1 **REPRODUCIBILITY MINI PROJECT: ALBERTA TREES**

2 Rolando Trejo Pérez 1^*

3   1. Institut de recherche en biologie végétale, Université de Montréal

4 *Corresponding Author

5 email: rolando.trejo.perez@umontreal.ca

6 **Author Contributions:** RTP is writting this mini reproducibility project.

10 **Conflict of Interest statement**

11 No conflicts of interest

Abstract

15 The goal of this manuscript is using reproducibility workflow in ecology and evolution. I used open access data, which

16 is available in Boreales (see references below). This data was rescued by Hesketh et al., 2021. This mini project was

17 created with the open-source software R, OSF and GitHub. A very simple question is stated: does species richness

18 varies along the soil temperature? All the code used here is available in a public GitHub repository.

19 **Key-words**: reproducibility, diversity, species richness, soil temperature.

Introduction

Global warming is predicted to significantly alter species physiology, biotic interactions and thus ecosystem functioning, as a consequence of coexisting species exhibiting a wide range of thermal sensitivities. The richness diversity of plants decreased with increasing soil temperature, driven by decreasing plant species richness (Robinson *et al.* 2018). Species richness and its relationship with temperature has been assessed along altitudinal and horizontal gradients revealing significant linear relationships between species richness and altitude and climatic variables (Pickering *et al.* 2008).

Here, in this manuscript, I used a very simple linear model to explore the relationship between vascular diversity and soil temperature from the Seasonal Dynamics (SEADYN) and later Annual Dynamics (ANNDYN) research project. The primary purpose of this project was to document seasonal changes in the vegetative composition during the snow-free season (May through October) and longer-term changes in vegetation and forest mensuration for boreal forest stands in Alberta, Canada dominated by Pinus banksiana (Lamb.) (Hesketh *et al.* 2022)

Methods

This mini reproducibility project was built using rescued data by Amelia Hesketh, Jenna Loesberg, Ellen Bledsoe, Justine Karst,and Ellen Macdonald in 2021 from an Alberta legacy dataset spanning four decades (1980-2015). We use two different datasets: 1) Hondo Vascular Cover (1980-2015) and 2) Hondo Soil Temperature (1980-2010). These datasets are available in Borealis at https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP3/PZCAVE

A very simple model is explored in this reproducible project: Does species richness varies along the soil temperature? For this purpose i only use 2010 data and the following linear model approach:

$$SR = ST + e$$

where SR denotes the species richness, ST is the soil temperature in Celsius,and e is the error.

This mini project was created with the open-source software R. Packages used were dataverse,tibble,dplyr,ggplot2, and mgcv. Statistical analyses were carried out in R 3.5.0 (R Core Team 2017). All code used in this manuscript is available on GitHub https://github.com/CIEE-Living-Data-Project/Rolando_Trejo_Reproducibility_LDP_2022.

I followed 4 steps to answer the question stated in this manuscript. It included dowloading data from the dataverse (Borealis) repository, data manipulation to get the final cleaned data, model codification, and grhapich visualization to

45 show the main result.

1. DOWNLOAD DATA FROM DATAVERSE (download in 2022-15-09)

47 Two databases were download from the Seasonal and annual dynamics of western Canadian boreal forest plant commu-

48 nities: a legacy dataset spanning four decades at https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP3/

49 PZCAVE

```r
# 1.1 Vascular cover from 1980 to 2015
library("dataverse")
Hondo_VascularCover_1980_2015 <- get_dataframe_by_name(
  filename = "Hondo_VascularCover_1980_2015.tab",
  dataset = "10.5683/SP3/PZCAVE",
  server = "https://borealisdata.ca/dataverse/ubc")


# 1.2 Soil temperature from 1980 to 2010


Hondo_SoilTemp_1980_2010 <- get_dataframe_by_name(
  filename = "Hondo_SoilTemp_1980_2010.tab",
  dataset = "10.5683/SP3/PZCAVE",
  server = "https://borealisdata.ca/dataverse/ubc")
```

2. DATA MANIPULATION

51 To answer my question, a cleaning procedure is required. It consists of merging the 2010 vascular cover and soil

52 temperature subsets, converting the abundances data in binary data, summarizing the number of species, and creating

53 a final file conserving only the number of species and soil temperature (Species_Temp).

```r
# 2.1. Generate a subset of data considering only 2010 data to simplify
# the statistical analyses.


    Hondo_VascularCover_2010 <- subset(Hondo_VascularCover_1980_2015,
                              year== "2010" )
    Hondo_SoilTemp_2010 <- subset(Hondo_SoilTemp_1980_2010,
```

3

```r
                              year== "2010" )
# 2.2. Merge Vascular cover and Soil temperature datasets by month,
# year, stand and quad.


    df=inner_join(Hondo_VascularCover_2010,

              Hondo_SoilTemp_2010,

              by = c("month","year","stand", "quad"))


# 2.3 It creates a subset ob non-binary data


df_non_binary <- subset(df,select = c(stand, month,year,

                              date,quad,temp_C))


#  2.4 It creates a subset of binary data: abundances needs to
# be transformed into 0 and 1
# in order to count the species richness in each row.


df_species <- subset(df,select = c(-stand, -month,-year,-date,-temp_C))


# 2.5 It transforms abundances into 0 and 1


df_binary <- data.frame(df_species[1], (df_species[-1] > 0) * 1)


# 2.6 It merges the non binary data, soil temperature, and species
# abundances transformed into 0 and 1.


df_trans <- data.frame(df_non_binary,df_binary)


# 2.7 It sums the number of species in each row


df_trans$species_no <- rowSums( df_trans[,8:221] )
```

```
# 2.8 It create the final table version: it includes only the
# species richness and soil temperature


Species_Temp <- subset(df_trans,select = c(species_no,temp_C))
```

3. MODEL: Linear model or not?

I compared the linear model and non-linear model approach using gam R package.

```
library(mgcv)


# The following code provides a linear and non-linear model approach:


linear_model <- gam(species_no  ~ temp_C, data = Species_Temp)
smooth_model <- gam(species_no  ~ s(temp_C), data = Species_Temp)
```

4. VISUALIZATION

To facilitate the interpretation of the linearity when modeling species richness and soil temperature, two graphics are proposed.

```
library(ggplot2)


# 4.1 It graphs the linear model (species richness vs soil temperature)


Linear<-ggplot(Species_Temp, aes(x = temp_C, y = species_no )) +
  geom_point() +
  geom_line(colour = "red", size = 1.2,
            aes(y = fitted(linear_model))) +
  xlab("Soil temperature (C)") +
  ylab("Species richness") +
  labs(title = "A) Linear model line")+
  theme_bw()
```

```
# 4.2 It graphs the non-linear model (species richness vs soil temperature)


Non_linear<-ggplot(Species_Temp, aes(x = temp_C, y = species_no )) +

  geom_point() +

  xlab("Soil temperature (C)") +

  ylab("Species richness") +

  labs(title = "B) Smooth model curve")+

  geom_line(colour = "blue", size = 1.2,

            aes(y = fitted(smooth_model))) +

  theme_bw()
```
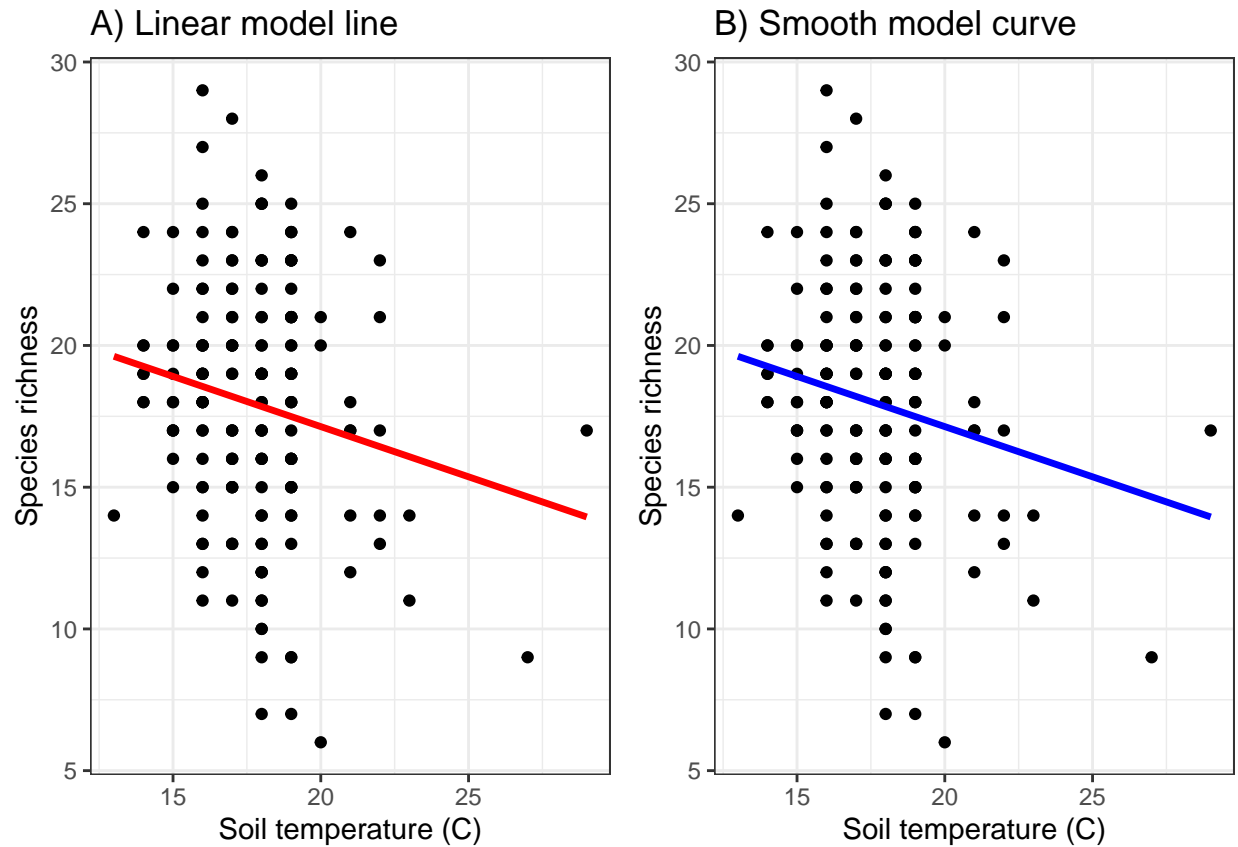
Results

A higher species richness is linked to a lower soil temperature according to this simple linear model. We can also see that linearity is supported by these data (see Figure 1). However, this is just a model considering soil temperature as the only predictor. Other variables must be considered to search if it is a real observed trend. See more detail in the Living Data Tutorials.

Discussion

The code used here in this mini reproducibility project can be used as guide through the creation and management of a fully reproducible manuscript using RMarkdown and Rstudio.

References

Hesketh, A.V., Loesberg, J.A., Bledsoe, E.K., Karst, J. & Macdonald, S.E. (2022). Seasonal and annual dynamics of western Canadian boreal forest plant communities: A legacy data set spanning four decades. *Ecology*.

Pickering, C., Hill, W. & Green, K. (2008). Vascular plant diversity and climate change in the alpine zone of the Snowy Mountains, Australia. *Biodiversity and Conservation*, 17, 1627–1644.

Robinson, S.I., McLaughlin, Ó.B., Marteinsdóttir, B. & O'Gorman, E.J. (2018). Soil temperature effects on the structure and diversity of plant and invertebrate communities in a natural warming experiment. *Journal of Animal Ecology*, 87, 634–646.

Figure 1. A) linear model approach and B) a non linear model approach. Linearity is respected.

Tables

There are not table to show.

Appendices

There are not appendices to show.