Analysis on Spatial Pattern of Douglas-fir in British Columbia

DATA 589: Special Topic

April 27, 2024

Jinxin Wang	95691002
Nan Tang	48735633
Xing Xu	42946970

Contents

Abstract

This study investigates a comprehensive spatial analysis of the distribution of Douglas-fir in British Columbia using a dataset from GBIF [1]. Through spatial point process analysis, we assess the influence of environmental factors such as elevation, forest cover, proximity to water, and human activities on the intensity of Douglas-fir distribution. A Poisson general linear regression model was constructed to quantify these relationships and predict spatial distribution patterns. Our findings indicate a non-homogeneous distribution with significant clustering at specific distances, suggesting particular habitat preferences and avoidance areas. Despite the model's success in capturing the general trends associated with elevation, Human Footprint Index, and distance to water, it requires further refinement to accurately represent the impact of forest cover. The study underscores the potential of incorporating additional covariates such as slope, soil characteristics, and climatic variables to enhance model robustness and contribute to forest management and conservation strategies.

Keywords: Douglas-fir, Spatial Point Process Analysis, British Columbia, Poisson Regression, Habitat Modeling

1 Introduction

Douglas-fir (Coast Douglas-fir, British Columbia Fir, or Pseudotsuga Menziesii Franco)[2], is one of the world's most important and valuable timber species, used for furniture, poles, fences, and flooring, among other things. It is a medium-to-large-sized (reaching nearly 100m in height), evergreen conifer, and at maturity has a moderately dense, conical crown with long branches, and deeply furrowed, dark reddish-brown bark with irregular, broad ridges.

Understanding the distribution patterns of Douglas-fir is crucial for ecological and conservation efforts. By analyzing these patterns, we can gain insights into the factors influencing their habitat preferences and population dynamics. From the Global Biodiversity Information Facility (GBIF) we employed a dataset [1] that includes coordinates of 113,610 Douglas-fir, 10,269 of which are in BC. Based on this dataset we intend to conduct an analysis on the distribution of Douglas-fir in BC(Figure 1 [3]).

There are many factors that might have impacts on the distributions of Douglas-fir. It occupies nearly all of southern British Columbia, western Washington, and western Oregon from sea level to 1,500m elevation and even higher. It inhabits all inland forests in the Greater Northwest except for the driest ponderosa pine types, juniper woodlands, and highest subalpine habitats. Along the eastern slope of the Continental Divide in northern Montana and southern Alberta, Douglas-fir, sometimes accompanied by limber pine, replaces ponderosa pine in forming low-elevation dry-site forests, in a climate that is too cold for ponderosa. Southward, Douglas-fir grows on the highest mountains ranges from Utah and Wyoming well into Mexico. It also spreads southward along the coast and Sierra Nevada to central California. Douglas-fir is classified as one timber specie with an intermediate level in shade tolerance, which means it has the ability to regenerate and grow up beneath other trees and eventually displace them.[4]

Therefore, we chose four potential factors to investigate: elevation, forest cover, distance to water, and human activities. We proposed three research questions as follows:

- What is the spatial intensity of Douglas-fir distribution across different regions of British Columbia, and is there any significant avoidance or clustering?
- How do the locations of Douglas-fir correlate with key geographical factors including elevation, forest cover, proximity to water, and human activities?
- Can we develop a Poisson general linear regression model to accurately predict the spatial distribution of Douglas-fir based on geographical and biological covariates?

By addressing these research questions, we aim to enhance our understanding of Douglas-fir ecology in BC and contribute valuable insights for effective forest management, biodiversity conservation, and sustainable resource utilization in BC's ecosystems.

Figure 1: The Distribution of Douglas-fir in British Columbia

2 Methodology

2.1 Data Description and Pre-processing

We employed the dataset of Douglas-fir from the GBIF data repository [1] and kept the information for locations of the Douglas-fir grown in BC by filtering the longitude, latitude, and region of each observation for spatial point process analysis. In addition, we tried to include some environment variables: elevation, forest cover, Human Footprint Index(HFI)[5], and the distance to water, as potential covariates for the further model construction. The coordinates data for these variables were measured based on the BC Albers projection. We converted the longitude and latitude from the original dataset to the same form as the ones for environment variables and excluded the missing values. So we can model the intensity of Douglas-fir with these covariates in a valid form.

2.2 Exploratory Data Analysis

To answer our research questions, we will use spatstat package [**?**] which provides classes and functions for handling and plotting spatial data in R. Before we specified and selected the model form, we inspected and explored data on a spatial scale through effective visualization of points patterns to obtain preliminary insights that can inspire the further modeling process. Then we performed detailed exploratory data analysis with the first and second moment descriptive statistics.

2.2.1 First Moment Analysis: Intensity

Investigation of the intensity of a point pattern is one of the first and most important steps in spatial data analysis. The intensity is a basic descriptive characteristic of a point process, an average ('expectation' or 'first moment') analogous to the average of a population of numbers.

We first calculated the average intensity of Douglas-fir in BC. Then we performed a quadrat test for checking homogeneity after visualizing the point pattern of Douglas-fir within the defined window for BC. Based on this test result, we estimated their intensity in BC through non-parametrically kernel density estimation and identifying hotspots by a scan test for highlighting the areas in BC with elevated intensity of Douglas-fir. Additionally, we estimated the relationship between the intensity of Douglas-fir and the covariates to explore their potential correlations respectively by `rho()` function, which can provide insights for identifying the form for model fitting.

2.2.2 Second Moment Analysis: Correlation

Second moment quantities for point processes are intimately related to count pairs of points, or adding up contributions from each pair of points in the process. Through second moment analysis, we can have an understanding of the relationships between points. There are several methods to analyze the spatial correlation including the Morisita index M, Ripley K's Function, and pair correlation function $g(r)$ [6].

In this project, we chose the pair correlation function $g(r)$ to investigate the correlation of Douglas-fir. $g(r)$ is the probability of observing a pair of points of the process separated by a distance r, divided by the corresponding probability for a Poisson process, and the value $g(r) = 1$ is consistent with complete spatial randomness because of the way the function has been standardized. A value $g(r) < 1$ indicates that interpoint distances equal to r are less frequent than would be expected for a completely random process, so this suggests regularity. A value $g(r) > 1$ indicates that this interpoint distance is more frequent than expected for a completely random pattern, which suggests clustering. We also used the bootstrap approach to include 95% confidence intervals in our analysis. Any deviation from the grey-shaded intervals shows possible correlations.

2.3 Intensity Modeling

2.3.1 Poisson Model Fitting

Under the assumption of independence, the intensity function of Douglas-fir can be modeled as a Poisson point process model in the similar form of Poisson GLM with some spatial covariates.

We started by examining the relationship between the intensity of Douglas-fir and each covariate and decided roughly what kinds of relationship they might have. To avoid collinearity we also conducted a correlation check between variables. Then, based on our pre-analysis above, we proposed a reasonable form for the model and fit it with our data.

2.3.2 Model Selection and Validation

To investigate if the proposed model form can well explain our data, we calculated the AIC values and performed a Likelihood Ratio Test to compare the proposed model and the model with only an intercept. It turned out that adding complexity in the model can support our observation better than no added terms in an intercept model.

Although the proposed model can well explain the provided data, we still need to evaluate how accurately this model can predict the given inputs. Thus we applied a Quadrat Test to estimate if the intensity of Douglas-fir fitted by the proposed model significantly deviates from the expected value. The last but not the least, we performed a partial residual analysis for all potential covariates to investigate if their partial effects on the intensity of Douglas-fir are well and accurately explained by the model, thus we can tell if the considered covariates in the model are reasonable and meaningful in fitting the intensity value of interest. To explore the partial effects of covariates in-depth, we will try the generalized additive model to improve the fitness and accuracy by adding more complexity to the proposed model.

3 Results

From the results based on the first moment analysis, the average intensity of Douglas-fir across BC is $0.011/m^2$. As shown in the Figure 3 below, Douglas-fir concentrates in the southwestern and southern parts of BC (the same pattern is also illustrated in the intensity graph), especially on Vancouver Island. Both graphs indicate that the intensity of Douglas-fir in BC may not be homogeneous. Therefore, we conducted a quadrat test on the data

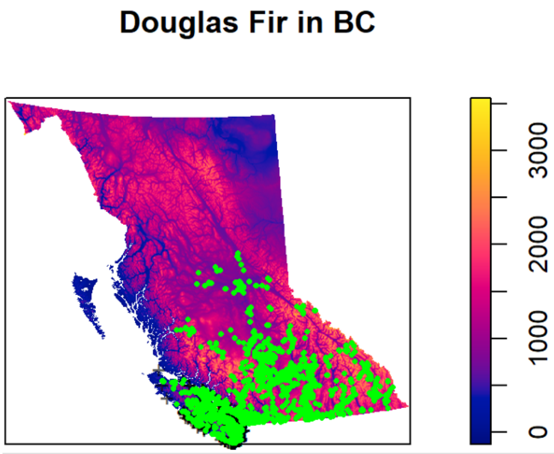

Figure 2: Spatial Pattern

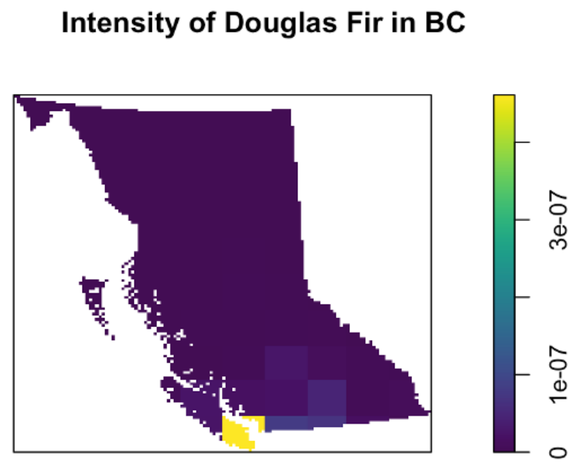

Figure 3: Intensity Plot

by splitting the whole window into 10-by-10 quadrats. The p-value of the test is less than 0.05, suggesting the intensity of Douglas-fir in BC is inhomogeneous. Based on our hotspot analysis, the hotspot (the blue spots in the Figure 4) shows that the intensity is significantly higher in Vancouver Island and the southwestern part of BC.

Our second moment analysis provided the correlations between each data point by pair correlation function. The following Figure 5 shows that Douglas-fir exhibits clustering at extremely small distances ($< 5,000m$) and avoidance between 5,000m to 20,000m. The distribution of Douglas-fir is independent at larger distances.

3.1 Model Building

To better build the model, we first examined the relation between the intensity of Douglas-fir and each covariate. The result of each covariate is displayed below (Figure 6).

Based on the correlations between covariates and intensity, we initially assumed that the intensity of Douglas-fir has a quadratic relation with elevation, an exponential relation with HFI, a quadratic relation with forest cover, and a linear relation with distance to water. However, the quadratic assumption of forest cover did not well explain the proposed model, thus we instead suggested it as a linear form and the modified model-fitting surprisingly provided significantly statistical results with 5% significance level as shown in the Table 1. Finally, our fitted model is:

$$\lambda(u) = e^{-0.1-3.8*10^{-3}elevation+16.5HFI-1.1*10^{-4}distwater+9.0*10^{-7}elevation^2+1.0*10^{-4}forest^2-6.8e^{HFI}}$$

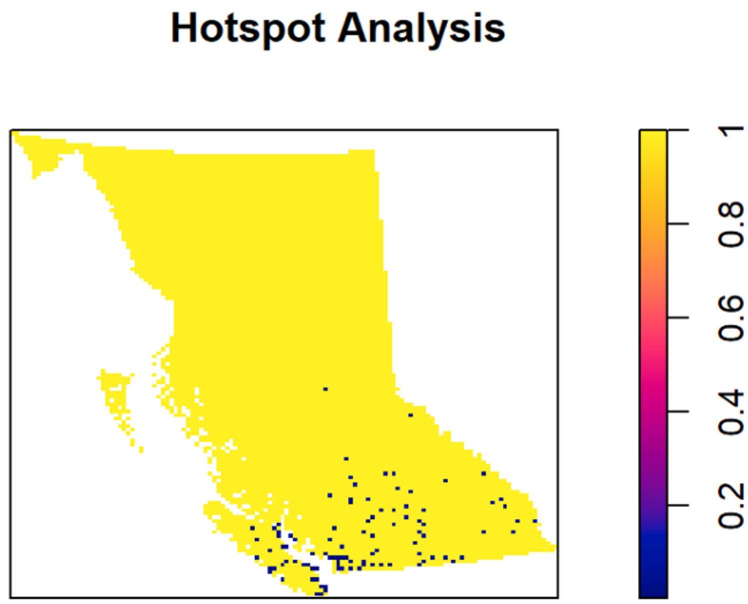

Figure 4: Hotspots

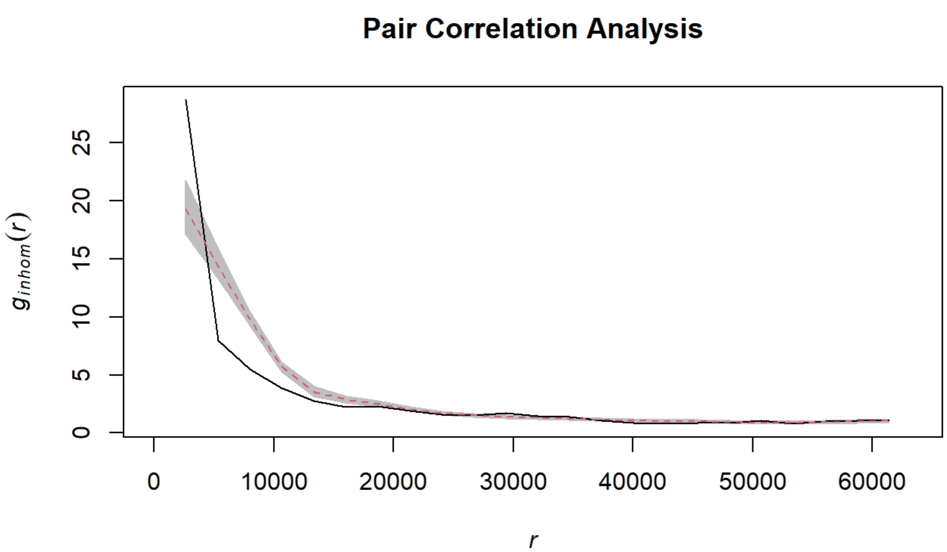

Figure 5: Pair Correlation Function Plot

The predicted intensity (Figure 7) shows that our model accurately captures the high-intensity regions of Douglas-fir in the southwestern and southern areas of BC, aligning closely with the true intensity graph. However, it also shows an overestimation of intensity in the northeastern BC, an area devoid of Douglas-fir. To further examine the accuracy of our model, we visualized the residuals (Figure 8) and zoomed in on the southwestern part of BC (Figure 9). The graphs suggest that our model accurately captures the intensity of Douglas

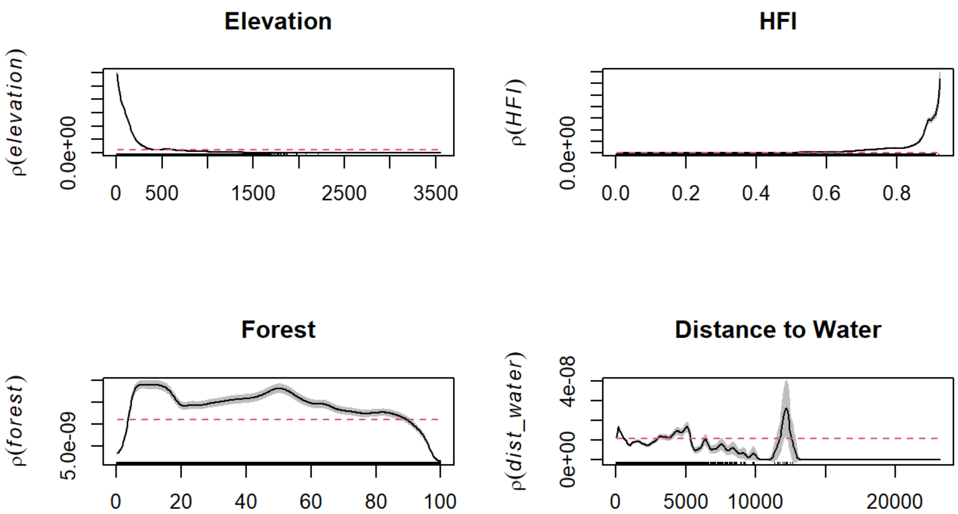

Figure 6: Correlations between Rho Estimate of Intensity with Covariates

	Estimate	S.E.	CI95.lo	CI95.hi	Z test	Z val
(Intercept)	-11.49	0.22	-11.93	-11.06	***	-51.67
$elevation$	-0.004	0.00	-0.004	-0.004	***	-51.56
HFI	16.45	0.38	15.70	17.19	***	43.31
$dist\_water$	-0.0001	0.00	-0.0001	-0.0001	***	-18.15
$elevation^2$	0.00	0.00	0.00	0.00	***	18.29
$forest^2$	0.0001	0.00	0.0001	0.0001	***	27.66
$exp(HFI)$	-6.81	0.24	-7.27	-6.34	***	-28.66

Table 1: Results Summary for Nonstationary Poisson Process

Fir in most areas of BC, but underestimates the intensity in southwestern and southern parts of BC. These discrepancies suggest that while our model performs well overall, there are limitations in its predictive accuracy, highlighting the need for further refinement and validation.

3.2 Model Validation Results

We validated our model's performance through various analyses. Firstly, a chi-squared test conducted across 15 quadrats revealed a significant deviation of predictions from actual values, with a p-value below 0.05. This underscores the need for enhancements in our model's predictive capabilities.

Further investigation via partial residual analysis of each covariate shed light on our model's strengths and areas for improvement (Figure 10). Notably, our model adeptly cap-

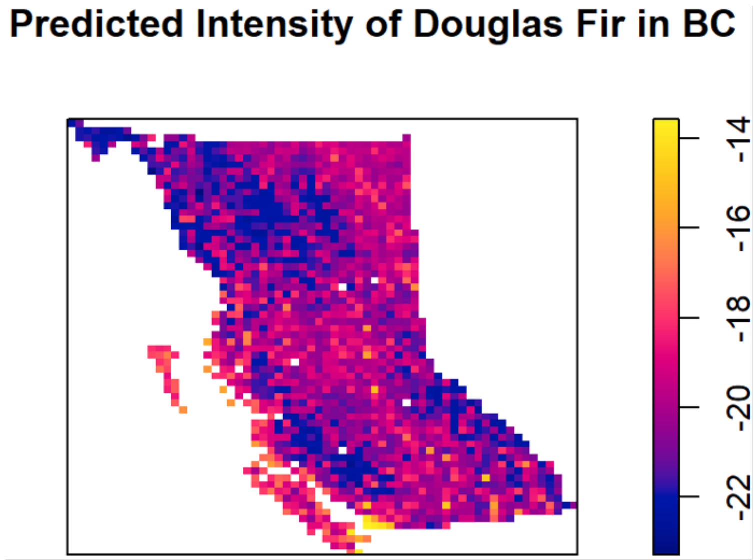

Figure 7: Prediction of Intensity by Proposed Model

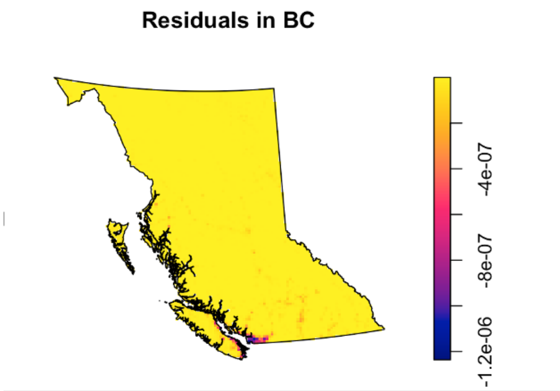

Figure 8: Residuals

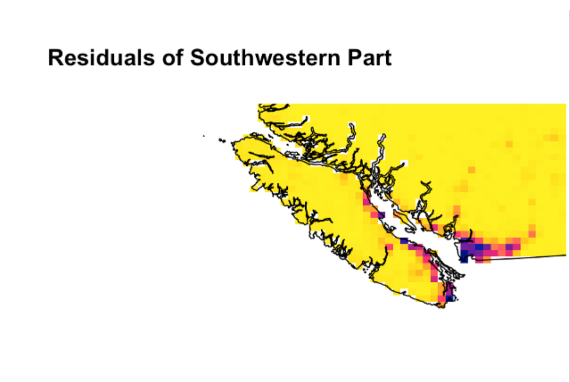

Figure 9: Zoom-in Residuals

tured the influence of elevation and HFI on Douglas-fir intensity, as evidenced by the alignment of the fitted line (blue dotted) with the true pattern (black solid). Although our model exhibited deviations in the intermediate values of distance to water, it generally captured this covariate's impact. However, our model struggled to reflect the forest's influence accurately, indicating room for refinement in this aspect. Attempts to address this included employing higher polynomials and GAM; however, these efforts yielded minimal improvements in terms of AIC or led to overfitting, making the coefficients less interpretable.

In summary, while our model demonstrated a certain level of accuracy in predicting Douglas-fir intensity and effectively captures trends in elevation, HFI, and distance to water, further enhancements are necessary, particularly in modeling the forest's influence.

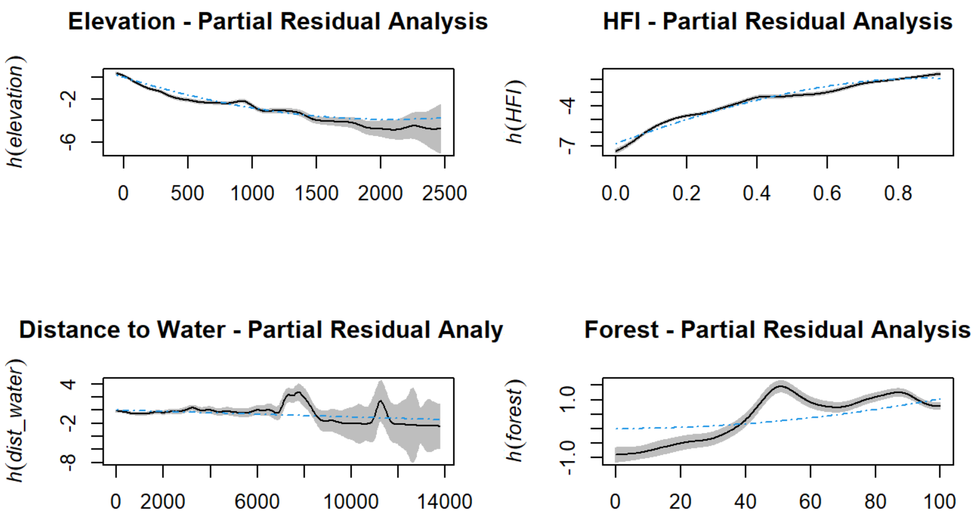

Figure 10: Partial Effects of Covaraites on the Predicted Intensity

4 Conclusion and Discussion

Through our investigation and analysis of the spatial distribution of Douglas-fir in BC, we have gained valuable insights into its point pattern to answer our research questions. Our findings revealed several key observations including:

- **Inhomogeneous Intensity**

  The intensity of Douglas-fir in BC shows significant spatial variation, concentrating in the southwestern and southern parts of BC, especially on Vancouver Island. Additionally, we observed significant clustering at distances less than 5,000 meters, indicating preferential habitat areas. Conversely, there is a trend of avoidance at distances between 5,000 meters and 20,000 meters, suggesting areas less conducive to Douglas-fir growth.

- **Relationship with Covariates**

  Elevation, HFI, forest, and distance to water are crucial factors influencing the intensity of Douglas-fir. Incorporating these variables into the modeling process allows for a more comprehensive analysis of Douglas-fir distribution patterns and habitat preferences.

- **Poisson Modeling**

  We developed a Poisson model to quantitatively describe the relationship between Douglas-fir intensity and four key covariates: Elevation, HFI, Distance to Water, and Forest. The model is represented by the equation:

  $$\lambda(u) = e^{-0.11-3.76*10^{-3}elevation+16.45HFI-1.05*10^{-4}distwater+9.01*10^{-7}elevation^2+1.03*10^{-4}forest^2-6.81e^{HFI}}$$

This model allows us to predict Douglas-fir intensity with a reasonable level of accuracy. However, we have identified a limitation where the model does not precisely capture the relationship with the Forest covariate.

To enhance the accuracy and robustness of our modeling approach, we plan to incorporate more potential covariates that may be relevant to Douglas-fir distribution. These could include slope, soil characteristics, precipitation patterns, and temperature variations. By integrating these new covariates into our model, we aim to refine our understanding of Douglas-fir distribution and improve the fitted performance of our models. This expanded dataset and refined modeling approach will contribute to a more comprehensive analysis of Douglas-fir habitat preferences and population dynamics in BC.

References

[1] GBIF Secretariat, “Pseudotsuga menziesii (mirb.) franco in gbif backbone taxonomy,” Checklist dataset, 2023, accessed via GBIF.org on 2024-04-28. [Online]. Available: https://www.gbif.org/species/5283989

[2] Board of Trustees of the Royal Botanic Gardens, Kew, “Pseudotsuga menziesii var. menziesii,” Plants of the World Online, 2017, retrieved 5 November 2020.

[3] British Columbia Ministry of Forests, “Douglas fir tree species compendium,” Online, 2024, accessed on 2024-04-28. [Online]. Available: https://www2.gov.bc.ca/gov/content/industry/forestry/managing-our-forest-resources/silviculture/tree-species-selection/tree-species-compendium-index/douglas-fir

[4] S. F. Arno and R. P. Hammerly, *Northwest Trees.* Seattle: Mountaineers, 1977, nature.

[5] P. W. Keys *et al.*, “Title of the article,” *Environmental Research Letters*, vol. 16, p. 044061, 2021.

[6] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R.* London: Chapman and Hall/CRC Press, 2015. [Online]. Available: https://www.routledge.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/