# Draft Manuscript

## Emma Menchions

2022-09-23

- 4 Title: Can gradients in annual rainfall & forest type predict the distribution & abundance of 4 Ericaceous shrubs? -
- 5 BIOL 548T LDP
- 6 Emma Menchions<sup>1</sup>\*
- 1. Dept. of Botany, University of British Columbia, Vancouver, BC, Canada
- \*Corresponding Author
- 9 email: emma.menchions@botany.ubc.ca
- Author Contributions: EM conceived of the study, conducted the analyses, and wrote the original, and revised drafts
- of the manuscript.
- Data Availability: The data and code that support the findings of this study are openly available on GitHub at https:
- 13 //github.com/emench/Menchions-BIOL548T.
- 14 Conflict of Interest statement
- No conflicts of interest

## 6 Abstract

- Patterns in plant distributions are driven by a multitude of factors, one of which is soil properties. For many ericaeceous
- shrubs, high acidity soils are often an important determinant of distribution. Here I used data collected from Thiffault et
- 19 al. (2015a) across plots in the understory of boreal forests of Quebec from 1985 to 2000 to determine if the abundance
- of four species of ericeceous shrubs (Chamaedaphne caluculata, Rhododenron groenlandicum, Kalmia angustifolia
- and Vaccinium sp. ) can be predicted by patterns in soil acidity with habitat. Specifically, since conifer forests with

high rainfall are typically associated with more highly acidic soils, I predicted this habitat type would contain the
highest abundance of of these shrubs, and in particular *C. caluculata and R. groenlandicum* since they also prefer
wetter habitats. For *K. angustifolia* and *Vaccinium sp.* I predicted that low rainfall coniferous forests would contain the
greatest abundance due to high acidity from the conifers and lower moisture which these species supposedly prefer.
I used box plots and pariwise Wilcox tests to compare medians between percent cover of different types of habitats
defined by high or low rain fall and forest type. I found that for *C. caluculata and R. groenlandicum*, my predictions
were quite close to the observed patterns, suggesting that perhaps soil acidity and moisture are primary drivers of
distribution. However, for *K. angustifolia* and *Vaccinium sp.* my predictions did not match the results. However,
conifer forests and non-forested (presumably boggy areas) plots did tend to contain the highest median abundance
which may also suggest soil acidity as a strong driver of distribution. Future studies and analyses should use more
direct observations of soil pH measurements to study these patterns rather than relying on the proxy variable of rainfall
and forest type used here to avoid the many confounding factors that this index introduces.

34 Key-words: Ecology, Ericaceae, LDP, Plant Ecology, Vegetation Mapping, Vegetation Patterns

## 35 Introduction

- Four species of Ericaceae dominate the understory in the northern boreal forests of Quebec, Canada. Two of these
- species Rhododenron groenlandicum and Chamaedaphne caluculata are said to prefer acidic soils and wetter habitats
- whereas the other dominant species Kalmia angustifolia, Vaccinium angustifolium, and Vaccinium myrtillodes are
- so known to prefer drier, but also acidic soils. Many factors contribute to soil acidity, but a few variables in particular,
- rainfall and forest type have been observed to drive broad patterns in soil pH in the United States "Soil acidity an
- overview | ScienceDirect topics" (n.d.a). Therefore, I ask, can annual rainfall and forest type (conifer, deciduous, or
- mixed) predict patterns of abundance for these species in the boreal zone of Northern Quebec between 1985 and 2000?
- 43 Acidic soils are often associated with conifer forests and higher rainfall (Hornung 1985; "Soil acidity an overview |
- 44 ScienceDirect topics" n.d.b). Therefore, species with a preference for more acidic soils should be more often found in
- conifer forest types with higher precipitation. Therefore, I predict that R. groenlandicum and C. calyculata, will have
- the highest mean percent cover in areas with conifer forest types and higher annual precipitation. In contrast, I predict
- that K. angustifolia and Vaccinium sp., will have the highest mean percent cover in conifer forests with lower annual
- 48 precipitation.
- <sup>49</sup> Specifically, I predict the following pattern of mean percent cover for R. groenlandicum and C. calyculata: (H = high
- precipitation, L = low precipitation, c= conifer forests, d=deciduous forests, m=mixed forest)
- 51 From highest to lowest predicted mean percent cover with high rainfall and conifer forests being indicative of higher
- abundance: H-c, L-c, H-m, H-d, L-m, L-d
- 53 For K. angustifolium and Vaccinium sp. I predict the following pattern, with less rainfall being of greater importance
- to abundance: L-c, L-m, L-d, H-c, H-m, H-d

## 55 Methods

#### 56 Data collection

- Data was retrieved from Thiffault et al. (2015b): https://datadryad.org/stash/dataset/doi:10.5061%2Fdryad.4767v.
- According to Thiffault et al. (2015c), the data on abundance of the 4 species were collected from 15,339 circular plots
- $_{59}$  across central and northern Quebec along 1.5-2 km long transects of which supported 5-7 plots. Forest type: was
- determined by visual estimations at the time that the plots were visited by assigning a type corresponding to the two
- 61 most dominant tree species within the plot. This lead to a total of 15 total forest types. Climate data was collected
- by identifying the closest meteorological station (51 in the study region) and correcting for differences in latitude,

63 longitude and elevation between the plot and the station using distance weighting and multiple linear regression.

#### Data download (script: "part-1 data download")

- 65 All data download, preparation and statistical analyses were carried out in R 4.2.1 (R Core Team 2022). All scripts
- 66 used are available on GitHub https://github.com/emench/Menchions-BIOL548T. I first retrieved this data from Dryad
- using a script that automates the download from Dryad.

#### Data preparation (script: "part-2\_data\_preparation")

- <sub>69</sub> I then prepared the data for analysis by altering the forest type classification and creating a variable indicating whether
- the plot was in a high or low rainfall area. For the forest type classifications, I used the metadata on forest type codes
- provided to create 4 categories of forest type: "c" (coniferous, if the two dominant tree species in the plot were both
- coniferous, n=1282), "d" (deciduous, if the two dominant species were both deciduous, n=45), "m" (mixed, if the two
- most dominant species contained a conifer and a deciduous species, n=69), and "n" (non-forested, n=26).

```
74 ## [1] 1282 49
75 ## [1] 45 49
76 ## [1] 69 49
```

77 ## [1] 26 49

- I then created a rainfall index. I calculated the median annual usable precipitation (cm) which was 5cm/ year. I used
- <sub>79</sub> this as a threshold and defined any plot with annual usable precipitation > 5cm to a high precipitation category (H) and
- below this threshold to low annual rainfall (L).

- Finally, I combined the forest type and rainfall variables into an index (rf index) that described the rainfall category
- followed by the forest type (e.g. H c = high rainfall, conifer forest).

```
# Combining and classifying as forest type - rainfall index

X_mat <-
    X_mat %>%
    unite(., rf_index, c("RainFall","Forest_Type"), sep = "_", remove =F, na.rm = FALSE)

# writing altered data as .csv ----
write.csv(X_mat, here::here("data","X_mat_processed.csv"))
```

#### 83 Plotting (script: "part-3 data analysis")

- 484 I constructed box plots displaying percent cover of each species as the response variable and rf (rainfall forest cover
- index) as the grouping variable using using ggplot2::ggplot. I decided to log-transform the response variable since the
- data was skewed with outliers making the display and comparison between groups difficult in all of the plots.

```
# type index in box plots
 CAL_plot <- ggplot(data=CAL, aes(x=rf_index, y=log(V_CAL))) + # log transforming the y axis
    geom_boxplot() # automatically removes observations with 0 % cover - which is ok
 # Repeating for Rhododendron groenlandicum
 RHG <- data.frame(Y_mat$V_RHG,X_mat$PRECU,X_mat$rf_index)</pre>
 colnames(RHG) <- c("V_RHG","PRECU","rf_index")</pre>
 RHG_plot <- ggplot(data=RHG, aes(x=rf_index, y=V_RHG)) +</pre>
   geom_boxplot()
 RHG_plot <- ggplot(data=RHG, aes(x=rf_index, y=log(V_RHG))) + # log transforming the y axis
   geom_boxplot()
## Repeating for Kalmia angustifolia
 # Plotting:
 KAA <- data.frame(Y_mat$V_KAA,X_mat$PRECU,X_mat$rf_index)</pre>
 colnames(KAA) <- c("V_KAA", "PRECU", "rf_index")</pre>
 KAA_plot <- ggplot(data=KAA, aes(x=rf_index, y=V_KAA)) +</pre>
   geom_boxplot()
 KAA_plot <- ggplot(data=KAA, aes(x=rf_index, y=log(V_KAA))) + # log transforming the y axis</pre>
   geom_boxplot()
## Repeating for Vaccinium sp.
   VAAM <- data.frame(Y_mat$V_VAAM,X_mat$PRECU,X_mat$rf_index)</pre>
    colnames(VAAM) <- c("V_VAAM", "PRECU", "rf_index")</pre>
   VAAM_plot <-ggplot(data=VAAM, aes(x=rf_index, y=V_VAAM)) +</pre>
      geom_boxplot()
   VAAM_plot <- ggplot(data=VAAM, aes(x=rf_index, y=log(V_VAAM))) + # log transforming the y axis
```

#### geom\_boxplot()

I then combined all of the above plots into one single chart for easier comparison.

## 88 Statistical analysis

- <sup>89</sup> I used a Kruksal-Wallis test to determine if any any groups were significantly different from the rest, or in other words,
- 90 if at least one rainfall forest cateogry contains a significantly different percent cover of a given species than the other
- groups. For all species, p < 0.05 indicating we can reject the null hypothesis that percent cover is the same across rainfall
- 92 forest type categories for all species. To further assess the relationships, I performed a pairwise wilcox test to determine
- whether percent cover differed significantly from each other in pairs. This was not in the initial pre-registration but is
- an extended step to further test the hypothesis.

## 95 Results & Discussion

- 96 ## Warning: The labeller API has been updated. Labellers taking `variable` and
- 97 ## `value` arguments are now deprecated. See labellers documentation.
- 98 ## Warning: Removed 1084 rows containing non-finite values (stat\_boxplot).

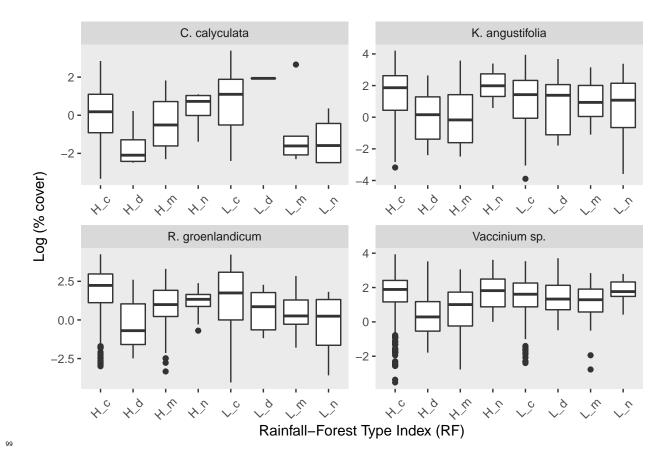


Figure 1. Box plots of *C. calyculata*, *K. angustifolia*, *R. groenlandicum*, and *Vaccinium sp.* percent cover in plots that fall in each of the rainfall-forest type categories. The thick horizontal lines represent the median for each category.

The y axis was log-transformed for display purposes.

Table 1. Mean % cover for *C. calyculata* in different rainfall forest type categories.

```
CAL_cover %>% group_by(rf_index) %>% summarise(median = median(cover, na.rm=T))%>%
    arrange(desc(median))
```

```
## # A tibble: 8 x 2
104
          rf_index median
    ##
105
          <chr>
                     <dbl>
    ##
106
    ## 1 H_c
                    0.0357
107
    ## 2 H_d
108
    ## 3 H_m
                    0
    ## 4 H_n
                    0
110
    ## 5 L_c
                    0
```

```
112 ## 6 L_d 0
113 ## 7 L_m 0
114 ## 8 L_n 0
```

For *C. calyculata*, the only rainfall forest type category that had median percent cover above 0 was high rainfall coniferous forests. This supports the prediction that this species would have a greater preference for this category possibly due to the higher acidity and moisture provided.

Table 2. Mean % cover for *R. groenlandicum* in different rainfall forest type categories.

```
RHG_cover %>% group_by(rf_index) %>% summarise(median = median(cover, na.rm=T)) %>%
arrange(desc(median))
```

```
## # A tibble: 8 x 2
         rf_index median
    ##
120
         <chr>
                     <dbl>
    ##
121
    ## 1 H_c
                     8.5
122
    ## 2 L c
                     3.67
123
    ## 3 H_n
                     3.48
124
    ## 4 H m
                     2.17
125
    ## 5 L n
                     1.66
    ## 6 L_m
                     1
127
    ## 7 H d
                     0.4
    ## 8 L_d
                     0.321
129
```

For *R. groenlandicum*, the order of my predictions also matched those observed. H\_c had significantly higher percent cover than any other rf type (p<0.05), followed by L-c, H\_m, L\_m H\_d and L\_d. This may indicate that *R. groenlandicum* indeed prefers more acidic soils created by higher rainfall, but mroe importantly coniferous forests.

However, this is confounded with other factors that may drive its preference for coniferous habitat such as other soil

properties associated with conifer needles and light availability.

**Table 3.** Mean % cover for *K. angustifolium* in different rainfall forest type categories.

```
KAA_cover %>% group_by(rf_index) %>% summarise(median = median(cover, na.rm=T))%>%
arrange(desc(median))
```

```
## # A tibble: 8 x 2
136
         rf_index median
    ##
         <chr>
                     <dbl>
138
    ## 1 H n
                     7.58
    ## 2 H_c
                     4.5
140
    ## 3 L_n
                     3.74
    ## 4 L_d
                     2.6
142
    ## 5 L_m
                     1.96
143
    ## 6 L c
                     1.33
144
    ## 7 H_d
                     0.667
145
    ## 8 H m
                     0.5
```

For *K. angustifolium* the patterns in abundance that I predicted (L-c, L-m, L-d, H-c, H-m, H-d) were not supported by those observed. The category with the highest median percent cover was non-forested area however this was not significantly different from abundance in high rainfall coniferous forests (p>0.05). This may indicate that light-availability is a more important determinant of its distribution. Additionally, non-forested plots may often be boggy areas with acidic soils from sphagnum growth, so perhaps this species' distribution is highly driven by soil acidity, so much so that the acidity produced by conifer needles and high rainfall is not sufficient. The fact that low rainfall coniferous forests had one of the lowest percent covers is curious. It suggests that other factors not included here should be included in further assessments of distribution.

```
VAAM_cover %>% group_by(rf_index) %>% summarise(median = median(cover, na.rm=T)) %>%
arrange(desc(median))
```

```
## # A tibble: 8 x 2
155
          rf_index median
    ##
156
          <chr>
                     <dbl>
    ##
157
    ## 1 H_c
                       6.6
158
    ## 2 H n
                       5.99
159
    ## 3 L_n
                       5.83
    ## 4 L_c
                       5
161
    ## 5 L_d
                       3.83
    ## 6 L_m
                       3.22
163
    ## 7 H m
                       2.75
```

- 165 ## 8 H d 1.33
- Finally, for Vaccinium sp. median percent cover again did not follow the pattern I predicted, except for H\_d and H\_m
- having the lowest values, however, they were not significantly different from each other (p<0.05). the categories with
- higher abundance were H c, H n and L n although they were not significantly different from each other (p>0.05).
- H\_c was significantly higher than L\_c (p<0.05) which I originally thought would contain the greatest abundance.
- This pattern might be similar in nature to the one observed for K. angustifolium, where coniferous and non-forested
- (presumably boggy) areas with high rainfall contain the most abundance, possibly due to the importance of acidity in
- these environments compared to those associated with deciduous or mixed forests.
- Analysis to test if these species distributions are driven partly by soil acidity could ideally be performed using direct
- measures of soil acidity at the plots rather than the proxy variables used here. Upon future visits to the sites, these
- measurements could be taken to construct models and compare preferences between species to better predict what
- drives their distribution patterns and regenerative ability in logging and fire disturbed habitats.

## References

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., et al. (2022). Rmarkdown: Dynamic documents
- 179 *for r*.
- <sup>180</sup> Aust, F. & Spitzer, L. (2022). Prereg: R markdown templates to preregister scientific studies.
- Hornung, M. (1985). Acidification of soils by trees and forests. Soil Use and Management, 1, 24–27.
- Müller, K. (2020). Here: A simpler way to find your files.
- Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., et al. (2022). Rfast: A collection
- of efficient and extremely fast r functions.
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Com-
- puting, Vienna, Austria.
- Rodríguez-Sánchez, F., Jackson, C.P. & Hutchins, S.D. (2022). Grateful: Facilitate citation of r packages.
- Simonsohn, U. & Gruson, H. (2022). Groundhog: Version-control for CRAN, GitHub, and GitLab packages.
- Soil acidity an overview | ScienceDirect topics. (n.d.a).
- Soil acidity an overview | ScienceDirect topics. (n.d.b).
- 191 Thiffault, N., Grondin, P., Noël, J. & Poirier, V. (2015b). Ecological gradients driving the distribution of four Ericaceae
- in boreal Quebec, Canada. *Ecology and Evolution*, 5, 1837–1853.
- 193 Thiffault, N., Grondin, P., Noël, J. & Poirier, V. (2015c). Ecological gradients driving the distribution of four Ericaceae
- in boreal Quebec, Canada. *Ecology and Evolution*, 5, 1837–1853.

- Thiffault, N., Grondin, P., Noël, J. & Poirier, V. (2015a). Ecological gradients driving the distribution of four Ericaceae in boreal Quebec, Canada. *Ecology and Evolution*, 5, 1837–1853.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., *et al.* (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686.
- Xie, Y. (2014). Knitr: A comprehensive tool for reproducible research in R. In: *Implementing reproducible computa*tional research (eds. Stodden, V., Leisch, F. & Peng, R.D.). Chapman; Hall/CRC.
- 201 Xie, Y. (2015). Dynamic documents with R and knitr. 2nd edn. Chapman; Hall/CRC, Boca Raton, Florida.
- $_{202}$  Xie, Y. (2022). Knitr: A general-purpose package for dynamic report generation in r.
- <sup>203</sup> Xie, Y., Allaire, J.J. & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC, Boca Raton, Florida.
- Xie, Y., Dervieux, C. & Riederer, E. (2020). R markdown cookbook. Chapman; Hall/CRC, Boca Raton, Florida.

## grateful::cite\_packages(output = "paragraph")

- <sup>206</sup> We used R version 4.2.1 (R Core Team 2022) and the following R packages: grateful v. 0.1.11 (Rodríguez-Sánchez
- et al. 2022), groundhog v. 2.0.1 (Simonsohn & Gruson 2022), here v. 1.0.1 (Müller 2020), knitr v. 1.40 (Xie 2014,
- <sup>208</sup> 2015, 2022), prereg v. 0.6.0 (Aust & Spitzer 2022), Rfast v. 2.0.6 (Papadakis et al. 2022), rmarkdown v. 2.16 (Xie et
- al. 2018, 2020; Allaire et al. 2022), tidyverse v. 1.3.2 (Wickham et al. 2019).

# 210 Appendix

# Table A1. P-Values for Pairwise Wilcox test for C. calyculata

## ${\tt CAL\_wt\_pframe}$

212	##	H_c	H_d	H_m	H_n	L_c	L_d	L_m
213	## H_d	0.0003184252	NA	NA	NA	NA	NA	NA
214	## H_m	0.1858401983	0.04590974	NA	NA	NA	NA	NA
215	## H_n	0.5004745729	0.17309149	1.0000000	NA	NA	NA	NA
216	## L_c	0.4543463732	0.00536943	0.4543464	0.5923667	NA	NA	NA
217	## L_d	0.0707266386	0.84421956	0.2318784	0.3633565	0.15627824	NA	NA
218	## L_m	0.0051409599	0.59236673	0.1562782	0.3633565	0.04590974	0.5923667	NA
219	## L_n	0.1562782421	0.36335646	0.5309553	0.5923667	0.34415728	0.4543464	0.5790773

# Table A2. P-Values for Pairwise Wilcox test for R. groenlandicum

## RHG\_wt\_pframe

221	##	H_c	H_d	H_m	H_n	L_c	L_d	L_m
222	## H_	d 3.378207e-10	NA	NA	NA	NA	NA	NA
223	## H_	m 4.416647e-04	0.01272471	NA	NA	NA	NA	NA
224	## H_	n 2.573869e-02	0.01272471	0.88769677	NA	NA	NA	NA
225	## L_	c 3.929857e-03	0.00012994	0.16511102	0.4562412	NA	NA	NA
226	## L_	d 4.451922e-04	0.75177373	0.09693363	0.1508120	0.01272471	NA	NA
227	## L_	m 7.730911e-07	0.33233183	0.09693363	0.1051708	0.00574138	0.4148958	NA
228	## L_	n 9.136257e-04	0.16672719	0.41489583	0.3561606	0.09693363	0.3561606	0.7530249

## Table A3. P-Values for Pairwise Wilcox test for K. angustifolium

## ${\tt KAA\_wt\_pframe}$

230	##	H_c	H_d	H_m	H_n	L_c	L_d
231	## H_d 1.9362	38e-03	NA	NA	NA	NA	NA
232	## H_m 5.6859	52e-04 0.664	338226	NA	NA	NA	NA

```
## H_n 2.376205e-01 0.001833148 0.001833148
                                                                               NA
                                                                     NA
233
   ## L_c 9.968487e-07 0.575015629 0.339490167 0.01756374
                                                                               NA
                                                                     NA
   ## L_d 4.283907e-01 0.408532106 0.339490167 0.18243733 0.6290323
                                                                               NA
235
   ## L_m 3.394902e-01 0.104272082 0.035632366 0.04610518 0.3394902 0.7356280
   ## L_n 7.356280e-01 0.135600896 0.113185867 0.20464047 0.3394902 0.6130469
237
   ##
                 L_m
   ## H_d
                  NA
239
   ## H_m
                  NA
240
   ## H_n
                  NA
241
   ## L_c
                  NA
242
   ## L_d
                  NA
   ## L_m
                  NA
244
   ## L_n 0.9793181
```

Table A4. P-Values for Pairwise Wilcox test for *Vaccinium sp.* 

## VAAM\_wt\_pframe

247	##		H_c	H_d	H_m	H_n	L_c	L_d
248	##	H_d	6.927476e-07	NA	NA	NA	NA	NA
249	##	H_m	1.873071e-05	0.3507908097	NA	NA	NA	NA
250	##	H_n	7.349944e-01	0.0763626626	0.20662594	NA	NA	NA
251	##	L_c	9.360567e-03	0.0001889474	0.00652134	0.9333849	NA	NA
252	##	L_d	4.632609e-01	0.0528171906	0.24486419	0.9654529	0.8776301	NA
253	##	L_m	9.360567e-03	0.1187779551	0.46326089	0.4632609	0.1814597	0.5917751
254	##	L_n	9.802611e-01	0.0040257523	0.01772142	0.8162312	0.4632609	0.4632609
255	##		L_m					
256	##	H_d	NA					
257	##	H_m	NA					
258	##	H_n	NA					
259	##	L_c	NA					
260	##	L_d	NA					
261	##	L_m	NA					
262	##	L_n	0.1248045					