects of spatial autocorrelation.

```{r}

#creating task function

str(ldb\_data)

ldb\_data$presence <- factor(ldb\_data$presence)

tsk\_grouse <- as\_task\_classif\_st(ldb\_data, target = "presence")

tsk\_grouse

#creating probability learner

lrn\_rf\_tune\_prob <- lrn("classif.ranger",

num.trees = to\_tune(100, 2000), predict\_type = "prob",

mtry = to\_tune(1, length(tsk\_grouse$feature\_names)),

importance = "impurity")

avail\_msrs <- as.data.table(msrs())[

task\_type == "classif" &

predict\_type == "prob" &

task\_properties == "twoclass"]

View(avail\_msrs)

useful\_msrs <- c("classif.auc", "classif.bbrier", "classif.prauc")

#Tuner design

df\_design <- expand.grid(

num.trees = c(100, 250, 500, 750, 1000, 1500, 2000),

mtry = 1:length(tsk\_grouse$feature\_names))

# Convert to a data.table (same structure as df\_design, different object type)

dt\_design <- data.table(df\_design)

# Create the tuner object

tnr\_design <- tnr("design\_points", design = dt\_design)

#Creating cross validation resampling objects

cv\_inner <- rsmp("cv", folds = 10)

cv\_outer <- rsmp("spcv\_coords", folds = 4)

#Creating auto tuner object

at\_prob\_brier <- auto\_tuner(

tuner = tnr\_design,

learner = lrn\_rf\_tune\_prob,

resampling = cv\_inner,

measure = msr("classif.bbrier"),

terminator = trm("none")

)

outer\_cores <- min(4, availableCores())

inner\_cores <- floor(availableCores() / outer\_cores)

plan(list(

tweak("multisession", workers = outer\_cores),

tweak("multisession", workers = inner\_cores)

))

#resampling - this takes a long time because i did not set up parallel env.

rr\_prob\_brier <- resample(tsk\_grouse, at\_prob\_brier, cv\_outer, store\_models = TRUE)

rr\_prob\_brier$aggregate(msrs(useful\_msrs))

conf\_prob\_brier <- rr\_prob\_brier$prediction()$confusion

conf\_prob\_brier

rr\_prob\_results\_brier <- extract\_inner\_tuning\_results(rr\_prob\_brier)

mod\_scores\_prob\_brier <- rr\_prob\_brier$score(msrs(useful\_msrs))

View(mod\_scores\_prob\_brier)

mod\_scores\_brier\_df <- mod\_scores\_prob\_brier %>%

select(c(7, 9, 10, 11))

write.csv(mod\_scores\_brier\_df, "mod\_scores\_brier\_df.csv")

#best model

best\_lrn\_prob\_brier <- rr\_prob\_brier$learners[[which.min(mod\_scores\_prob\_brier$classif.bbrier)]]$learner

best\_lrn\_prob\_brier

imp <- best\_lrn\_prob\_brier$importance()

imp <- data.frame(Variable = factor(names(imp), levels = rev(unique(names(imp)))),

Importance = imp, row.names = NULL)

imp

imp\_plot\_prob <- ggplot(imp, aes(x = Importance, y = Variable)) +

geom\_bar(stat = "identity", fill = "burlywood3") +

labs(x = "Importance", y = "Variable") +

theme\_minimal() +

theme(axis.text = element\_text(size = 9),

axis.title = element\_text(size = 10, face = "bold"),

plot.title = element\_text(size = 11, face = "bold", hjust = 0.5))+

scale\_y\_discrete(labels = c(Z = "Canopy Height Model", mrvbf = "MRVBF", openness\_neg = "Negative Topographic Openness", slope = "Slope", aspect = "Aspect", tri = "Terrain Ruggeness Index", twi = "Topographic Wetness Index", openness\_pos = "Positive Topographic Openness", mrrtf = "MRRTF", o\_flow = "Overland Flow Distance"))

imp\_plot\_prob

# Write variable importance to .csv file:

write.ftable(ftable(conf\_prob\_brier), file = "confusion\_matrix\_prob\_brier.csv", sep = ",",

quote = FALSE)

# Save that plot to a file (for use in reporting)

ggsave("Variable importance\_prob\_brier.png", imp\_plot\_prob, width = 1920, height = 1440,

units = "px", dpi = 300)

#The final plot!!!

ranger\_model\_brier <- best\_lrn\_prob\_brier$model

fun <- function(model, ...) predict(model, ...)$predictions

prediction\_prob\_terra\_brier <- terra::predict(

model\_layers, ranger\_model\_brier, fun = fun, na.rm = TRUE)

prediction\_prob\_present\_brier <- prediction\_prob\_terra\_brier$TRUE.

plot(prediction\_prob\_terra\_brier)

writeRaster(prediction\_prob\_present\_brier, "prediction\_brier.tif", overwrite = TRUE)

```

```{r}

#Bibliography

library(grateful)

cite\_packages(output = "file", citation.style = NULL, out.dir = getwd(), out.format = "docx", omit = "grateful", cite.tidyverse = TRUE, include.RStudio = TRUE, out.file = "Citations")

```