Expansion of Canadian Agriculutre and the Relationship with Wetlands in Boreal Shield Provinces

Simranjeet Bilkhu (400611361)

Dataset Sourcing

The data used in this exploration is available through the Open Government Portal. Here are the following filtering options used to source the data:

- **Portal Type:** Open Data
- **Resource Type:** Dataset
- **Format:** GeoTIF
- **Update Frequency:** Annually

The data is comprised of digital crop type maps across several provinces. Crop type maps were reconstructed each year from 2009 to 2024. However, this paper will discuss crop type maps that were constructed in 2013-2023 for the following provinces: Alberta, Manitoba, Newfoundland and Labrador, Ontario, Quebec and Saskatchewan. The provinces selected contain the Boreal shield which is of ecological significance ("Canadian Biodiversity: Ecozones: Boreal Shield" n.d.).

Application Description

Wetlands play an integral role in maintaining global environmental stability ("Wetlands 101" n.d.), and may play a key role in fighting climate change (Pope 2017). However, they also serve as a key resource for agriculture since they provide water for crops, livestock and aquaculture ("Wetlands and Agriculture: Impacts of Farming Practices and Pathways to Sustainability" n.d.). This paper will examine the relationship between land usage in agriculture and the degradation of wetlands in selected Canadian provinces. Specifically, this paper will examine land usage data collected between every year between 2013 to 2023 for Boreal Shield provinces which house approximately 25 percent of the world's wetlands ("Canadian Biodiversity: Ecozones: Boreal Shield" (n.d.)).

Data Transformation and Pre-processing

The data came in the form of 66 separate GeoTIF files, one file for each province and year. Here are the steps used in pre-processing:

For each GeoTIF file:

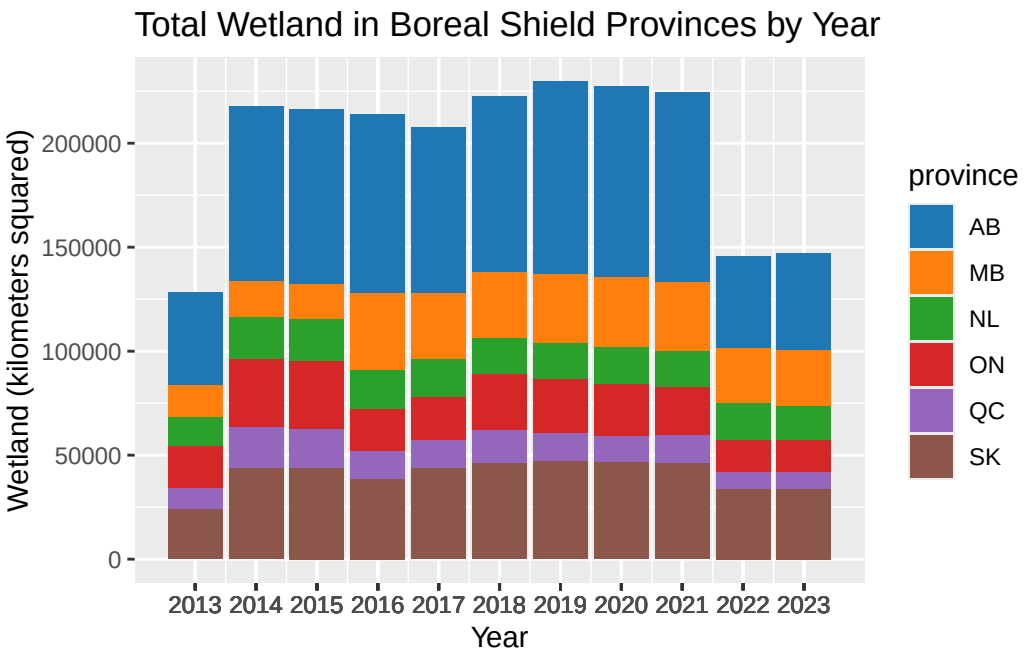
1. Convert the GeoTIF files into SpatRaster files using the R terra library (Hijmans 2024).
2. Count the frequency of codes corresponding to agriculture as well as the frequency for the codes corresponding to wetland.
3. Take the sum of the frequency of agriculture codes to get an estimate of the amount of land used for agriculture in a given province and a given year.
4. Take the sum of the frequency of wetland codes to get an estimate of the amount of wetland in a given year and province.
5. Store the estimates in a tidy dataframe (Wickham, Vaughan, and Girlich 2024).
6. Scale the estimate of the amount of wetland to square kilometers using dplyr (Wickham et al. 2023).
7. Scale the estimate of the amount of land used for agriculture to get an estimate of the amount of wetland.

No missing data imputation methods were used in the analysis (missing data was ignored). This may be problematic since certain geographic conditions might effect satellite imagery which might make certain land types less detectable than others through satellite imagery. Thus the data may not be missing at random. Unfortunately, I'm not familiar with missing data imputation methods, so I will be ignoring missing values in this analysis. Fortunately though, there were no outliers found in the exploration.

Single Variable Analysis

Question: Is there a downward trend in the amount of wetlands in the Boreal shield provinces from 2013 to 2023?

The goal of the single variable analysis is to determine whether or not the amount of wetlands in the Boreal shield provinces as decreased over from 2013 to 2023. Below is a stacked barplot showing the amount of wetlands in the Boreal shield provinces per year.



It appears that there might not be a downward trend in the amount of total wetland in Boreal Shield provinces from 2013 to 2023. Nevertheless, there seems to be below average total wetland in 2013, 2022 and 2023. Further research may need to be gathered to determine whether there were any changes in wetland conservation policies that may have impacted the amounts of wetland in the Boreal shield throughout those years.

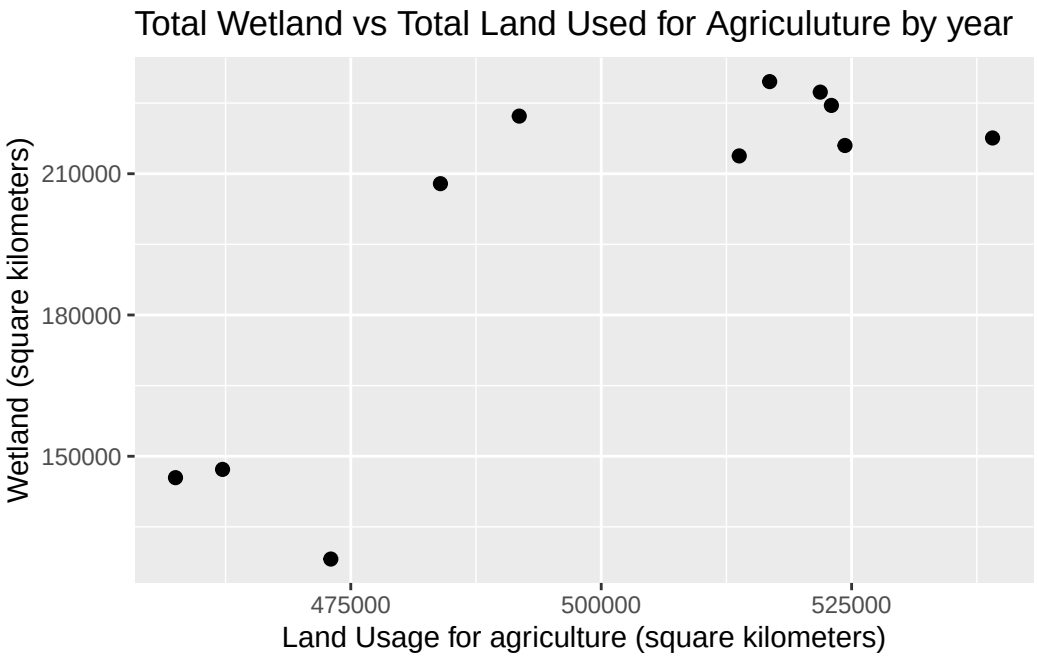
Interestingly, it appears that the the highest amounts of total wetland in the Boreal Shield provinces were seen in 2019, 2020 and 2021. It may be possible that these years were associated with higher amounts of rainfall compared to the others; further research is needed.

It's interesting to see that even though that 2013 appears to be the year with the lowest amounts of wetland, 2022 and 2023 appear to be the years that Quebec and Ontario had the lowest amounts of wetland between 2013 and 2023.

Multivariable Analysis

Question: Is there a relationship between land use for agriculture and wetland in the Boreal Shield?

To investigate this question, I made a scatter plot between the the amount of land used for agriculture and the amount of wetland per year in the Boreal Shield provinces.



Rather unexpectedly, it appears that, if anything, that higher amounts of wetland tend to be associated with higher land usages of agriculture. However, there are important external variables that have not been considered in this analysis such as rainfall, artificial wetland construction, average temperature and so on.

Regardless, it might very well be the case that higher amounts of wetland facilitate agriculture as wetland may serve as a crucial water source for agriculture ("Wetlands and Agriculture: Impacts of Farming Practices and Pathways to Sustainability" n.d.).

Ultimately, the findings of this exploration should be validated with a more rigorous analysis as it contradicts other findings that associate agricultural development with wetland degradation such as those found in Ballut-Dajud et al. (2022).

Webpage

For more exploration, an interactive page is available here which was made using the shiny R package (Chang et al. 2024).

The site allows users to further explore the amount of wetland over the years for each individual province. This may provide insight as to if there are any specific provinces that wherein wetland degradation is prevalent. The site also allows users to do a multivariable exploration to examine the relationship between land area of wetlands and land usage in agriculture for each individual province. This may inform users on whether or not there are certain provinces whose agricultural practices are more harmful to wetlands than others. Finally, the site provides users the opportunity to examine the trends in agriculture land usage for each province which can give an idea on whether or not wetlands may be impacted more in the future due to a growing agricultural industry.

To use the site, two drop downs are available. The first drop down provides the user a list of provinces to explore. The options to choose from are: Alberta, Manitoba, Newfoundland and Labrador, Ontario, Quebec, Saskatchewan and all (for all of the above). The second drop down allows the user to select which variable(s) to explore and the options include: agriculture, wetland and both. If agriculture or wetland are chosen then a bar plot will be constructed showing the amounts of agricultural land or wetland for each year between 2013 and 2023. If both are chosen then a scatter plot will be created plotting amounts of agricultural land against amounts of wetland.

References

Ballut-Dajud, Gastón Antonio, Luis Carlos Sandoval Herazo, Gregorio Fernández-Lambert, José Luis Marín-Muñiz, María Cristina López Méndez, and Erick Arturo Betanzo-Torres. 2022. "Factors Affecting Wetland Loss: A Review." *Land* 11 (3): 434.

"Canadian Biodiversity: Ecozones: Boreal Shield." n.d. *Canadian Biodiversity*. Accessed September 25, 2024. https://canadianbiodiversity.mcgill.ca/english/ecozones/borealshield/borealshield.htm.

Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2024. *Shiny: Web Application Framework for r*. https://CRAN.R-project.org/package=shiny.

Hijmans, Robert J. 2024. *Terra: Spatial Data Analysis*. https://CRAN.R-project.org/package=terra.

Pope, Alexandra. 2017. "Canada's Boreal Wetlands Are Key to Fighting Climate Change: Report." *Canadian Geographic*. https://canadiangeographic.ca/articles/canadas-boreal-wetlands-are-key-to-fighting-climate-change-report/.

"Wetlands 101." n.d. *Nature Conservancy Canada*. Accessed September 25, 2024. https://www.natureconservancy.ca/en/what-we-do/resource-centre/conservation-101/../../wetlands-101.html.

"Wetlands and Agriculture: Impacts of Farming Practices and Pathways to Sustainability." n.d. Accessed September 25, 2024. https://www.ramsar.org/sites/default/files/documents/library/bn13_agriculture_e.pdf.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Supplementary Material

Preprocessing

```
library(terra)
library(dplyr)
library(stringr)
library(here)
library(ggplot2)
library(tidyr)
options(scipen = 999) # don't use scientific notation
help(viridis)

# Loading in Data ------------------------------------------------------------

# Load multiple GeoTIFF files
folder_path <- here('Extracted')


# List all the tiff files in the folder

file_names <- list.files(
  folder_path,
  pattern = "\\.tif$",
  full.names = TRUE)

head(file_names, 3)

data_names <- str_extract(basename(file_names), "^[^.]+")

# Shortening datafile names to year_province
shorthand_names = rep(NA, length(data_names))
for (i in 1:length(data_names)) {
  shorthand_names[i] = substr(data_names[i], 5, 11)
}

# creating a frame with the shorthand names and filenames
df <- data.frame(
  year_province = shorthand_names,
  file_name = file_names)

# Create an empty dataframe that will store counts for agriculture and wetlands
data = tibble(
  year = c(rep(2013, 6), rep(2014, 6), rep(2015, 6), rep(2016, 6), rep(2017, 6),
   ↪  rep(2018, 6),
           rep(2019, 6), rep(2020, 6), rep(2021, 6), rep(2022, 6), rep(2023, 6)),
  province = rep(c('AB', 'MB', 'NL', 'ON', 'QC', 'SK'), 11),
  name = df$year_province,
  agriculture = rep(NA, 66),
  wetlands = rep(NA, 66)
)

# read each file one by one and store the counts
for (i in 1:66) {
  src = rast(df$file_name[[i]])
```

```
  # make a frequency table of the codes
  f = freq(src)
  # pull counts for agriculture
  agriculture = f |>
    filter(value %in% c(120, 122, 131:191, 193:199)) |>
    pull(count) |>
    sum()
  # pull counts for agriculture
  wetland = f |>
    filter(value == 80) |>
    pull(count) |>
    sum()
  data$agriculture[i] = agriculture
  data$wetlands[i] = wetland
  print(i/66)
  print(paste(data$year[i], data$province[i]))
}
# scale the data so that units are in square kilometers (original resolution is 30m x
 ↪  30m)
data_scaled = data |> mutate(
  agriculture = agriculture*(900/1000000),
  wetlands = wetlands*(900/1000000)
)
# save the dataframe into a csv
write.csv(data_scaled, file = 'land_use.csv')
```

Code for report:

```
library(dplyr)
library(ggplot2)
data = read.csv('land_use.csv')
options(scipen = 999) # Don't use scientific notation
```

```
ggplot(data, mapping = aes(x = year, y = wetlands, fill = province)) +
  geom_bar(stat = 'identity') +
  scale_x_continuous(breaks = data$year) +
  labs(title = 'Total Wetland in Boreal Shield Provinces by Year',
       x = 'Year',
       y = 'Wetland (km sq)') +
  scale_fill_manual(values = c("#1f77b4", "#ff7f0e", "#2ca02c", "#d62728",
                               "#9467bd", "#8c564b", "#e377c2", "#7f7f7f",
                               "#bcbd22", "#17becf", "#aec7e8"))
```

```
data_by_year = data |> group_by(year) |>
  summarise(
    total_agriculture = sum(agriculture),
    total_wetland = sum(wetlands))

ggplot(data = data_by_year, mapping = aes(x = total_agriculture, y = total_wetland)) +
  geom_point(size = 2) +
  labs(title = "Total Wetland vs Total Land Used for Agriculuture by year",
       x = "Land Usage for agriculture (square kilometers)",
```

```
       y = "Wetland (square kilometers)")
```

Code for Shiny site:

```
library(shiny)
library(ggplot2)
library(dplyr)
library(shinythemes)
options(scipen = 999)

# this will be in my ui:
graphs = div(
  titlePanel("Visualize Wetland Area by Year, Agriculture Area by Year and more"),
  sidebarLayout(
    sidebarPanel(
      # input to choose which province to display
      selectInput(inputId = "prov", label = 'Province',
                  choices = list("Alberta" = "AB", "Manitoba" = "MB", "Newfoundland and
                   ↪  Labrador" = "NL",
                              "Ontario" = "ON", "Quebec" = "QC", "Saskachewan" = "SK",
                               ↪  "All" = "all")),
      # input to choose which variable is chosen (or do a comparison for both variables)
      selectInput(inputId = "var_chosen", label = "Choose the variable you want to
       ↪  investigate (choose both if you want to compare wetland vs agriculture)",
                  choices = list("Agriculture" = "agriculture", "Wetland" = "wetlands",
                   ↪  "Both" = "both"))

    ),
    mainPanel(
    conditionalPanel(
      condition = "input.var_chosen == 'both'",
      plotOutput(outputId = "multivar")
    ),
    conditionalPanel(
      condition = "input.var_chosen != 'both'",
      plotOutput(outputId = "univar")
    )
    )
  )
)

# have a small information panel giving a link to the data
about = div(
  titlePanel("Information"),
  mainPanel(
    textOutput(outputId = 'txt')
  )
)

# Load in the data set
DATASET = tibble(read.csv("land_use.csv", header = TRUE))

# Dictionary for province abreviations
```

```
PROVINCE_CODES = list('AB' = 'Alberta', 'MB' = 'Manitoba', 'NL' = 'Newfoundland and
 ↪  Labrador',
                      'ON' = 'Ontario', 'QC' = 'Quebec', 'SK' = 'Saskachewan')

# dictionary for province color codes
PROVINCE_COLOUR_CODES = list('AB' = '#1f77b4', 'MB' = '#ff7f0e', 'NL' = '#2ca02c',
                             'ON' = '#d62728', 'QC' = '#9467bd', 'SK' = '#8c564b')

ui = navbarPage(
  # I like cerulean
  theme = shinytheme("cerulean"),
  title = "Expansion of Canadian Agriculutre and the Relationship with Wetlands in Boreal
   ↪  Shield Provinces",
  tabPanel("Data Visualization", graphs),
  tabPanel("Info", about),
  id = "navbar"
)

server = function(input, output) {
  # store the inputs into reactive variables to be accessed later
  province_chosen = reactive({
    input$prov
  })
  variable_chosen = reactive({
    input$var_chosen
  })

  data = reactive({
    # Summarize the data set if the user just wants to show an overall scatterplot
    if (input$prov == 'all' && input$var_chosen == 'both') {
      df = DATASET |>
        group_by(year) |>
        summarise(agriculture = sum(agriculture), wetlands = sum(wetlands))
    # Filter by province of the user's choosing
    } else if (input$prov != 'all') {
      df = DATASET |>
        filter(
          province == input$prov
        )
    } else {
    # dataframe should stay the same for the stacked barplot
      df = DATASET
    }

    df
  })

  output$univar = renderPlot({
    # Verify that the plot does not render if 'both' is chosen
    if (variable_chosen() != 'both'){
      # If all provinces are chosen for univariate analysis make a stacked barplot
      if (province_chosen() == 'all'){
        ggplot(data = data(), mapping = aes(x = factor(year),
```

```
          y = eval(parse(text = paste0(variable_chosen()))), fill = province)
          ) + geom_bar(stat = 'identity') +
          labs(
            title = paste("Amount of land classified as", variable_chosen(),
                          "by year for all provinces"),
            x = "Year",
            y = paste(variable_chosen(), '(in square kilometers)')
          ) +
          scale_fill_manual(values = c("#1f77b4", "#ff7f0e", "#2ca02c", "#d62728",
                                       "#9467bd", "#8c564b"))

      } else if (province_chosen() != 'all') {
        # Make a barplot
        ggplot(data = data(), mapping = aes(x = factor(year),
        y = eval(parse(text = variable_chosen())))) +
        geom_bar(stat = 'identity', fill = PROVINCE_COLOUR_CODES[province_chosen()]) +
        labs(
          title = paste("Amount of land classified as", variable_chosen(),
                        "by year for", PROVINCE_CODES[province_chosen()]),
          x = "Year",
          y = paste(variable_chosen(), '(in square kilometers)')
        )
      }
    }
  })
  output$multivar = renderPlot({
    # Make a scatter plot (the title changes depending on province_chosen())
    if (province_chosen() == 'all') {
      ggplot(data = data(), mapping = aes(x = agriculture, y = wetlands)) +
        geom_point(size = 3) +
        labs(title = "Wetland vs Land Used for Agriculuture in Boreal Shield Provinces",
           x = "Land Usage for agriculture (square kilometers)",
           y = "Wetland (square kilometers)")
    } else {
      ggplot(data = data(), mapping = aes(x = agriculture, y = wetlands)) +
        geom_point(size = 3, col = PROVINCE_COLOUR_CODES[province_chosen()]) +
        labs(title = paste("Wetland vs Land Used for Agriculuture by Year in",
                           PROVINCE_CODES[province_chosen()]),
             x = "Land Usage for agriculture (square kilometers)",
             y = "Wetland (square kilometers)")
    }
    }) # ending renderplot
  output$txt = renderText({
    print("The data used in this exploration is available through the Open Government
     ↪ Portal: https://open.canada.ca/data/.")
  })
} # ending server

shinyApp(ui, server)
```