Dissertation for the degree of

MSc in Earth Observation & Geoinformation Management

Benjamin McLellan
August 2025

Statement of Copyright and Originality

I declare that this dissertation represents my own work, and that where the work of others has been used it has been duly accredited.

I further declare that the length of the components of this dissertation is 4,972 words for the Research Paper and 3,249 words for the Technical Report.

Copyright of this dissertation is retained by the author and The University of Edinburgh. Ideas contained in this dissertation remain the intellectual property of the author and their supervisors, except where explicitly otherwise referenced.

All rights reserved. The use of any part of this dissertation reproduced, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a retrieval system without the prior written consent of the author and The University of Edinburgh (Institute of Geography) is not permitted.

I agree that this dissertation and associated electronic documents, web pages, data, files and computer programs which are not limited by the signed data agreement can be retained by the University.

YES

I agree that, with the permission of my supervisor(s) or the Programme Director, these materials be made available for the purposes of preparing a publication.

YES

I agree that, with the permission of my supervisor(s) or the Programme Director, these materials can be used within the University of Edinburgh for continued research or teaching.

YES

Benjamin McLellan

August 2025

Acknowledgements

First and foremost, many thanks go to my supervisor Dr Steven Hancock for his time, reassurance, and continuous support throughout this project. Also, for joining me for some very wet data collection!

I would also like to thank Dr Calum Brown and Cathy Atkinson from Highlands Rewilding for their expert support and guidance, and for making this dissertation possible.

Further thanks go to Dr Christopher Ellis from the Royal Botanical Garden Edinburgh for introducing me to the fascinating world of lichen and bryophytes, and for taking me with him to see some in the wild.

Thanks also to Archie Morgan, Oscar Conway, and Oscar Jessup for making the first trip to Tayvallich possible (and enjoyable!).

Finally, I would like to thank Ciara Mitchell for her constant support, for letting me bore her endlessly about all things temperate rainforest, and for putting up with the lichen and bryophytes in the kitchen.

Part I: Research Paper

The Influence of Lidar-Derived Forest Structure and Geographic Variables on Temperate Rainforest Quality: A Case Study on Tayvallich, Scotland.

Abstract

Temperate rainforests are globally rare biodiversity hotspots containing unique species of lichen and bryophyte species. With the required conditions existing on less than 1% of the Earth's surface, Scotland has some of the best examples of temperate rainforests globally. However, these woodlands have been fragmented by overgrazing, invasive species, and deforestation. Expansion and regeneration of these forests is therefore essential for achieving national biodiversity targets. Whilst national drivers of biodiversity and favourable tree species have been identified, the aim of this study is to identify site-specific drivers which can directly inform planting and management strategies. This is achieved by exploring how lidar-derived woodland structure and geographical variables influence temperate rainforest quality, focussing on a site in Tayvallich, Scotland. Quality across 64 sites was assessed using two indices: (1) a standardised rainforest assessment methodology, to measure overall woodland condition, and (2) species richness of unique temperate rainforest lichen and bryophytes, a direct measure of ecological response. Using generalised additive models, canopy cover and proximity to water were identified as the most important variables to consider in planting and management strategies. Proximity to a biodiverse species pool was also found to be highly significant, but further research is required to confirm the role of this variable. Based on these findings, recommendations are made for how to best create and enhance biodiversity in these globally rare ecosystems.

Contents

Statement of Copyright and Originality i

Acknowledgements ii

Part I: Research Paper 1

- Abstract 1
- Introduction 3
- Methods 5
  - Study Site 5
  - Airborne Laser Scanning 7
  - Data and Preprocessing 8
  - Structural and Spatial Metrics 8
  - Quality Indices 9
  - Data Analysis 12
- Results 14
  - General 14
  - Significant Metrics 15
- Discussion 19
  - The Role of Structural and Spatial Variables 19
  - Implications for Management 21
  - Limitations 22
  - Future Research 23
- Conclusion 24
- Research Paper References 25

Part II: Technical Report 1

Introduction

Temperate rainforests are a globally rare ecosystem due to the restrictive bioclimatic conditions that define them: over 1,400 mm annual precipitation, a summer isotherm under 16 °C, and at least 10% of the annual precipitation occurring during the summer (Alaback, 1991). These conditions exist on less than 1% of the Earth's surface, with 15% of this area situated in Europe (DellaSala, 2011). Of this European extent, 40% lies within the British Isles where the oceanic climate provides the aforementioned bioclimatic conditions, as seen in Figure 1 (Ellis, 2016). As highlighted, much of the of the UK's temperate rainforest zone is located in western Scotland.

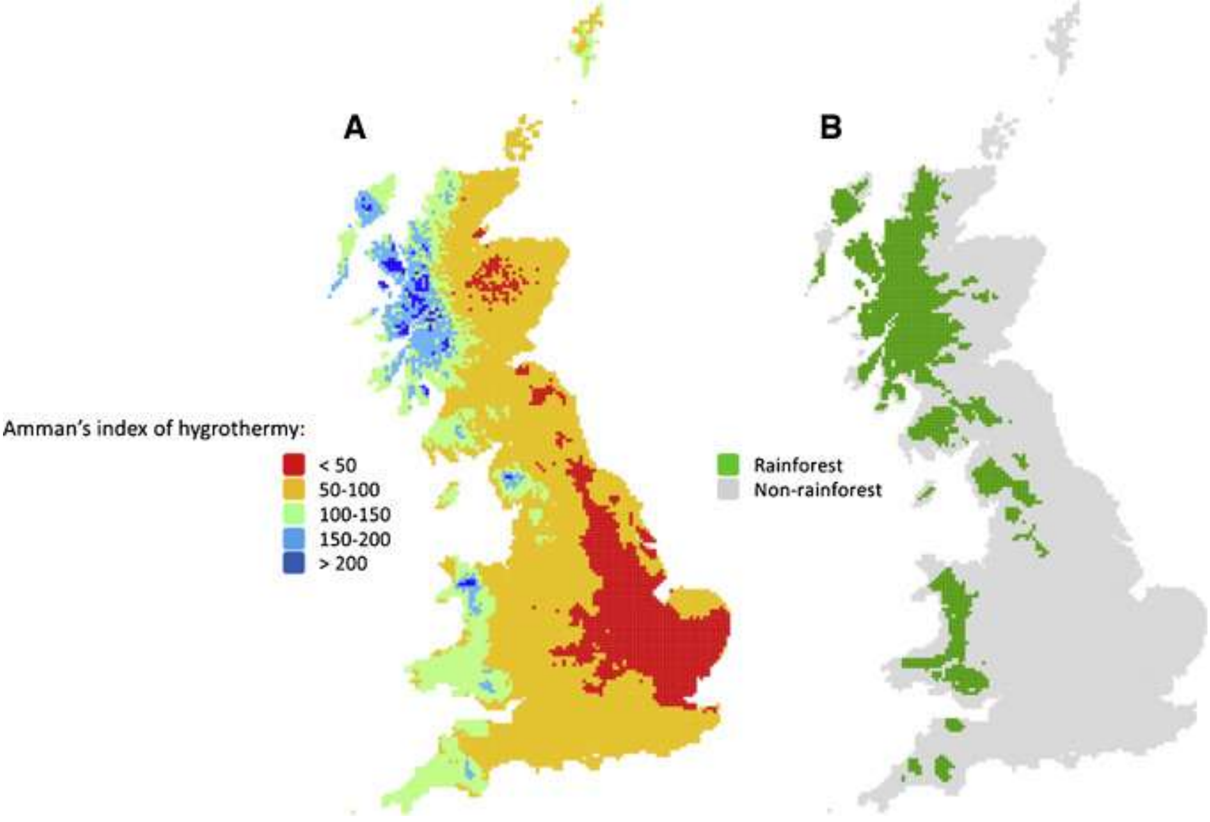


Figure 1: From Ellis (2016). A) highlights Amman's index of hygrothermy, suitable for mapping oceanic (>100) and continental (<100) conditions. B) highlights the temperate rainforest zone in the UK.

Scottish temperate rainforests are of particular ecological concern as they are characterised by unique species of globally rare lichen and bryophytes, making them highly biodiverse ecosystems (Ellis and Eaton, 2016). Lichens and bryophytes are 'poikilohydric', meaning they do not actively regulate their water like vascular plants, responding directly to ambient environmental conditions. Consequently, these species require very specific conditions for survival, making them sensitive to microclimatic conditions and niches. Because of this, they have been used indicator species for woodland condition (Ellis et al., 2015) and, in this study, high-quality temperate rainforests are defined as those with the optimum conditions for these species.

However, in Scotland high quality temperate rainforest only remains in fragments as a result of overgrazing, invasive species, and deforestation for timber and land-use change (Woodland Trust, 2021; Ellis and Eaton, 2016). Consequently, the regeneration and expansion of these ecosystems is essential for promoting biodiversity and meeting national conservation targets, such as those highlighted in NatureScot's 2045 Scottish Biodiversity Strategy (NatureScot, 2022).

Understanding how and where to plant in regeneration and expansion projects is therefore important for creating high-quality temperate rainforest that promotes biodiversity. Exploring how woodland structure and environmental and geographic (spatial) variables impact forest quality can help achieve this. Land managers can influence *how* forests are structured through planting design and management practices, such as woodland thinning or coppicing, and spatial variables can help identify *where* conditions are most suitable for successful restoration.

Previous research has explored the role of such variables in predicting indicator species richness across Scottish temperate rainforests. Focussing on 20 woodlands, Ellis and Eaton (2021) examined how variables such as tree species, woodland structure, climate, topography, and distance to water influence large-scale indicator species richness. They identified national drivers of richness, finding that tree species, climate, and proximity to water were the most important predictors, whilst structural variables were not important at this scale.

Based on these findings, they suggested the optimum tree species for planting and management as those with higher bark pH: *Fraxinus excelsior, Corylus avellana, Salix* spp., *Sorbus aucuparia,* and *Ulmus* spp. Further research has identified a negative influence of non-native conifers, reinforcing the importance of woodland composition (Broome et al., 2021).

However, whilst favourable tree species and national-scale drivers of richness have been identified, research has not explored how structural and spatial variables influence rainforest quality at finer scales. This highlights a critical knowledge gap: understanding how finer-scale, site specific variables affect forest quality is essential for informing practical planting and management strategies that support biodiversity in these rare and threatened ecosystems.

Therefore, the aim of this project is to identify important structural and spatial metrics at this scale, understand their influence on temperate rainforest quality, and apply this understanding to guide evidence-based restoration and management.

Focussing on a site in western Scotland, Airborne laser scanning (ALS) data is used to derive fine-scale and site-specific structural and spatial metrics. Quality is defined using two complimentary indices: (1) a standardised rainforest assessment, to understand general habitat suitability for lichen and bryophytes; and (2) indicator species richness, representing the diversity of important temperate rainforest species.

Methods

Study Site

This study focusses on woodland sites across Highland Rewilding's Tayvallich Estate (Figure 2). Figure 2b highlights the proximity of Taynish National Nature Reserve (NNR), a temperate rainforest of special scientific interest (SSSI) (NatureSoct, 2011). Tayvallich receives 1545 mm rainfall in a typical year, an annual average high and low temperature of 12.4°C and 6.4°C, and a July isotherm of 14.75°C (Gazetteer for Scotland, 2021). In addition to temperate rainforest, it contains many different habitats: a national vegetation classification (NVC) identified 69 NVC communities and 1,743 sub-communities, 902 of which were mosaics of two or more vegetation types (Highlands Rewilding, 2024). Figure 3 highlights the land-use classifications according to the 2023 Land Cover Map (UKCEH, 2023). In addition to the complex vegetation, the site also has highly varying topography, with many small parallel ridges separated by glens and hollows.

Temperate rainforest exists in fragments across the estate, with dominant tree species including, *Betula* spp., *Quercus* spp., and *Corylus avellana*. *Fraxinus excelsior*, *Salix* spp., and *Alnus glutinosa* are also present throughout. There is also productive non-native conifer woodland, dominated by *Picea sitchensis* and *Larix decidua* which encroach on areas of native temperate rainforest.

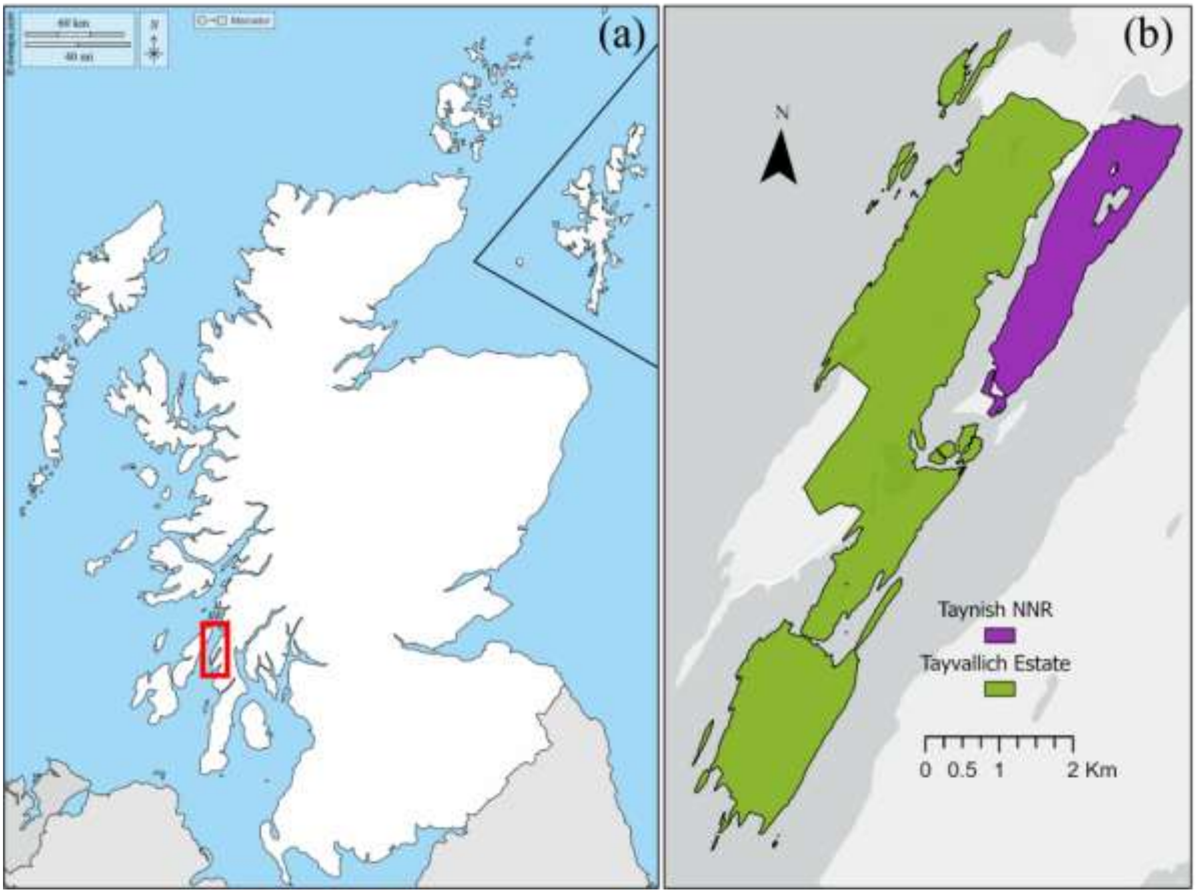


Figure 2: (a) The general study area within Scotland (d-maps, 2025) and (b) Tayvallich estate and Taynish woods National Nature Reserve (NNR).

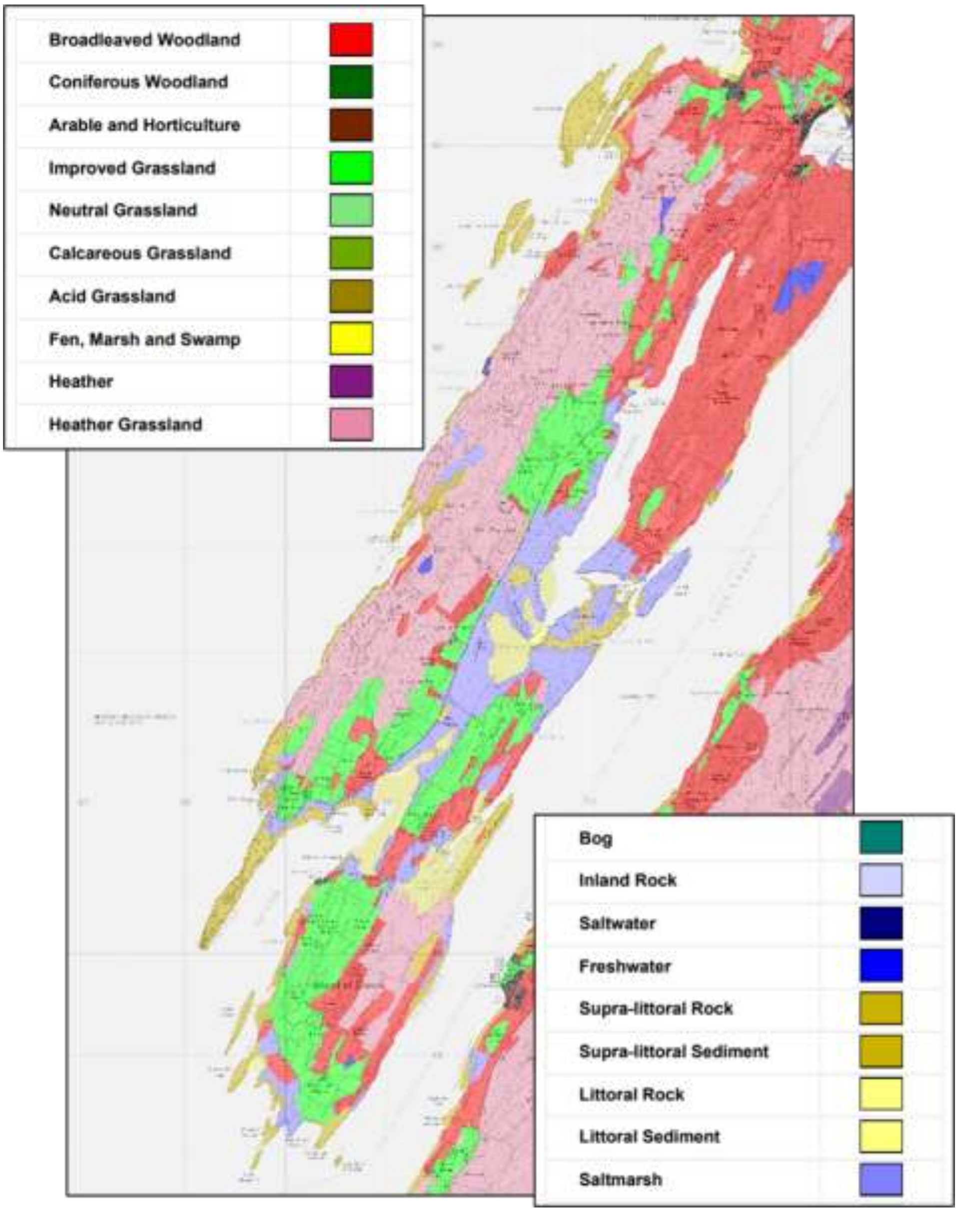


Figure 3: Land classification for Tayvallich, according to UKCEH (2023).

Airborne Laser Scanning

The development of ALS has greatly improved our ability to describe, monitor and map structural aspects of vegetation. ALS utilises light detection and ranging (lidar) technology, an active remote sensing tool which can measure the 3-D location of targets. This is achieved by emitting pulses of near-infrared energy and measuring the time taken for them to return to the system. Using an onboard Global Navigation Satellite System (GNSS) and an Inertial Measurement Unit (IMU), to account for platform orientation, the target's location can be recorded with decimetre accuracy (Coops et al., 2016).

Consequently, ALS data can be used to accurately map forests by recording the distribution of returns reflected off the stems, branches and foliage, from the canopy top to the forest floor. As such, ALS has been widely used in research for quantifying structure at forest, stand, and individual tree levels (Coops et al., 2016; Naesset, 2002; Yu et al., 2011). Pulses which reach the ground can also be used to create high resolution digital terrain models (DTMs) (Kopecky et al., 2021).

Data and Preprocessing

The data used in this project was provided by Highlands Rewilding. The ALS survey was conducted on 9th June 2023, recording leaf-on forest structure. The scanner used was the Riegl VQ780i which provided approximately 10 points per square metre, stored in 3-D point clouds. Data covers the 5 extents presented in Figure 4. Preprocessing was required to create a DTM, a digital surface model (DSM), and to normalise the point clouds.

Structural and Spatial Metrics

As introduced, structural and spatial metrics were created to understand *how* and *where* expansion, reforestation, and afforestation should take place. Previous research has grouped structural metrics into 3 categories: (1) metrics focused on the overall canopy height; (2) metrics representing cover of the horizontal arrangement of biomass; and (3) metrics highlighting the vertical distribution of biomass (Coops et al., 2016). To ensure all aspects of structure were represented, metrics from each of these categories were calculated.

Spatial metrics include climatic and geographic metrics, which either influence microclimatic conditions in a plot or are associated with lichen and bryophyte survival and colonisation requirements. The metrics created and what they show can be seen in Table 1. All metrics were

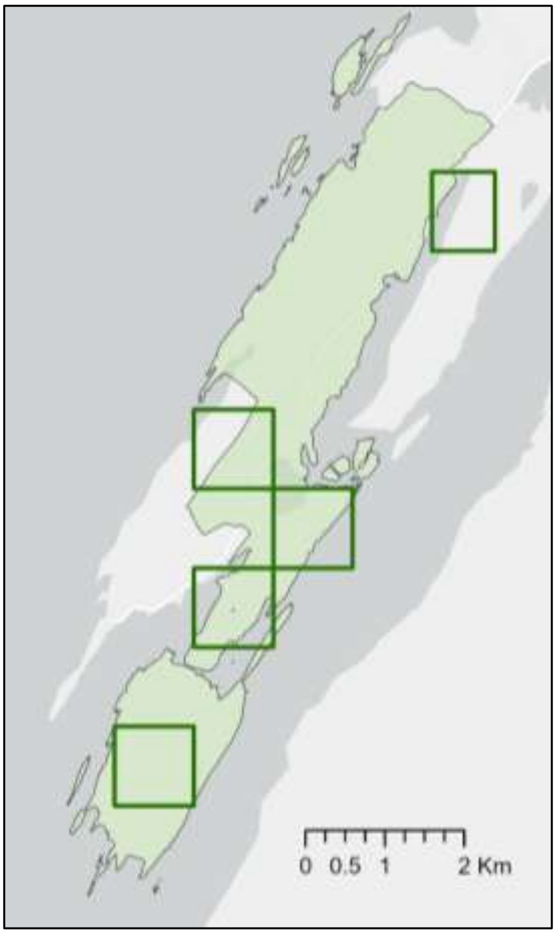


Figure 4: Map highlighting the extents of the 5 point clouds used in this study.

calculated at 15 m resolution. Please see the Technical Report for more information regarding the creation and justification of the metrics used.

Table 1: Important information regarding the structural and environmental metrics created. All calculated at 15 m resolution.

Metric	Description
Mean canopy height	Mean return height per pixel.
Max canopy height	Max return height per pixel.
Canopy cover	Percentage of returns above height threshold (calculated at 2 m and 6 m).
Canopy cover variance	Standard deviation of canopy cover per pixel.
Vegetation below 4 m	Percentage of points between 0.3 and 4 m.
Effective number of layers (ENL)	Number of distinct layers in a plot's vertical profile.
Height variance	Standard deviation of height values per pixel.
Slope/aspect/elevation	Slope, aspect and elevation per pixel.
Topographic wetness index (TWI)	Area wetness based on runoff and pooling.
Topographic position index (TPI)	Topographic position based on surrounding area. Calculated at multiple scales (30, 90, and 500 m).
Wind exposure	Categorical exposure index (protected/exposed).
Solar radiation	Annual and growing season solar radiation.
Proximity to species pool	Proximity of plots to Taynish, a biodiverse species pool located to the west of Tayvallich (Figure 2b).
Distance from water	Distance from standing water, major rivers and/or the sea.

Quality Indices

Rapid Rainforest Assessment

The Rapid Rainforest Assessment (RRA) is a standardised methodology created by Plantlife (2023a). Originally developed for land managers to understand overall woodland condition, the objective of the RRA is to assess the habitat suitability for the unique temperate rainforest lichen and bryophyte species introduced earlier. It was deemed appropriate for this study as it provides a comprehensive coverage of important habitat features, quantifying overall habitat condition and suitability.

The RRA aims to assess the potential suitability of a woodland plot for lichen and bryophyte colonisation and survival (Plantlife, 2023a). It is broken down into several categories which assess the species' habitat requirements: (i) Woodland composition: focusing on the tree species mix; (ii) Woodland structure: estimating tree age, canopy cover, and field layer diversity; (iii) Habitat features: recording veteran tree features, dead wood, rocky substrates, and any wet features; (iv) Lichen and bryophytes: noting the coverage of lichen and bryophytes in the woodland; (v) Grazing and browsing: estimate grazing and browsing levels using specified indicators; (vi) Invasive species: identifying and the noting the abundance of any native and/or non-native invasive species; and (vii) Ash dieback: for forests where ash is present, determine the potential risk for lichen and bryophyte species.

Although woodland structure is considered within the RRA, it is assessed in a broad and observational manner, estimating light availability and the amount of open space. As such, it does not capture the full complexity of vertical or horizontal structure that ALS can quantify.

A score is provided for each of the seven categories, added together to get an overall RRA score. This can be interpreted using the provided quality classes, highlighted in Table 2.

Table 2: The different Rapid Rainforest Assessment (RRA) score quality classes (2023b).

Score	Class	Description
25+	Very good	Suggests a site that is currently in very good condition to support a range of temperate rainforest lichens and bryophytes. There might still be management issues that need addressing, and attention should be focussed on conditions around important habitat features and species, to ensure these remain favourable.
15-24	Good	Suggests a site that is currently in good condition to support some temperate rainforest lichens and bryophytes. However, there are aspects that could be improved and there may be a number of management issues that need addressing.
5-14	Fair	Suggests a site that has potential to support temperate rainforest lichens and bryophytes but current condition is not optimal. There are a range of management issues that need to be addressed, likely including restructuring, grazing adjustments and management of invasive species.
<5	Poor	Suggests a site that is not currently in good condition to support temperate rainforest lichens and bryophytes. There are a range of complex and extensive management issues that need to be addressed, likely including considerable restructuring, grazing adjustments and management of invasive species.

Indicator Species Richness

Although the RRA score provides a comprehensive assessment of overall woodland condition, it does not explicitly capture the extent of temperate rainforest indicator species. Therefore, indicator species richness was measured alongside the RRA, using the lichen and bryophyte species highlighted in their guides (Plantlife, 2023b; Plantlife, 2024a; Plantlife, 2024b).

Lichen and bryophyte richness are widely recognised as indicators of habitat condition and have been frequently used to assess temperate rainforest condition (Ellis et al., 2015; Ellis and Eaton, 2021; Brosnan and Ellis, 2020).

However, using richness alone could misrepresent quality in plots where one substrate has a high richness but the plot is generally of poor condition. Therefore, overall woodland quality in this study is defined using both metrics: RRA score provides a general overview of habitat condition and suitability, whilst indicator species richness reflects current ecosystem condition.

Sampling Method

Data was collected on the 8th and 9th June 2025. Sample sites were selected using random stratified sampling to ensure that sites with varying structures were sampled. 16 sites were sampled, with each site being split into 4 quadrants (hereby referred to as plots). Plots were 15 by 15 m. In each of the 64 plots, the RRA was conducted and the number of indicator species on substrates up to 2 m were recorded. Sites were sampled working anti-clockwise and starting in the northwestern plot, as highlighted in Figure 5. A handheld GNSS was used to geolocate the centre of each site.

Tree Type

As non-native conifers dominated some of the plots, it was important to account for their influence in the data analysis. Therefore, the dominant tree type (conifer or broadleaf) was recorded in each plot.

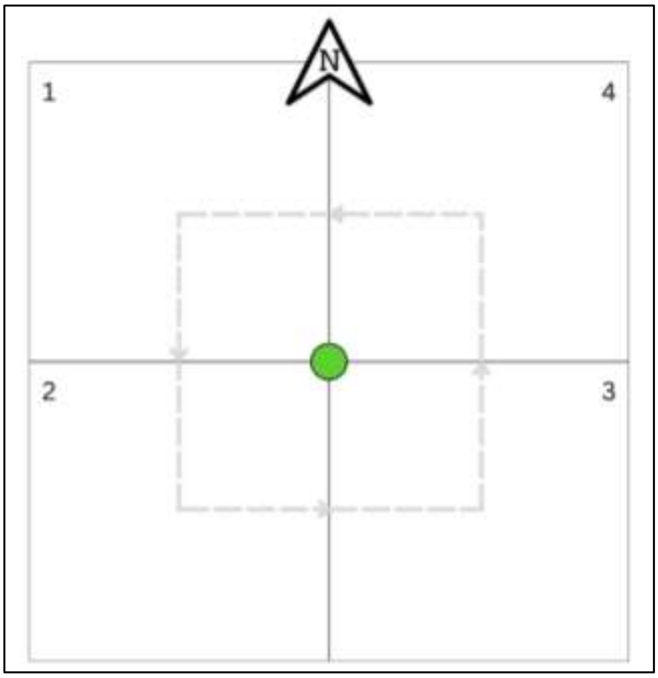


Figure 5: Example site survey method. Quadrants are 15 by 15 m.

Data Analysis

Generalised Additive Models

Generalised Additive Models (GAMs) were used to explore how the structural and spatial metrics influence woodland quality. GAMs are statistical models which can be used to reveal complex relationships between explanatory variables and a response variable. This is achieved by fitting flexible curves (smooth functions) to the explanatory variables, highlighting complex relationships that would be missed by linear models.

Since GAMs are additive, they present the effect of each variable on the response whilst accounting for the other terms in the model. For example, a model including canopy cover and elevation would highlight any effect of canopy cover on forest quality whilst accounting for elevation (and vice versa). GAMs can also include interaction terms, which explore how combinations of variables jointly influence the response. For instance, a model including an interaction between canopy cover and solar radiation would explore how the effect of canopy cover depends on solar radiation, whilst still accounting for other terms in the model. The influence of a term can be seen in its corresponding partial effect graph, which highlights the relationship between the explanatory variable and the response.

By including coordinates of data points as an interaction term (referred to as the spatial smooth), GAMs can quantify the influence of spatial patterns and autocorrelation. This allows the user to understand the extent to which terms are independent of location. For example, a GAM in which the partial effect of location is non-significant suggests that variance in the response has been explained by the other terms in the model, whereas a GAM in which the effect of location is significant suggests location is still a dominant factor (Wood, 2017). Because of this explanatory power, they are frequently applied in spatial ecology research where relationships are often non-linear and difficult to represent in traditional models (Yee and Mitchell, 1999; Frescino et al., 2001; Kosicki, 2020). For these reasons, GAMs were deemed the most appropriate model for the purpose of this study.

Correlation Matrix

To mitigate multicollinearity in the model, a Pearson correlation matrix was created. Variables with a correlation $r>0.7$ would only be considered for inclusion in the same model if they offer substantially different information, and their effect on the model was monitored closely.

Model selection

Before complex models were created, individual GAMs were created for each of the structural metrics whilst accounting for spatial effect and tree type. For spatial metrics, individual GAMs were created with and without the spatial term to identify variables which account for spatial effect and are significant independently. Interaction terms were then explored and complex models were built based on the findings from the individual models. Please refer to the Technical Report for more information on model building and selection.

Cross Validation

To assess the robustness and predictive performance of the models, a leave-one-site-out cross-validation (LOSO) approach was used for all GAMs. Specifically, for each iteration, all plots from one site were excluded whilst the model was trained using data from the remaining sites. The fitted model was then used to predict values for the left-out site, and the predictive error was recorded. This process was repeated so that each site served as the validation set once, ensuring that all spatial clusters were withheld in turn. Cross-validated performance metrics, including root mean squared error (RMSE) and mean absolute error (MAE) were calculated by averaging results across all sites. These metrics provide an honest measure of how well the model performs on unseen data, highlighting potential overfitting.

Model Performance

Table 3 summarises the values used to assess a model's performance and a brief description of what they show. These values should be interpreted alongside the partial effect graphs to get a full understanding of a variable's influence.

Table 3: The values used to determine the relative contribution of terms in a model and evaluate the model's performance.

Statistic	**Description**
p value	Statistical significance of each predictor and smooth term; values below 0.05 indicate strong evidence for an effect.
EDF	Effective degrees of freedom for each smooth term; reflects model complexity and the 'wiggliness' of fitted relationships. An EDF of 1 denotes a linear effect.
Adjusted R^2	Proportion of variance in the response explained by the model (adjusted for model complexity).
Deviance Explained	Percent deviance explained by the model; analogous to R^2 for generalized additive models, measuring goodness-of-fit.
AIC	Akaike Information Criterion; a measure of model fit that penalizes complexity, lower values indicate a better balance of fit and parsimony.
CV RMSE	Average magnitude of prediction error on left-out data, lower values indicate better predictive performance. Units are the same as the response variable.
CV MAE	Average absolute difference between observed and predicted values in cross-validation, more robust to outliers than RMSE. Units are the same as the response variable.
Partial Effect Graphs	Display the effect of a variable after controlling for other terms in the model. Y-axis values are relative and show how response varies according to the explanatory variable, holding other variables constant.

Results

General

The mean RRA score was 6.47 (fair), the highest was 25 (very good), and the lowest score was -9 (poor). The mean richness was 7.56, with the highest richness being 16. This was found at the same site with the highest RRA score. In fact, RRA score and indicator richness are highly correlated (Spearman's $\rho = 0.78$).

It is also worth stating that the majority of the lichen species identified belong to the Parmelion group. Whilst members of the Lobarion group were present at some sites, they were far less frequent and species such as *Lobaria Pulmonaria* and *L. virens* were not identified at any of the plots. *L. pulmonaria* was observed in areas of woodland outside of the plots.

Significant Metrics

Structural Metrics

Table 4 highlights the explanatory power of the significant structural metrics whilst accounting for tree type and the spatial smooth. As highlighted, canopy cover over 2 m had the highest explanatory power for both RRA score and richness. All terms in the table had a linear relationship (EDF = 1), which was positive for all except canopy cover variance.

Table 4: Performance of GAMs including individual structural metrics, tree type, and the spatial smooth.

Metric	**p value**	**Adj. R^2**	**Dev. Explained (%)**
Canopy Cover (> 2 m)	RRA < 0.001 Rich = 0.005	RRA = 0.709 Rich = 0.537	RRA = 76.6 Rich = 59.6
Canopy Cover (> 6 m)	RRA < 0.001 Rich = 0.035	RRA = 0.693 Rich = 0.507	RRA = 75 Rich = 57.2
Mean Height	RRA = 0.011 Rich = 0.013	RRA = 0.581 Rich = 0.513	RRA = 61.4 Rich = 58.3
Canopy Cover Variance	RRA = 0.014 Rich = 0.022	RRA = 0.579 Rich = 0.515	RRA = 61.2 Rich = 57.7

Since canopy cover over 6 m and mean height were highly correlated with canopy cover over 2 m (r = 0.83 and 0.8 respectively), these metrics could not be used in the same model. Although not strongly correlated (r = -0.32), an additive model including canopy cover over 2 m (hereon referred to as canopy cover) and canopy cover variance highlighted multicollinearity in their explanatory power: both terms were no longer significant, suggesting shared explanation of response variance.

Given the structural metrics' high multicollinearity and similar influence on forest quality, canopy cover was selected as the most important variable, with the other metrics not contributing unique explanation of variance.

No interaction terms were significant after accounting for the effect of the individual terms.

It is worth mentioning that before accounting for tree type, canopy cover did not have a linear relationship and an optimum canopy cover of between 75 and 80% was identified for both quality indices.

Spatial Metrics

In their individual GAMs, the only significant spatial variables when including the spatial smooth were TPI calculated at 500 m (hereon referred to as TPI) (RRA p value = 0.007, richness $p < 0.001$) and proximity to standing water (RRA $p = 0.018$, richness $p < 0.001$). The TPI model identified that areas with lower TPI were of higher quality than high areas.

Distance from Taynish, standing water, and the sea all reduced the explanatory power of the spatial smooth and thus were explored independently. All of which were highly significant for both RRA Score and richness ($p < 0.001$) and had similar explanatory power. Additive models and the correlation matrix revealed multicollinearity between the distance-based metrics, whereas TPI was no longer significant after accounting for each of the distance-based metrics.

Unlike with the structural metrics, it was not possible to identify the most important spatial metric as they represent distinct ecological processes and thus would have unique influences on habitat quality.

As with the structural metrics, no interaction terms were significant after accounting for the individual terms' effects.

Unlike with the structural metrics, it was not possible to identify the most important spatial metric as they represent distinct ecological processes and have similar explanatory power.

Combined Models and Interactions

Additive models exploring the effects of the distance based metrics and canopy cover whilst accounting for one another were then explored. A model including the spatial smooth, tree type, canopy cover, and TPI was also created.

Canopy cover, distance from Taynish, the sea, and standing were the most important variables, independent of tree type, location, and the other variables. TPI 500 was no longer significant after accounting for the effect of canopy cover.

Table 5 and 6 highlight the best-performing models, which have similar performance on unseen data and limited overfitting given the small data set.

Table 5: Model performance for RRA score. TAY = distance from species pool (Taynish), CC = canopy cover above 2 m, SW = distance from standing water, and SEA = distance from the sea.

Terms	**p values**	**EDFs**	**Adj. R^2**	**Dev. Explained (%)**	**AIC**	**CV RMSE**	**CV MAE**
TAY and CC	TAY < 0.001; CC = 0.013	TAY = 1; CC = 1	0.58	60.6	398	4.89	4.33
SW and CC	SW = 0.003; CC < 0.001	SW = 1.9; CC = 1	0.586	61.8	398	5.29	4.54
SEA and CC	SEA = 0.010; CC = 0.003	SEA = 1; CC = 1	0.537	56.6	404	5.38	4.79

Table 6: Model performance for indicator richness. TAY = distance from Taynish, CC = canopy cover above 2m, SW = distance from standing water, and SEA = distance from the sea.

Terms	**p values**	**EDFs**	**Adj. R^2**	**Dev. Explained (%)**	**AIC**	**CV RMSE**	**CV MAE**
TAY and CC	TAY <0.001; CC = 0.011	TAY = 1; CC = 1	0.538	59.7	303	2.7	2.22
SW and CC	SW <0.012; CC = 0.006	SW = 2.3; CC = 1.1	0.5	59.4	308	2.84	2.42
SEA and CC	SEA = 0.031; CC = 0.005	SEA = 1; CC = 1.4	0.449	54.9	312	3.04	2.53

Figures 6, 7 and 8 highlight the partial effects of terms in each of these models. For all models, canopy cover has a positive linear influence on habitat quality. Distance-based metrics all have negative influences, with distance to standing water being slightly non-linear.

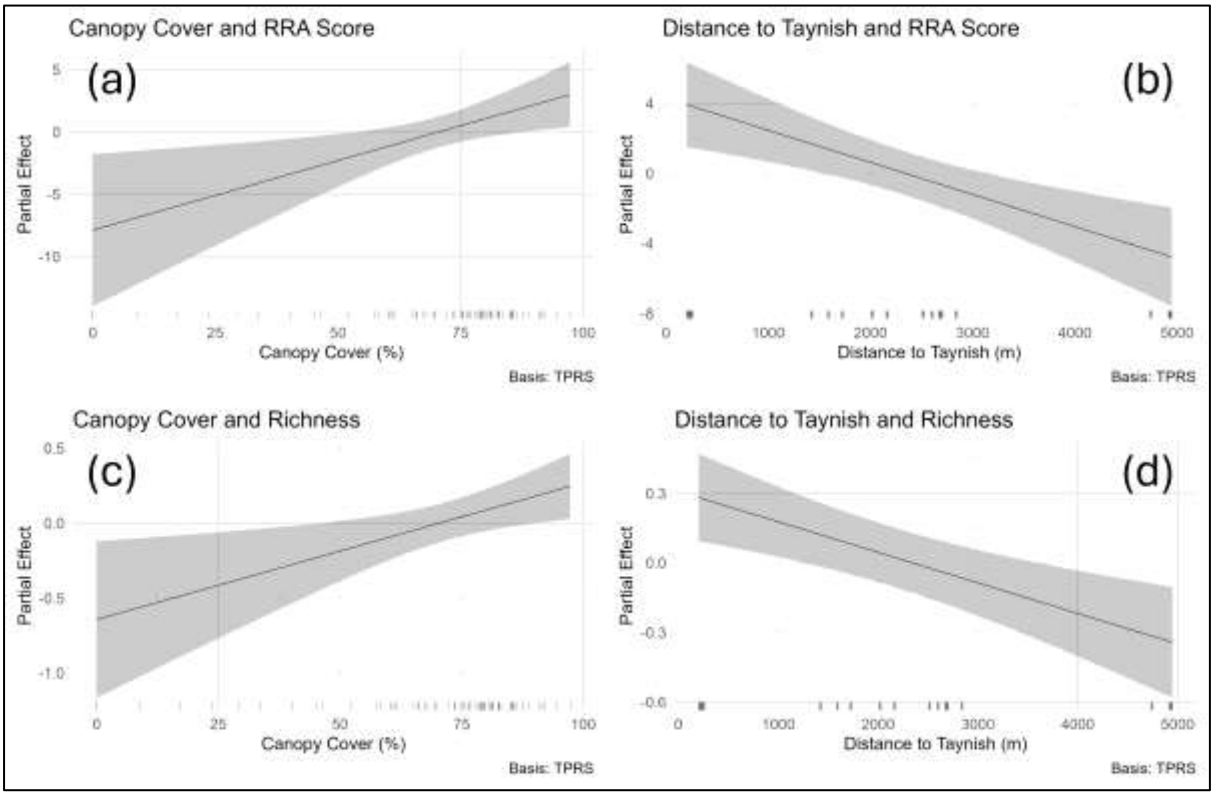


Figure 6: Partial effect graphs for model including canopy cover and distance to species pool (Taynish) for RRA score (a and b) and indicator species richness (c and d).

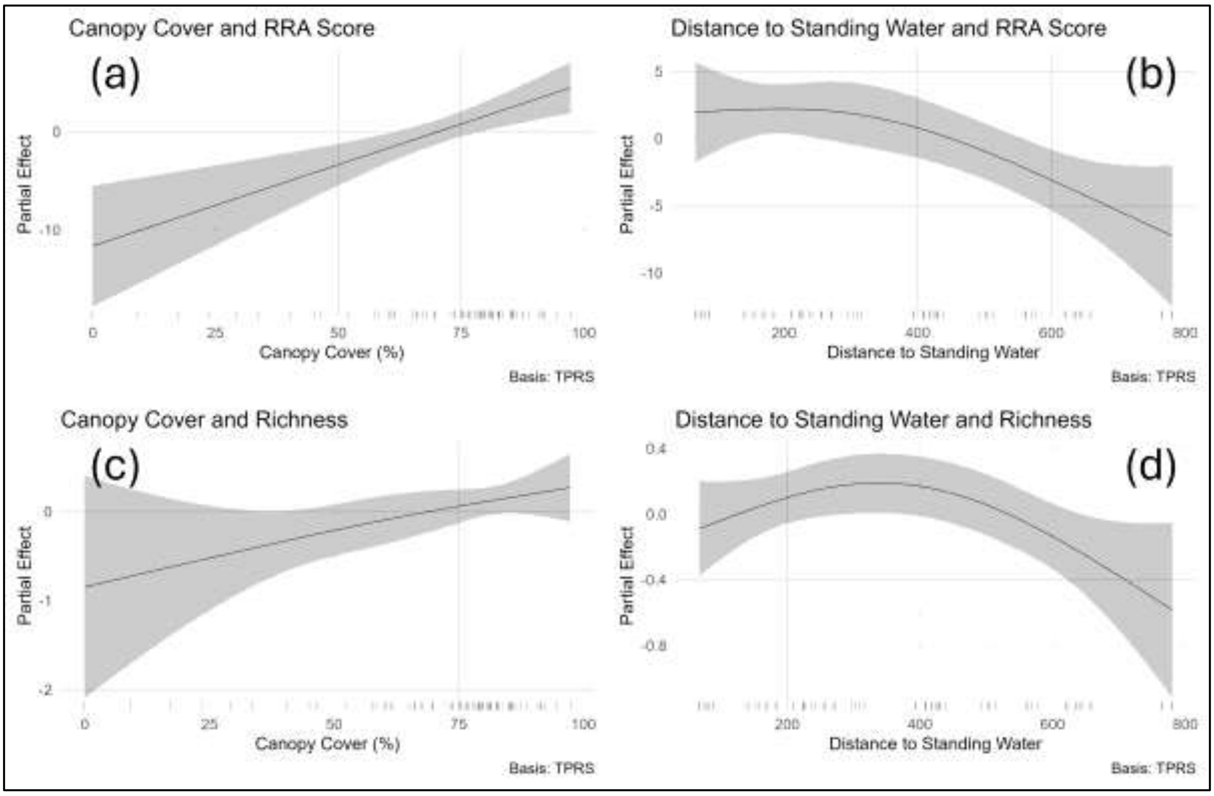


Figure 7: Partial effect graphs for model including canopy cover and distance to standing water for RRA score (a and b) and indicator species richness (c and d).

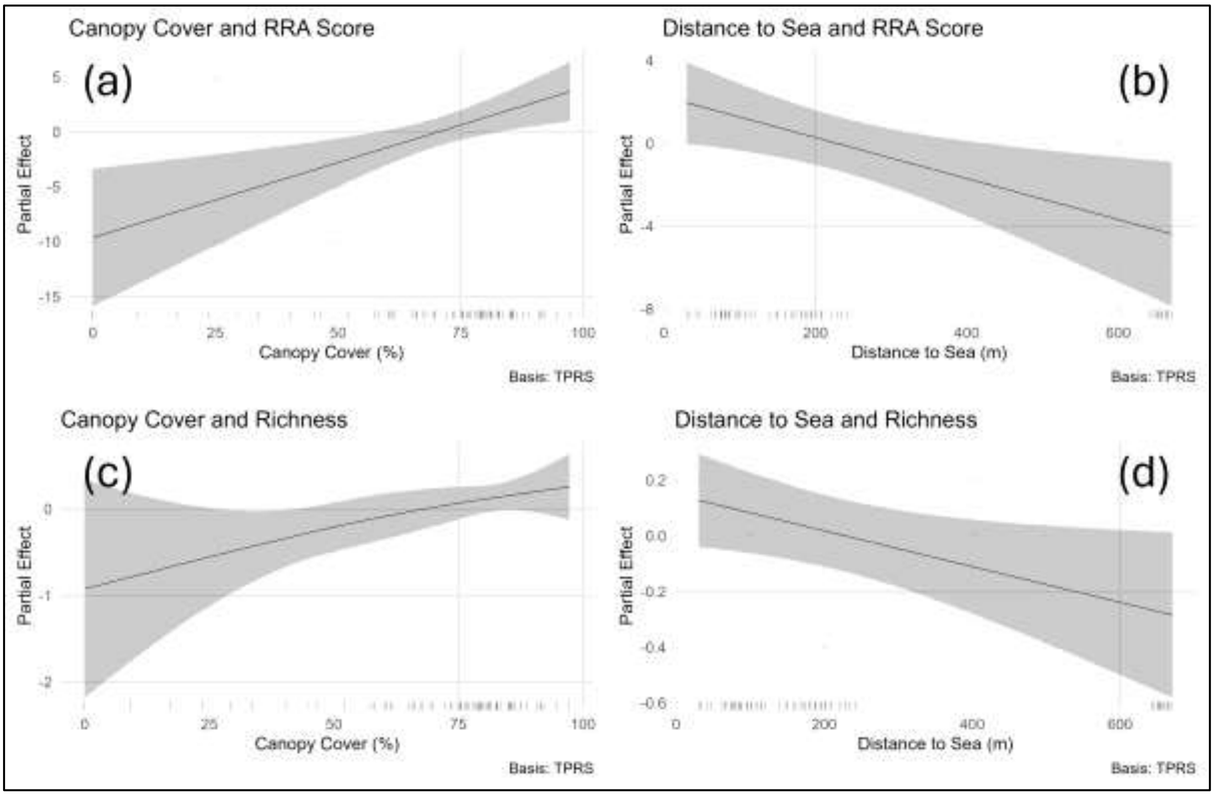


Figure 8: Partial effect graphs for model including canopy cover and distance to the sea for RRA score (a and b) and indicator species richness (c and d).

Discussion

The Role of Structural and Spatial Variables

Structural

As stated, canopy cover was the most important structural variable: it was the most significant of the structural metrics in the individual GAMs and remained significant after accounting for location and the spatial metrics. The role of canopy cover at this scale is a novel finding. As mentioned in the introduction, Ellis and Eaton (2021) found that canopy cover was not significant after accounting for climatic and geographic terms in their large-scale study. This highlights the value of conducting site-specific studies to understand local drivers of habitat quality, showing that canopy cover should be considered in forest management and plantation practices.

The linear relationship suggests that higher broad leaf canopy cover promotes forest quality. However, this can only be interpreted up to 90% as all plots over this were dominated by conifers, making it hard to determine the true driver of quality beyond this point. Additionally, the optimum canopy cover identified before accounting for tree type (75-80%) may suggest a drop in quality beyond this point, again highlighting the need for further research. Having said that, research supports the benefit of increasing canopy cover up to 90%: a higher canopy cover would promote suitable conditions in the understorey by increasing local humidity whilst protecting lichen and bryophytes from desiccation and exposure (Ellis et al, 2015; Ellis and Eaton, 2021; Frey et al., 2023). Canopy cover beyond this point could result in too much shading for most lichen and bryophyte species. Regardless, the results of this study and previous research strongly suggest that achieving a broadleaf cover of around 90% could promote habit quality by creating favourable understorey conditions for temperate rainforest species.

Spatial

As mentioned, the distance-based metrics were identified as the most important drivers for Tayvallich. These metrics reduced the significance of the spatial smooth term, suggesting they account for part of the variation attributed to location and spatial effect. If purely location and spatial effect were the main drivers of habitat quality, the spatial smooth would remain highly significant, supporting the theory that these drivers are real effects rather than spatial artefacts (Wood, 2017).

As identified, the partial effect graphs highlight that habitat quality reduces with increasing distance from each of the metrics. However, because of the multicollinearity between these variables for Tayvallich, differentiating between their effect is challenging.

Furthermore, the significance of species pool proximity must be interpreted carefully as the majority of plots were southwest of Taynish, meaning there is a possibility it is acting as a proxy for another southwest-northeast gradient. Since the prevailing wind is from the southwest, this could be exposure and the slight significance of landscape TPI could be capturing part of this potential relationship. Furthermore, although exposure was not significant in this study, it only provided categorical information based on prevailing wind direction so may not capture complex effects.

That said, there is research supporting the significance of species pool proximity in other ecosystems resulting from differences in lichen and bryophyte dispersal ranges (Benson and Coxson, 2002; Ronnas et al., 2017; Bartemucci et al., 2022; Baldwin and Bradfield, 2007; Frahm, 2008). Research has also identified *L. pulmonaria* as having a shorter dispersal range which helps to explain why it was only found in plots near Taynish (Walser, 2004; Juriado et al., 2011). This suggests that the distance from a species pool could be a real, and important, driver of temperate rainforest quality and warrants further investigation, but the results from this study are not enough to confirm this influence.

The high significance of water proximity reinforces the findings of Ellis et al. (2015) and Ellis and Eaton (2021) who identified this as a significant driver of Scottish temperate rainforest richness. Therefore, although the effect is hard to separate from distance to Taynish, the supporting literature highlights the importance of considering distance to water in plantation and management. The influence could be clarified by exploring the role of smaller waterways and rivers, present at Tayvallich, as the present study focussed on major rivers which were not present on the estate. Further research could explore this to better distinguish between the effect of species pool and water proximity.

Non-significant

As mentioned, this study identified canopy cover as the most important structural variable. However, this may be because other variables, such as ENL, were less varying across the broadleaf plots. Therefore, although we can confidently say that canopy cover had a significant role here, other structural properties may be more influential at other sites. Therefore, future research should explore other sites with more varying structural properties before the influence of metrics, such as ENL, can be disregarded as non-significant.

Among the spatial metrics, TWI, exposure, and solar radiation were not significant in this study. Given the small size of the study area, variation in solar radiation was minimal, which likely explains its lack of influence. In contrast, Ellis and Eaton (2021) found TWI to be a significant predictor at a 50 m scale, whereas it showed no effect at the finer 15 m resolution used here. This suggests that topographic wetness may have a greater influence at broader spatial scales, and its role at Tayvallich could have been underestimated due to the resolution used. Similarly, the influence of exposure may not have been fully captured by the categorical metric applied in this study. Therefore, despite their lack of significance here, both TWI and exposure may still be important for management, particularly given their relevance at larger spatial scales (Ellis and Eaton, 2021).

Implications for Management

Species and Structure

Based on the findings of the present study and previous research, recommendations can be made about *how* temperate rainforests can be created and expanded to create the optimum conditions for the unique species of lichen and bryophytes. Firstly, following Ellis and Eaton's (2021) study, it is recommended that landowners plant a diverse mixture of native broadleaf species with high bark pHs.

In terms of planting structure, the findings of the present study suggest that creating higher canopy cover (up to 90%) using these broadleaf species would promote temperate rainforest quality. This could be achieved by adopting a higher planting density than traditional practices. One example of this is the 'Miyawaki Forest Method (MFM)', which involves dense planting (2-7 trees/m^2) of a diverse mixture of native saplings with well-developed root systems (Miyawaki and Golley, 1993; Qi et al., 2024). Although there is not yet any research exploring the effectiveness of MFM technique for temperate rainforest regeneration, UK-based trial studies have revealed promising results in terms of woodland health and biodiversity (Butfoy, 2024). For these reasons, the usage of the MFM for temperate rainforest afforestation, reforestation, and expansion warrants further exploration.

Location

Using the current study and previous studies, recommendations on *where* planting should take place can be made. Firstly, expansion of existing woodlands would promote the highest biodiversity gain as species with limited dispersal range (e.g. *L. pulmonaria)* would be able to spread into the newly planted woodland.

However, where this is not feasible afforestation and reforestation should be targeted at areas shielded from exposure and as close to water as possible. In scenarios such as this, research has found that lichen and bryophyte colonisation can be promoted through translocation. For lichen, this process involves collecting samples from one site and adhering them to an appropriate substrate in the new site. This process has been successful for lichen in UK temperate rainforests (Plantlife, 2021). For bryophytes, translocation of dead wood from one site to another has proven to increase species richness and promote the colonisation of rare species (Tranberg et al., 2024; Tranberg et al., 2025). These methods should therefore be considered in reforestation and afforestation projects beyond the dispersal range of important temperate rainforest species.

Limitations

Before any results can be used to inform management, it is important to consider the limitations of the study. Some limitations have already been identified, primarily the difficulty in confirming the role of species pool proximity, the lack of complexity in the exposure and water proximity metrics. Furthermore, important variables such as woodland patch size and soil properties were not used in the analysis due to a lack of data.

In addition, the dataset used was relatively small (64 samples), with limited samples representing high quality forest. This means that there may be other variables that create high-quality forest which were not captured in the data set. Furthermore, interpretation is limited by not exploring species-specific response to structural traits, which may reveal more nuanced relationships. That said, using both RRA score and indicator richness means the drivers identified for Tayvallich are robust, ecologically meaningful, and influence overall condition. Increasing the number of data points and exploring species specific-response would also help to differentiate the effects of multicollinear variables and the role of species pool proximity.

There may also be potential issues and uncertainty with the sampling design. Firstly, although the RRA is standardised, the methodology is inherently subjective meaning there is potential for researcher bias. However, the high correlation between RRA score and indicator species richness suggests that this bias was not problematic, and that RRA score still captures overall plot condition. Uncertainty is also introduced by GNSS geolocation error (up to ~10 m). However, the robust and realistic relationships identified suggest that these effects are not artefacts of sampling or geolocation error and represent real ecological response.

Future Research

Whilst this study provides an exploratory analysis of habitat response to structural and environmental metrics, further research is required to fully understand how these metrics relate to ecological processes.

As mentioned, influence of species pool proximity warrants further explanation to confirm its significance and better understand the mechanisms behind it. In addition, future research should explore species-specific responses, rather than overall richness, to identify which species are sensitive to which metrics. This would also help better our understanding of the functional benefits of water proximity.

Future research could also explore the role of smaller waterways to understand this further, by using more detailed datasets such as the Ordnance Survey (OS) MasterMap (OS, 2025). Additionally, collecting accurate wind data for Tayvallich or using a continuous metric could highlight more nuanced influences of exposure. Finally, future research should also explore the role of soil properties and patch size, which can also inform planting strategy.

Conclusion

To summarise, this study offers a novel and informative exploration of how site-specific structural and spatial variables influence temperate rainforest quality. For Tayvallich, the most important drivers of habitat quality were canopy cover and proximity to water. Whilst proximity to a species pool was also significant, further research is needed to confirm this relationship and understand the underlying mechanisms.

Although canopy cover beyond 90% was not well represented in the dataset, the results support existing research suggesting that broadleaf canopy cover up to this threshold enhances understorey conditions for lichen and bryophytes. Similarly, the significance of water proximity aligns with previous findings and highlights its potential functional importance.

Based on these findings and supporting literature, it is recommended that land managers prioritise planting in sheltered areas close to water, using a diverse mix of broadleaf species with high bark pH and aiming to achieve up to 90% canopy cover. This study highlights the value of exploring site-specific drivers of temperate rainforest quality, and future research should explore how structural and spatial metrics influence quality in other temperate rainforests.

Research Paper References

Alaback, P. (1991). Comparative Ecology of Temperate Rainforests of the Americas Along Analogous Climate Gradients. *Revista Chilena de Historia Natural,* 64, pp. 399-412.

Baldwin, L.K. and Bradfield, G.E. (2007). Bryophyte responses to fragmentation in temperate coastal rainforests: a functional group approach. *Biological Conservation*, 136(3), pp. 408–422. DOI: https://doi.org/10.1016/j.biocon.2006.12.006.

Bartemucci, P., Lilles, E. and Gauslaa, Y. (2022). Silvicultural strategies for lichen conservation: Smaller gaps and shorter distances to edges promote recolonization. *Ecosphere*, 13(1). DOI: https://doi.org/10.1002/ecs2.3898.

Benson, S. and Coxson, D.S. (2002). Lichen colonization and gap structure in wet-temperate rainforests of northern interior British Columbia. *The Bryologist*, 105(4), pp. 673–692. DOI: https://doi.org/10.1639/0007-2745(2002)105[0673:LCAGSI]2.0.CO;2.

Broome, A., Inchboard, L.L., Perks, M., Clarke, T.-K., Park, K.J. and Thompson, R. (2021). Can epiphytic lichens of remnant Atlantic oakwood trees in a planted ancient woodland site survive early stages of woodland restoration? *Annals of Forest Science*, 78(3). DOI: https://doi.org/10.1007/s13595-021-01069-w.

Brosnan, V. and Ellis, C.J. (2020). Epiphyte Response to Woodland Habitat Condition Assessed Using Community Indicators: a Simplified Method for Scotland's Temperate Rainforest. *Edinburgh Journal of Botany*, 77(3), pp.519–541. DOI: https://doi.org/10.1017/s096042862000013X.

Butfoy, L. (2024). Natural England. *Testing the Miyawaki Method in Our Urban Greenspaces.* Available at: https://naturalengland.blog.gov.uk/2024/09/19/testing-the-miyawaki-method-in-our-urban-greenspaces/. Accessed: 06/08/2025.

Coops, N.C., Tompaski, P., Nijland, W., Rickbeil, G.J.M., Nielsen, S.E., Bater, C.W. and Stadt, J.J. (2016). A forest structure habitat index based on airborne laser scanning data. *Ecological Indicators*, 67, pp.346–357. DOI: https://doi.org/10.1016/j.ecolind.2016.02.057.

DellaSala, D.A. (2011). *Temperate and Boreal Rainforests of the World: Ecology and Conservation*. Island Press, Washington, D.C.

d-maps (2025). *Scotland, United Kingdom*. [online] Available at: https://d-maps.com/carte.php?num_car=15866&lang=en. Accessed: 06/08/2025.

Ellis, C.J. (2016). Oceanic and temperate rainforest climates and their epiphyte indicators in Britain. *Ecological Indicators*, 70(1), pp.125–133. DOI: https://doi.org/10.1016/j.ecolind.2016.06.002.

Ellis, C.J. and Eaton, S. (2016). Future Non-Analogue Climates for Scotland's Temperate Rainforest. *Scottish Geographical Journal*, 132, pp. 257–268. DOI: https://doi.org/10.1080/14702541.2016.1197964.

Ellis, C.J. and Eaton, S. (2021). Microclimates Hold the Key to Spatial Forest Planning Under Climate Change: Cyanolichens in Temperate Rainforest. *Global Change Biology*, 27(9). DOI: https://doi.org/10.1111/gcb.15514.

Ellis, C.J., Eaton, S., Theodoropoulos, M. and Elliot, K. (2015). *Epiphyte Communities and Indicator Species: An Ecological Guide for Scotland's Woodlands.* Royal Botanic Garden, Edinburgh.

Frahm, J.P. (2008). Diversity, dispersal and biogeography of bryophytes (mosses). *Biodiversity and Conservation*, 17(2), pp. 277–284. DOI: https://doi.org/10.1007/s10531-007-9251-x.

Frescino, T.S., Edwards, T.C. and Moisen, G.G. (2001). Modeling spatially explicit forest structural attributes using generalized additive models. *Journal of Vegetation Science*, 12(1), pp.15–26. DOI: https://doi.org/10.1111/j.1654-1103.2001.tb02613.x.

Frey, J., Holter, P., Kinzinger, L., Schindler, Z., Morhart, C., Kolbe, S. and Seifert, T. (2023). Detailed mapping of below canopy surface temperatures in forests reveals new perspectives on microclimatic processes. *Agricultural and Forest Meteorology*, 341. DOI: https://doi.org/10.1016/j.agrformet.2023.10965.

Gazetteer for Scotland. (2021). *Climate Information for Tayvallich.* Available at: https://www.scottish-places.info/towns/townclimate3579.html. Accessed: 06/08/2025

Highlands Rewilding. (2024). *Fourth Natural Capital Report*: *Monitoring Nature Recovery.* Available at: https://static1.squarespace.com/static/621f9623d02fad4ef3e6b253/t/677bf84bc6de9437f6c7f537/1736177761265/Fourth+natural+capital+report+20.12.24-compressed_medium.pdf. Accessed: 06/08/2025.

Jüriado, I., Liira, J., Csencsics, D., Widmer, I., Adolf, C., Kohv, K. and Scheidegger, C. (2011). Dispersal ecology of the endangered woodland lichen *Lobaria pulmonaria* in managed hemiboreal forest landscape. *Biodiversity and Conservation*, 20(9), pp. 1803–1819. DOI: https://doi.org/10.1007/s10531-011-0062-8.

Kopecký, M., Macek, M. & Wild, J. (2021). Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of the Total Environment*, 757, 143785. DOI: https://doi.org/10.1016/j.scitotenv.2020.143785.

Kosicki, J.Z. (2020). Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. *Environmental and Ecological Statistics*, 27(2), pp. 273–292. DOI: https://doi.org/10.1007/s10651-020-00445-5.

Miyawaki, A. and Golley, F.B. (1993). Forest reconstruction as ecological engineering. *Ecological Engineering*, 2(4), pp. 333–345. DOI: https://doi.org/10.1016/0925-8574(93)90002-W.

Naesset, E. and Oklund, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79(1), pp.105–115. DOI: https://doi.org/10.1016/s0034-4257(01)00243-7.

NatureScot (2011). Taynish Woods Site of Special Scientific Interest. *Site Code 1522 Citation*. Available at: https://www.nature.scot/sites/default/files/site-special-scientific-interest/1522/sssi-citation.pdf. Accessed: 06/08/2025.

NatureScot. (2022). *Scotland's Biodiversity Strategy 2022-2045*. Available at: https://www.nature.scot/scotlands-biodiversity/scottish-biodiversity-strategy/scotlands-biodiversity-strategy-2022-2045.

Ordnance Survey. (2025). OS MasterMap Networks. *Water Layer*. Available at: https://www.ordnancesurvey.co.uk/products/os-mastermap-networks-water-layer. Accessed: 06/08/2025.

Plantlife (2021). Translocating Lobarian Lichens to Mitigate Losses from Ash Dieback. Available at: https://rise.articulate.com/share/Xs2UNbluYN_Xwa_xrWC4vGJAIOYB78To#/lessons/6bOx55XEN8Z_zyVIcUFm8itf590zVnrR. Accessed on: 06/08/2025.

Plantlife (2023a) *Rapid Rainforest Assessment Guidance*. Available at: https://www.plantlife.org.uk/wp-content/uploads/2023/03/Rapid-Rainforest-Assessment-GUIDANCE-1.pdf. Accessed: 06/08/2025.

Plantlife (2023b) *Rapid Rainforest Assessment Survey Form*. Available at: https://www.plantlife.org.uk/wp-content/uploads/2023/03/Rapid-Rainforest-Assessment-SURVEY-FORM-1.pdf. Accessed: 06/08/2025.

Plantlife (2024a) Lichens of Scotland's Rainforest. *Guide 1 Lichens on ash, hazel, willow, rowan and old oak.* Available at: https://static1.squarespace.com/static/5f8efc26ea62c54078a69a6f/t/67ae4a242aa7026d5d9d108a/1739475512602/Lobarion+Lichen+Guide+1.pdf

Plantlife (2024b) Lichens of Scotland's Rainforest. *Guide 2 Lichens on birch, alder and oak.* Available at: https://static1.squarespace.com/static/5f8efc26ea62c54078a69a6f/t/67ae4a5c4d54cb38161b8d84/1739475566819/Parmelion+Lichen+Guide+2.pdf

Qi, H., Dempsey, N. and Cameron, R. (2024). Seeing the forest for the trees? An exploration of the Miyawaki forest method in the UK. *Arboricultural Journal*, pp.1–13. DOIi: https://doi.org/10.1080/03071375.2024.2394355.

Ronnås, C., Werth, S., Ovaskainen, O., Várkonyi, G., Scheidegger, C. and Snäll, T. (2017). Discovery of long-distance gamete dispersal in a lichen-forming ascomycete. *New Phytologist*, 216(1), pp. 216–226. DOI: https://doi.org/10.1111/nph.14714.

Tranberg, O., Hekkala, A.-M., Lindroos, O., Löfroth, T., Jönsson, M., Sjögren, J. and Hjältén, J. (2024). Translocation of deadwood in ecological compensation: A novel way to compensate for habitat loss. *Ambio*, 53(3), pp. 482–496. DOI: https://doi.org/10.1007/s13280-023-01934-0.

Tranberg, O., Hekkala, A.-M., Lindroos, O., Löfroth, T., Jönsson, M., Sjögren, J. and Hjältén, J. (2025). Enhanced bryophyte communities, but challenges for lichens following translocation of deadwood in ecological compensation. *Science of the Total Environment*, 858, 159597. DOI: https://doi.org/10.1016/j.scitotenv.2022.159597.

UKCEH. (2023). UK Centre for Ecology and Hydrology. *Information Products*. Available at: https://www.ceh.ac.uk/data/information-products.

Walser, J.C. (2004). Molecular evidence for limited dispersal of vegetative propagules in the epiphytic lichen *Lobaria pulmonaria*. *American Journal of Botany*, 91(8), pp. 1273–1276. DOI: https://doi.org/10.3732/ajb.91.8.1273.

Wood, S.N. (2017). *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC Press, New York. DOI: https://doi.org/10.1201/9781315370279.

Woodland Trust. (2021). *Temperate Rainforests in the UK.* Available at: https://www.woodlandtrust.org.uk/trees-woods-and-wildlife/habitats/temperate-rainforest/.

Yee, T.W. and Mitchell, N.D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2(5), pp. 587–602. DOI: https://doi.org/10.2307/3236170.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M. and Viitala, R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS journal of photogrammetry and remote sensing*, 66(1), pp.28–37. DOI: https://doi.org/10.1016/j.isprsjprs.2010.08.003.

Part II: Technical Report

Contents

Part II: Technical Report 1

Table of Tables 2

Table of Figures 2

Introduction 3

Preprocessing 3

Individual Tree Detection 3

Structural Metrics 4

Height-Based Metrics 5

Horizontal Cover 5

Vertical Complexity 6

Spatial Metrics 7

Topographic Variables 7

Climatic Variables 8

Distance-Based Metrics 9

Sampling, Fieldwork and Geolocation 11

Random Stratified Sampling 11

Geolocation and Digitisation 12

Correlation Matrix 13

GAMs 13

Notable Results 14

Technical Report References 19

Appendices to the Technical Report 22

A – Scripts 22

B – Scatter Plots 59

C – Data Availability and Structure 85

Table of Tables

Table 1: Performance of spatial metrics with spatial smooth for RRA score. 15
Table 2: Performance of spatial metrics with spatial smooth for richness..............................16
Table 3: Performance of spatial metrics without spatial smooth for RRA score. 17
Table 4: Performance of spatial metrics without spatial smooth for richness. 17
Table 5: dissfinal folder structure...85

Table of Figures

Figure 1: Polygon used to determine Taynish boundary (NatureScot, 2007). 10
Figure 2: Study site locations across Tayvallich Estate. 12
Figure 3: Pearson's correlation matrix for the structural and spatial metrics..........................13
Figure 4: Graphs highlighting the influence of canopy cover on RRA score (a) and richness (b), before accounting for tree type...15

Introduction

This report accompanies Part I and provides a more in-depth description of the methods used. Important scripts, scatter plots and the location of data can be found in the appendix. Whilst comments have been added to the scripts, they are limited and as such the scripts should be interpreted alongside the main body of this report.

Preprocessing

As stated in the research paper, preprocessing was required before further analysis could take place. The ALS data was received with the ground classified, so this step was not required in this instance. Data preprocessing was conducted in R, using the `lidR` package (Roussel et al., 2020; Roussel and Auty, 2025). The 5 point clouds were read in together using the `readLAScatalog` function, which allows all 5 point clouds to be processed together, providing one raster as the output using the `opt_merge` function.

The `rasterize_terrain` and `rasterize_canopy` functions were used to create the DTM and DSM. For the DTM, the invert distance weighting (IDW) algorithm was used as it is more robust to edge effects than other methods (Roussel et al., 2024). The DTM was created at 1 m resolution for normalising the point clouds and then again at 15 m for the spatial metrics.

For the DSM, the pit-free algorithm was used. This methodology was developed by Khosravipour et al. (2014) and uses sequential height thresholds, applying Delauny triangulations to first returns. For each threshold, the triangulation is cleaned of triangles that are too large. This creates partial rasters which are then stacked, retaining the highest value per pixel. The output is a DSM which is natively free from spikes and depressions (pits) caused by noise rather than real surface features. The DSM was created at 15 m resolution, using the `lidR` default thresholds (0, 2, 5, 10, 15) which were deemed appropriate based on site observations and known tree heights.

Individual Tree Detection

Ideally, to best understand how woodland structure influences quality metrics would be calculated at the tree level; exploring how stand distribution relates to quality would directly inform how to plant trees. Consequently, individual tree detection was attempted using the local maxima filter (LMF) algorithm on a pit-free canopy height model (CHM), a normalised DSM where height values are relative to the ground, rather than absolute elevation. This method was chosen as it has been proven to be the most accurate method for mixed and broadleaf forests with ALS data of similar point density (Yang et al., 2019).

This was also done in `lidR`, using the `locate_trees` function which allows the user to input a CHM and select an algorithm. As stated, the algorithm selected was LMF, which uses a moving window on a CHM and identifies local high points as treetops. Window size should roughly represent the size of a tree crown. The ITD process was optimised using simulated annealing from the `GenSA` R package (Xiang et al., 2013). Arguments in the pit-free CHM creation process (such as resolution or thresholds), the extent of CHM smoothing and the LMF window size were all optimised. The LMF window size was variable, meaning the size could increase or decrease depending on CHM values. This allows for more realistic window sizes, assuming taller trees having larger crowns. The accuracy of the tree detection was assessed following the method of Eysn et al. (2015). This was used as the cost function and to report the overall accuracy of the model.

Using this pit-free LMF method, it was not possible to get a usable accuracy to achieve the aim of this study. Further research could explore some of the other algorithms highlighted by Yang et al. (2019), however this was not feasible within the time constraints of the present study and was beyond the present study's aim of analysing structural and spatial metrics.

Structural Metrics

Since tree-level metrics could not be calculated, plot-level structural metrics were created. Whilst not as directly applicable to planting design, plot-level metrics still provide valuable information that can inform planting strategies. As mentioned in the Research Paper, structural metrics can be grouped into 3 categories: (1) height-based metrics, (2) metrics concerning horizontal vegetation cover, and (3) metrics reflecting the vertical complexity of vegetation.

The `lidR` packaged was used to calculate metrics for each of the categories. This was done using the `pixel_metrics` function. The `pixel_metrics` function is applied to a point cloud, calculating a custom or predefined function at a desired resolution. `pixel_metrics` was applied to the normalised point cloud catalogue. Using `opt_merge`, the function output one raster layer for all point clouds with each pixel contains the calculated metric value. `lidR` allows for easy calculation of commonly used metrics through its `stdmetrics` function. All of the structural metrics used in this study were calculated using the `pixel_metrics` function at 15 m (unless specified otherwise), using both predefined functions and custom-build functions. The calculations used to create each metric are described below.

Height-Based Metrics

Height-based metrics were extracted from the `stdmetrics` output, these were max height per pixel and mean height per pixel. Whilst canopy height can be somewhat controlled through forest management (e.g. through coppicing and pollarding), height metrics were calculated so their role can be accounted for when interpreting the influence of other, more controllable metrics.

Although max height does not account for noise, summary statistics in the Research Paper highlight that max values were not unreasonable.

Horizontal Cover

Canopy Cover

Canopy cover was calculated as the percentage of points over a height threshold, as per Coops et al. (2016). This was calculated at using 2 m and 6 m thresholds. The 2 m threshold was selected to include shrubs and smaller trees, representing total canopy coverage. This threshold has been frequently used to calculate forest structural metrics in ecological research (Naesset, 2002; Yu et al., 2011; Kankare et al., 2015; Palmroos et al., 2023). This threshold was also validated through observations in the field. The 6 m threshold was selected to represent the canopy coverage of emergent trees. This threshold was decided by using local allometry as trees were defined as woody plants with a diameter at breast height (dbh) > 7 cm, as per Hamilton (1988). Using UK-based allometry for the common tree species at the site, 6 m was identified as the average height where dbh = 7 cm (Evans et al., 2015).

Lower-Understorey Dominance

The proportion of points below 4 m was also calculated to understand how dominant the lower and understorey are.

Canopy Cover Variance

Canopy cover was then calculated again at 1m, using the 2 m threshold. This was then upscaled to 15 m resolution to facilitate the calculation of standard distribution of canopy coverage, to reflect variance in canopy cover per pixel.

Vertical Complexity

Effective Number of Layers

Effective number of layers (ENL) was created by Ehbrecht et al. (2016) to overcome limitations in other vertical complexity metrics, which can be ambiguous and hard to interpret (Ehbrecht et al., 2016; Aalto et al., 2023). For example, foliage height diversity (FHD) applies Shannon entropy to foliage height profiles and the proportions of material in each profile, deriving a single metric describing the vertical structure. However, this means that stands with equal filling in the vertical strata (homogenous structure) produce higher FHD values than stands with distinct canopy layers (diverse structure) (Aalto et al., 2023).

ENL overcomes this problem by using Hill numbers (0D, 1D, and 2D), a family of diversity indices initially designed to quantify the diversity and dominance of a community (Hill, 1973). In ENL, these numbers are used to apply a different weighting to woodland layers: ENL0D applies no weighting and translates to total number of layers, ignoring dominance; ENL1D is the exponential of Shannon entropy, which applies equal weighting to all dominant layers; ENL2D applies the inverse Simpson index to allocate more weighting to the most dominant layers and less to rarer ones.

Before ENL was calculated, points below 0.3 were removed from the point cloud to remove any ground-based noise and shorter vegetation. ENL was calculated using the following equations:

$$\mathrm{ENL1D} = exp\left(-\sum_{i=1}^{n} p_i \ln(p_i)\right)$$

$$\mathrm{ENL2D} = 1\Big/\sum_{i=1}^{n} p_i^2$$

where p_i is the proportion of filled voxels in the i_{th}vertical layer to the sum of filled layers in the vertical profile analysed. n refers to the canopy top, and ln is the natural logarithm.

To ensure ENL values are comparable between studies and results are easily interpretable, Ehbrecht et al. (2016) states that layers should be 1 m thick. As such ENL0D was not calculated as this would equate to canopy height.

Height Variance

Since ENL was designed using terrestrial laser scanning (TLS), the standard deviation of return heights was also calculated to quantify vertical structural variability if the outputs of the ENL seemed unrealistic (although this was not the case). Height variance was also extracted from the `stdmetrics` output.

Spatial Metrics

As described in the research paper, spatial metrics were created in order to understand how topographic, climatic, and distance-based variables influence temperate rainforest quality. Metrics relevant to lichen and bryophyte habit from each of these categories were calculated, and the procedure for each is described below. As with the structural metrics, all metrics were created at 15 m resolution.

Topographic Variables

Topographic variables describe those created using the 15 m resolution DTM. Firstly, the DTM was used to calculate aspect and slope using the `terra` R package (Hijmans, 2025).

Topographic Wetness Index

A TWI was created to explore how rainwater would accumulate according to the DTM. As lichen are poikilohydric, understanding how water availability differs across a site is an important factor to consider. In fact, previous research has identified TWIs influence lichen assemblage in Scottish temperate rainforests (Ellis and Eaton, 2021). The TWI was calculated using the `whitebox` R package (Wu and Brown, 2022; Lindsay, 2016).

This was done using the FD8 algorithm (the `wbt_fd8_flow_accumulation` function) which has been found to be the most accurate algorithm for studying ecological relationships in temperate forest systems (Kopecky et al., 2021). The FD8 algorithm provides the specific catchment area for a pixel, meaning the potential water supply to that cell based on the surrounding topography, which is used in the following equation to derive TWI:

$\text{TWI} = \ln(\mathit{specific\ catchment\ area} / \tan(\mathit{slope}))$

Topographic Position Index

TPI is a more general metric which describes the elevation of a location relative to its surrounding terrain. Since there is significant topographic variation at Tayvallich, it was important to explore how this may contribute to habitat quality.

TPI was calculated using the `MultiscaleDTM` R package and its `TPI` function (Ilich et al., 2021; Ilich et al., 2023). This function allows the user to define the shape and size of a moving window, which evaluates each cell's position relative to the other values in the window. This was done using a circular window at different spatial scales, using 500 m, 90 m, and 30 m radii to capture landscape, intermediate, and fine-scale topographic variation. These values were selected based on measurements conducted in ArcGIS Pro (Esri, 2024), using the Measure tool on the DTM to understand at which scales topography varied.

Climatic Variables

Climatic variables include wind exposure and the amount of solar radiation received, which would influence plot microclimate. It is therefore important to consider this effect in management and planting design.

Wind Exposure

Wind exposure was calculated using EXPOS model from the `ExposR` R package (Boose, 2024). This model has previously been used to explore relationships between exposure and ALS-derived forest structure (Ankori-Karlinsky et al., 2024). The DSM was used to calculate exposure, accounting from wind shielding from vegetation. Slope and aspect were calculated and used within the function, and a wind direction of 225 degrees (southwest) was set as the prevailing wind direction (Met Office, 2016).

The EXPOS model provides a categorical raster as the output with the following values (0 = missing data, 1 = protected, 2 = exposed). Whilst this is a simplistic model, creating an accurate model without accurate wind data for the estate would be challenging.

Solar Radiation

Solar radiation was calculated using both the DTM and the DSM. This was computed using the ArcGIS Pro Raster Solar Radiation spatial analyst tool, capable of calculating solar radiation at different time scales. The DTM was used to calculate total solar radiation for 2023, to align with the ALS data. To account for vegetation, solar radiation was also calculated using the DSM. Since the ALS data is leaf-on, the DSM was used to calculate solar radiation within the growing season (May – September) of 2023.

Distance-Based Metrics

Distance from Species Pool (Taynish)

As highlighted in the research paper, Tayvallich is adjacent to Taynish NNR, a species-rich temperate rainforest. Taynish contains a diverse range of lichen and bryophyte species, including 475 species of lichen (148 of which are considered notable species) and more than 250 species of bryophytes (around 25% of all British species) (NatureSoct, 2011).

Since lichen and bryophytes are anemochorous, meaning their spores are dispersed by the wind, it is important to consider how the distance from a species pool such as Taynish influences quality.

To calculate this, data was downloaded from the Native Woodland Survey of Scotland (NWSS) (Scottish Forestry, 2024). As data from the NWSS contains many polygons representing sub-communities, polygons existing within the boundary highlighted in Figure 1 were aggregated into one polygon in ArcGIS Pro using the Dissolve tool. Polygons. The ArcGIS Pro Euclidian Distance tool was then used to measure the distance from the centre of each cell to the Taynish polygon.

Figure 1: Polygon used to determine Taynish boundary (NatureScot, 2007).

Distance from water

As stated, proximity to water has been found to be a significant contributing factor to lichen assemblage (Ellis and Eaton, 2021)

Data for standing water and foreshore water were sourced from Ordnance Survey (OS) OpenMap Local (OS, 2025a). In terms of running water, the OS Open Rivers (OS, 2025b) was used but no major rivers were present on the estate. However, based on satellite imagery more detailed datasets should have been used to capture the smaller rivers and water ways observed on the estate. The standing water and foreshore water polygons were merged into one dataset, also using the Dissolve ArcGIS Pro tool. Distance standing water, the foreshore, and the combined dataset was then measured using the Euclidian Distance tool. This revealed that all plots were closer to the sea than standing water.

Sampling, Fieldwork and Geolocation

As described in the Research Paper, the RRA and species richness were used as quality indices. Before collecting the data, the RRA was practiced in other woodlands to ensure familiarity with the methodology. Temperate rainforest examples were also visited with lichenologist Dr Christopher Ellis from the Royal Botanic Gardens Edinburgh, who ensured familiarity with the indicator species. Tayvallich Estate was visited prior to data collection to inform the development of an effective sampling strategy and assess the accessibility of sites.

Random Stratified Sampling

As mentioned, random stratified sampling was used to ensure study sites had varying conditions. Sites were originally going to be 30x30 m as this has been identified as the scale at which remotely sensed woodland structure and spatial variables influence lichen communities (Palmroos et al., 2023). However, quality varied at a finer scale than this at Tayvallich, likely because temperate rainforests are more dynamic than the Finnish boreal forests examined by Palmroos et al. (2023). Consequently, the 30 m sites were split into 15 m quadrats (plots) which better reflected the variance in quality. The distribution of the 16 sites can be seen in Figure 2.

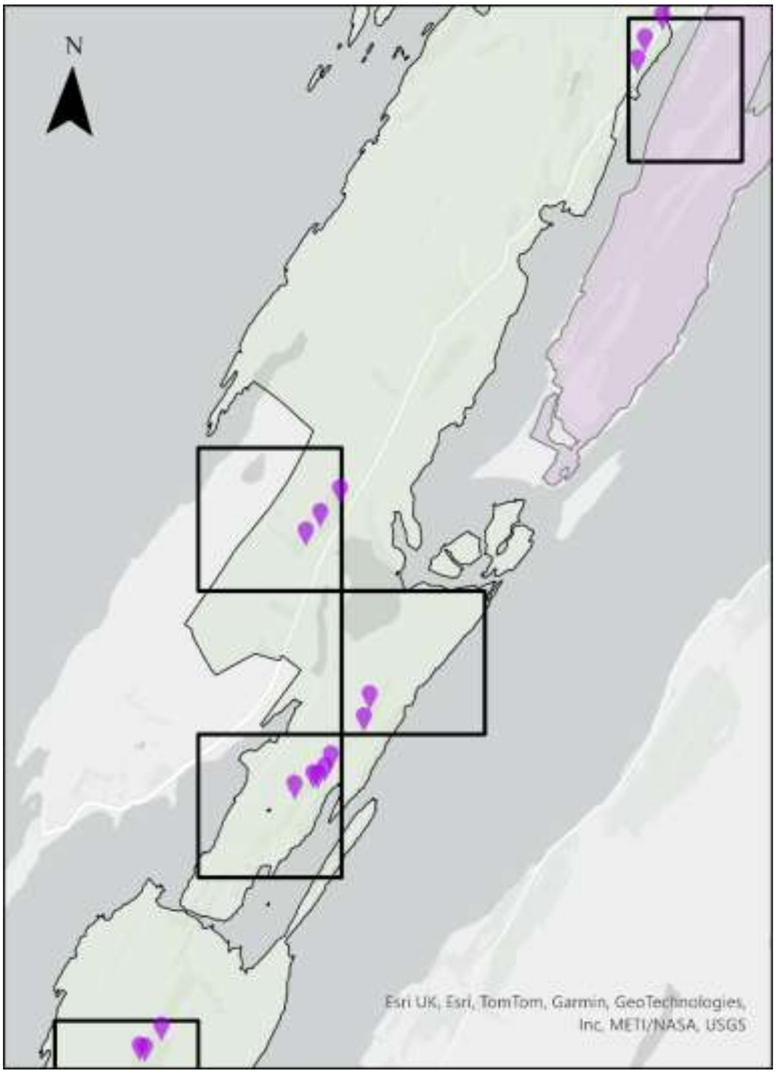


Figure 2: Study site locations across Tayvallich Estate.

Geolocation and Digitisation

As described, the location of each site was recorded using a handheld GNSS. This was set to average its location at the centre of the site for the duration of the survey (around 30 minutes for each site). The GPS coordinates were imported from the GPX files, combined, and filtered to retain only valid sampling locations.

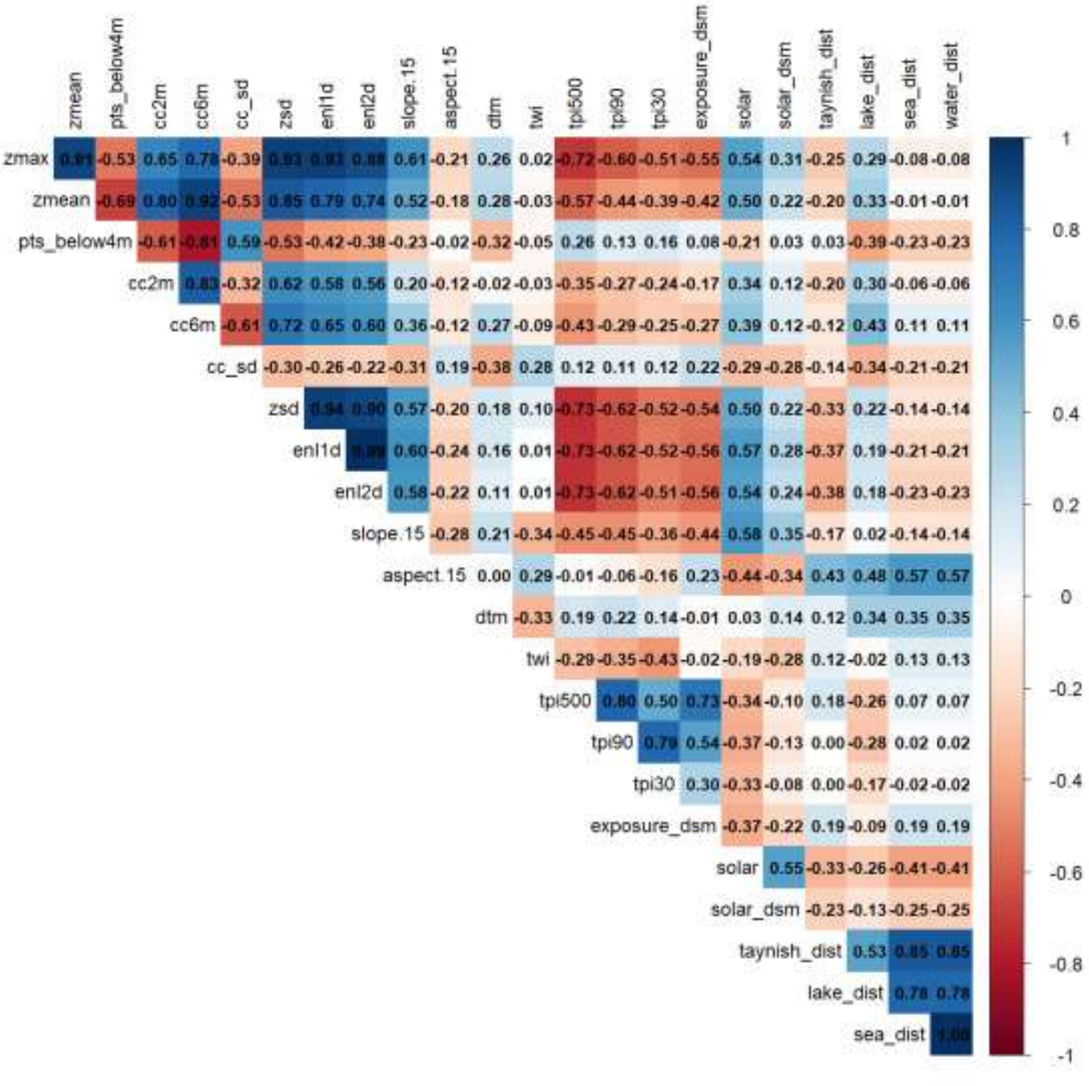
The centroid of each 15x15 m plot was then calculated based on spatial offsets from the site point, producing four plot centroids per site. The RRA score and indicator richness was then appended as attributes in a GeoPackage.

The structural and spatial raster layers were stacked and values from each raster were extracted for every plot, ensuring that only data relevant to sampled plots were retained. The result was a data frame containing coordinates and their corresponding quality scores and metric values.

These steps were also completed in R, using the `terra` and `SF` packages (Pebesma, 2018; Pebesma and Bivand, 2023).

Correlation Matrix

As mentioned, a Pearson’s correlation matrix was created to mitigate multicollinearity in the GAM. The result of this can be seen in Figure 3.



	zmean	pts_below4m	cc2m	cc6m	cc_sd	zsd	enl1d	enl2d	slope.15	aspect.15	dtm	twi	tpi500	tpi90	tpi30	exposure_dsm	solar	solar_dsm	taynish_dist	lake_dist	sea_dist	water_dist
zmax	0.91	-0.53	0.65	0.78	-0.39	0.93	0.93	0.88	0.61	-0.21	0.26	0.02	-0.72	-0.60	-0.51	-0.55	0.54	0.31	-0.25	0.29	-0.08	-0.08
zmean		-0.69	0.80	0.92	-0.53	0.85	0.79	0.74	0.52	-0.18	0.28	-0.03	-0.57	-0.44	-0.39	-0.42	0.50	0.22	-0.20	0.33	-0.01	-0.01
pts_below4m			-0.61	-0.81	0.59	-0.53	-0.42	-0.38	-0.23	-0.02	-0.32	-0.05	0.26	0.13	0.16	0.08	-0.21	0.03	0.03	-0.39	-0.23	-0.23
cc2m				0.83	-0.32	0.62	0.58	0.56	0.20	-0.12	-0.02	-0.03	-0.35	-0.27	-0.24	-0.17	0.34	0.12	-0.20	0.30	-0.06	-0.06
cc6m					-0.61	0.72	0.65	0.60	0.36	-0.12	0.27	-0.09	-0.43	-0.29	-0.25	-0.27	0.39	0.12	-0.12	0.43	0.11	0.11
cc_sd						-0.30	-0.26	-0.22	-0.31	0.19	-0.38	0.28	0.12	0.11	0.12	0.22	-0.29	-0.28	-0.14	-0.34	-0.21	-0.21
zsd							0.94	0.90	0.57	-0.20	0.18	0.10	-0.73	-0.62	-0.52	-0.54	0.50	0.22	-0.33	0.22	-0.14	-0.14
enl1d								[illegible]	0.60	-0.24	0.16	0.01	-0.73	-0.62	-0.52	-0.56	0.57	0.28	-0.37	0.19	-0.21	-0.21
enl2d									0.58	-0.22	0.11	0.01	-0.73	-0.62	-0.51	-0.56	0.54	0.24	-0.38	0.18	-0.23	-0.23
slope.15										-0.28	0.21	-0.34	-0.45	-0.45	-0.36	-0.44	0.58	0.35	-0.17	0.02	-0.14	-0.14
aspect.15											0.00	0.29	-0.01	-0.06	-0.16	0.23	-0.44	-0.34	0.43	0.48	0.57	0.57
dtm												-0.33	0.19	0.22	0.14	-0.01	0.03	0.14	0.12	0.34	0.35	0.35
twi													-0.29	-0.35	-0.43	-0.02	-0.19	-0.28	0.12	-0.02	0.13	0.13
tpi500														0.80	0.50	0.73	-0.34	-0.10	0.18	-0.26	0.07	0.07
tpi90															0.79	0.54	-0.37	-0.13	0.00	-0.28	0.02	0.02
tpi30																0.30	-0.33	-0.08	0.00	-0.17	-0.02	-0.02
exposure_dsm																	-0.37	-0.22	0.19	-0.09	0.19	0.19
solar																		0.55	-0.33	-0.26	-0.41	-0.41
solar_dsm																			-0.23	-0.13	-0.25	-0.25
taynish_dist																				0.53	0.85	0.85
lake_dist																					0.78	0.78
sea_dist																						1.00

Figure 3: Pearson's correlation matrix for the structural and spatial metrics.

GAMs

The `mgcv` R package was used to create GAMs via the `gam` function (Wood, 2025). This function was applied to model both RRA score and richness, using the same terms in each iteration. A Gaussian family was selected for RRA score as it is a continuous variable, while a Poisson family was used for richness since it represents count data. The family specifies the assumed distribution of the response variable and guides how the model relates predictors to the outcome. The Gaussian family assumes a continuous, normally distributed response, making it appropriate for RRA scores, whereas the Poisson family is designed for discrete count data like richness.

Interaction terms are added using the `ti` smooth function, and individual terms (main effects) are added using the `s` smooth function. The `s` smooth function can be used to model interactions if the variables are in the same unit, by including both variables in the brackets, separated by a comma. For example, spatial smooths are added using `s(x, y)`, where x and y are the coordinates.

The `gratia` R package was used to create the partial effect graphs of each term via the `draw` function (Simpson, 2024).

As described in the research paper, a LOSO cross-validation method was adopted to assess the robustness of the model and identify overfitting. This was achieved in R, using the `RMSE` and `MAE` functions from the `Metrics` package (Hamner et al., 2018).

Notable Results

All GAM outputs can be found in the ‘gams’ folder, see Appendix C for more details.

Structural

As mentioned in the report, canopy cover before accounting for tree type exhibited a different influence on the quality indices and an optimum was identified, as highlighted in partial effect graphs highlighted below.

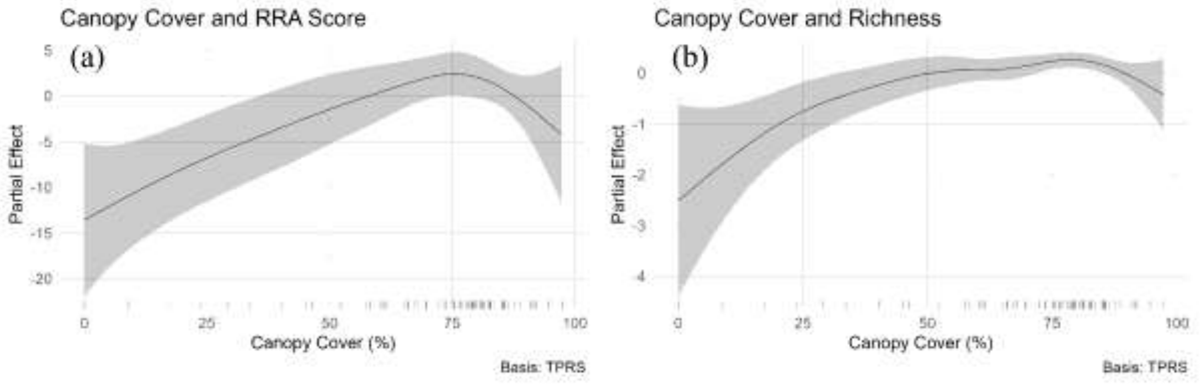


Figure 4: Graphs highlighting the influence of canopy cover on RRA score (a) and richness (b), before accounting for tree type.

These graphs highlight the optimum for both quality indices. It is worth mentioning that other metrics, such as ENL, were also significant for both indices before accounting for tree type. However, the insignificance after accounting for tree type suggests that these are not genuine ecological influences but artefacts of tree type.

Additive models including all structural metrics and tree type revealed multicollinearity in their explanation. However, ENL became significant in a model including the spatial smooth, tree type, and canopy cover showing that, after accounting for the effects of these variables, ENL was significant. However, ENL was no longer significant in GAMs including spatial metrics (distance to sea and distance to Taynish) and as such it was not considered to offer high explanatory power.

Spatial Metrics

Please see the GAM results of spatial metrics with the spatial smooth for both RRA score (Table 1) and richness (Table 2).

Table 1: Performance of spatial metrics with spatial smooth for RRA score.

Terms	**p values**	**EDFs**	**Adj R2**	**Dev. Explained**	**AIC**
s(x, y) + s(dtm)	s(x,y) <0.001; s(dtm) 0.459	s(x,y) 5.9; s(dtm) 1.0	0.428	49.1	423
s(x, y) + s(slope.15)	s(x,y) <0.001; s(slope.15) 0.643	s(x,y) 3.3; s(slope.15) 1.2	0.378	42.5	401
s(x, y) + s(aspect.15)	s(x,y) <0.001; s(aspect.15) 0.729	s(x,y) 3.1; s(aspect.15) 1.4	0.375	42.3	402
s(x, y) + s(solar)	s(x,y) <0.001; s(solar) 0.325	s(x,y) 3.4; s(solar) 1.0	0.384	43	400

s(x, y) + s(solar_dsm)	s(x,y) <0.001; s(solar_dsm) 0.493	s(x,y) 3.2; s(solar_dsm) 1.0	0.373	41.7	401
s(x, y) + exposure_dsm	s(x,y) <0.001; exposure_dsm 0.235	s(x,y) 2.4	0.351	38.6	425
s(x, y) + s(twi)	s(x,y) <0.001; s(twi) 0.345	s(x,y) 3.2; s(twi) 3.2	0.408	46.8	424
s(x, y) + s(tpi500)	s(x,y) <0.001; s(tpi500) 0.007	s(x,y) 2.0; s(tpi500) 4.5	0.491	54.3	414
s(x, y) + s(tpi90)	s(x,y) <0.001; s(tpi90) 0.112	s(x,y) 3.3; s(tpi90) 2.4	0.421	47.3	422
s(x, y) + s(tpi30)	s(x,y) <0.001; s(tpi30) 0.269	s(x,y) 3.1; s(tpi30) 1.0	0.367	40.8	425
s(x, y) + s(taynish_dist)	s(x,y) 0.099; s(taynish_dist) 0.478	s(x,y) 8.6; s(taynish_dist) 1.0	0.51	58.5	416
s(x, y) + s(lake_dist)	s(x,y) 0.002; s(lake_dist) 0.018	s(x,y) 2.0; s(lake_dist) 2.9	0.437	48	418
s(x, y) + s(sea_dist)	s(x,y) 0.019; s(sea_dist) 0.316	s(x,y) 7; s(sea_dist) 1	0.464	53.2	420

Table 2: Performance of spatial metrics with spatial smooth for richness.

Richness Terms	**p values**	**EDFs**	**Adj R2**	**Dev. Explained**	**AIC**
s(x, y) + s(dtm)	s(x,y) <0.001; s(dtm) 0.205	s(x,y) 7.5; s(dtm) 1.0	0.439	46	337
s(x, y) + s(slope.15)	s(x,y) <0.001; s(slope.15) 0.661	s(x,y) 7.9; s(slope.15) 1.0	0.455	46.5	320
s(x, y) + s(aspect.15)	s(x,y) <0.001; s(aspect.15) 0.685	s(x,y) 7.4; s(aspect.15) 1.3	0.448	45.8	321
s(x, y) + s(solar)	s(x,y) <0.001; s(solar) 0.260	s(x,y) 7.8; s(solar) 1.0	0.467	47.2	319
s(x, y) + s(solar_dsm)	s(x,y) <0.001; s(solar_dsm) 0.520	s(x,y) 7.8; s(solar_dsm) 1.0	0.457	46.4	320
s(x, y) + exposure_dsm	s(x,y) <0.001; exposure_dsm 0.160	s(x,y) 6.6	0.434	42.9	339
s(x, y) + s(twi)	s(x,y) <0.001; s(twi) 0.309	s(x,y) 8.3; s(twi) 2.4	0.477	50	336
s(x, y) + s(tpi500)	s(x,y) <0.001; s(tpi500) <0.001	s(x,y) 2.0; s(tpi500) 4.5	0.406	44.2	333
s(x, y) + s(tpi90)	s(x,y) <0.001; s(tpi90) 0.422	s(x,y) 6.8; s(tpi90) 2.0	0.435	45.1	339
s(x, y) + s(tpi30)	s(x,y) <0.001; s(tpi30) 0.443	s(x,y) 7.7; s(tpi30) 1.0	0.447	45.2	338
s(x, y) + s(taynish_dist)	s(x,y) 0.070; s(taynish_dist) 0.800	s(x,y) 8.1; s(taynish_dist) 1.0	0.467	47.7	335

s(x, y) + s(lake_dist)	s(x,y) <0.001; s(lake_dist) <0.001	s(x,y) 2.0; s(lake_dist) 3.1	0.454	42.2	333
s(x, y) + s(sea_dist)	s(x,y) <0.001; s(sea_dist) 0.494	s(x,y) 7.8; s(sea_dist) 1.0	0.454	46.6	337

Please see the GAM results of spatial metrics without the spatial smooth for both RRA score (Table 3) and richness (Table 4).

Table 3: Performance of spatial metrics without spatial smooth for RRA score.

Terms	**p values**	**EDFs**	**Adj R2**	**Dev. Explained**	**AIC**
s(dtm)	s(dtm) 0.158	s(dtm) 3.3	0.092	14	448
s(slope.15)	s(slope.15) 0.993	s(slope.15) 1	-0.017	0	426
s(aspect.15)	s(aspect.15) 0.062	s(aspect.15) 1	0.043	5.9	423
s(solar)	s(solar) 0.484	s(solar) 1	-0.009	0.9	426
s(solar_dsm)	s(solar_dsm) 0.665	s(solar_dsm) 1	-0.014	0.3	426
exposure_dsm	exposure_dsm 0.588	NA	-0.011	0.5	451
s(twi)	s(twi) 0.277	s(twi) 2.3	0.046	8	449
s(tpi500)	s(tpi500) 0.959	s(tpi500) 1	-0.016	0	451
s(tpi90)	s(tpi90) 0.235	s(tpi90) 1	0.007	2.3	450
s(tpi30)	s(tpi30) 0.209	s(tpi30) 1	0.01	2.5	450
s(taynish_dist)	s(taynish_dist) <0.001	s(taynish_dist) 1	0.354	36.5	422
s(lake_dist)	s(lake_dist) <0.001	s(lake_dist) 3.2	0.335	36.9	427
s(sea_dist)	s(sea_dist) <0.001	s(sea_dist) 1	0.276	28.7	430

Table 4: Performance of spatial metrics without spatial smooth for richness.

Richness Terms	**p values**	**EDFs**	**Adj R2**	**Dev. Explained**	**AIC**
s(dtm)	s(dtm) 0.003	s(dtm) 6.2	0.154	20.1	370
s(slope.15)	s(slope.15) 0.398	s(slope.15) 1	-0.011	0.5	370
s(aspect.15)	s(aspect.15) 0.011	s(aspect.15) 4.9	0.093	14.2	359
s(solar)	s(solar) 0.460	s(solar) 1.1	-0.012	0.5	370
s(solar_dsm)	s(solar_dsm) 0.049	s(solar_dsm) 1	0.014	2.6	366
exposure_dsm	exposure_dsm 0.077	NA	0.01	2	385
s(twi)	s(twi) 0.133	s(twi) 2.1	0.019	4.1	385
s(tpi500)	s(tpi500) 0.266	s(tpi500) 1.5	0.001	2.1	387
s(tpi90)	s(tpi90) 0.764	s(tpi90) 1	-0.015	0.1	388
s(tpi30)	s(tpi30) 0.414	s(tpi30) 1.7	-0.003	1.9	388
s(taynish_dist)	s(taynish_dist) <0.001	s(taynish_dist) 1	0.354	30.8	341
s(lake_dist)	s(lake_dist) <0.001	s(lake_dist) 3.5	0.314	31	347
s(sea_dist)	s(sea_dist) <0.001	s(sea_dist) 1	0.212	20.7	357

Interaction Terms

As mentioned in the Research Paper, interaction terms were explored - none of which were significant. The terms were selected based on their applicability to management and practice. Within the `gam` function interaction terms are to be included in addition to the main effects, to ensure the significance of the interaction is independent of the term on its own. As mentioned, this was attempted for a number of meaningful interactions such as canopy cover and canopy cover variance, canopy cover and TWI, canopy cover and TPI (all scales), canopy cover and ENL, ENL and points below 4 m, canopy cover and points below 4 m, and distance from Taynish and TPI.

Other interactions were attempted, the results of which can be found in the ‘gams’ folder, with folders including ‘ti’. Please see Appendix C for more information.

Technical Report References

Aalto, I., Aalto, J., Hancock, S., Valkonen, S. and Maeda, E.E. (2023). Quantifying the impact of management on the three-dimensional structure of boreal forests. *Forest Ecology and Management*, 535. DOI: https://doi.org/10.1016/j.foreco.2023.120885.

Ankori-Karlinsky, R., Hall, J., Murphy, L., Muscarella, R., Martinuzzi, S., Fahey, R., Zimmerman, J.K. and Uriarte, M. (2024). Chronic winds reduce tropical forest structural complexity regardless of climate, topography, or forest age. *Ecosystems*, 27(3), pp. 479–491. DOI: https://doi.org/10.1007/s10021-024-00900-5.

Boose, E. (2024). *ExposR: Models Topographic Exposure to Hurricane Winds*. R package version 1.2. https://cran.r-project.org/web/packages/ExposR/.

Coops, N.C., Tompaski, P., Nijland, W., Rickbeil, G.J.M., Nielsen, S.E., Bater, C.W. and Stadt, J.J. (2016). A forest structure habitat index based on airborne laser scanning data. *Ecological Indicators*, 67, pp.346–357. DOI: https://doi.org/10.1016/j.ecolind.2016.02.057.

Ehbrecht, M., Schall, P., Juchheim, J., Ammer, C. and Seidel, D. (2016). Effective number of layers: A new measure for quantifying three-dimensional stand structure based on sampling with terrestrial LiDAR. *Forest Ecology and Management*, 380, pp.212–223. DOI: https://doi.org/10.1016/j.foreco.2016.09.003.

Ellis, C.J. and Eaton, S. (2021). Microclimates Hold the Key to Spatial Forest Planning Under Climate Change: Cyanolichens in Temperate Rainforest. *Global Change Biology*, 27(9). DOI: https://doi.org/10.1111/gcb.15514.

Esri. (2024). ArcGIS Pro version 3.3.0. Available at: https://www.esriuk.com/en-gb/arcgis/products/arcgis-pro/overview.

Evans, M., Moustakas, A., Carey, G. and others (2015). Allometry and growth of eight tree taxa in United Kingdom woodlands. *Scientific Data*, 2, 150006. DOI: https://doi.org/10.1038/sdata.2015.6.

Eysn, L., Hollaus, M., Lindberg, E., Berger, F., Monnet, J.-M., Dalponte, M., Kobal, M., Pellegrini, M., Lingua, E., Mongus, D. and Pfeifer, N. (2015). A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space. *Forests*, 6(5), pp. 1721–1747. DOI: https://doi.org/10.3390/f6051721.

Hamilton, G.J. (1988). Forestry Commission Booklet 39. *Forest Mensuration Handbook*. https://cdn.forestresearch.gov.uk/2022/02/fcbk039.pdf.

Hamner, B., Frasco, M., and LeDell, E. (2018). *Metrics: Evaluation Metrics for Machine Learning*. R package version 0.1.4, https://cran.r-project.org/web/packages/Metrics/index.html.

Hijmans, R. (2025). *terra: Spatial Data Analysis*. R package version 1.8-61, https://github.com/rspatial/terra.

Hill, M.O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54(2), pp.427–432. DOI: https://doi.org/10.2307/1934352.

Ilich, A., Misiuk, B., Lecours, V., and Murawski, S. (2021). *MultiscaleDTM: Multi-Scale Geomorphometric Terrain Attributes.* R package version 1.0. https://cran.r-project.org/web/packages/MultiscaleDTM/index.html.

Ilich, A., Misiuk, B., Lecours, V., and Murawski, S. (2023). "MultiscaleDTM: An open-source R package for multiscale geomorphometric analysis." *Transactions in GIS*, 27(4). DOI: https://doi.org/10.1111/tgis.13067.

Kankare, V., Liang, X., Vastaranta, M., Yu, X., Holopainen, M. and Hyyppä, J. (2015). Diameter distribution estimation with laser scanning based multisource single tree inventory. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, pp. 161–171. DOI: https://doi.org/10.1016/j.isprsjprs.2015.07.007.

Khosravipour, A., Skidmore, A.K., Isenburg, M., Wang, T. and Hussin, Y.A. (2014). Generating Pit-free Canopy Height Models from Airborne Lidar. *Photogrammetric Engineering & Remote Sensing*, 80(9), pp.863–872. DOI: https://doi.org/10.14358/pers.80.9.863.

Kopecký, M., Macek, M. & Wild, J. (2021). Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of the Total Environment*, 757, 143785. DOI: https://doi.org/10.1016/j.scitotenv.2020.143785.

Lindsay, J.B. (2016). "Whitebox GAT: A case study in geomorphometric analysis." *Computers & Geosciences*, 95, 75-84. DOI: http://dx.doi.org/10.1016/j.cageo.2016.07.003.

Met Office. (2016). Western Scotland: Climate. *Wind*. Available at: https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/weather/learn-about/weather/regional-climates/western-scotland_-climate-met-office.pdf. Accessed: 06/08/2025.

Naesset, E. and Oklund, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79(1), pp.105–115. DOI: https://doi.org/10.1016/s0034-4257(01)00243-7.

NatureScot. (2007). Taynish Woods Site of Special Scientific Interest. *Map.* Available at: https://www.nature.scot/sites/default/files/site-special-scientific-interest/1522/sssi-map.pdf. Accessed: 06/08/2025.

NatureScot. (2011). Taynish Woods Site of Special Scientific Interest. *Site Code 1522 Citation*. Available at: https://www.nature.scot/sites/default/files/site-special-scientific-interest/1522/sssi-citation.pdf. Accessed: 06/08/2025.

Ordnance Survey. (2025a). OS OpenMap. *Local.* Available at: https://www.ordnancesurvey.co.uk/products/os-open-map-local. Accessed: 06/08/2025.

Ordnance Survey. (2025b). OS Open. *Rivers.* Available at: https://www.ordnancesurvey.co.uk/products/os-open-rivers. Accessed: 06/08/2025.

Palmroos, I., Norros, V., Keski-Saari, S., Mäyrä, J., Tanhuanpää, T., Kivinen, S., Pykälä, J., Kullberg, P., Kumpula, T. and Vihervaara, P. (2023). Remote sensing in mapping biodiversity – A case study of epiphytic lichen communities. *Forest Ecology and Management*, 538. DOI: https://doi.org/10.1016/j.foreco.2023.120993.

Pebesma, E. (2018). *sf: Simple Features for R*. R package version 1.0-21. https://cran.r-project.org/web/packages/sf/index.html.

Pebesma, E., and Bivand, R. (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC Press, New York. DOI: doi:10.1201/9780429459016.

Roussel J, Auty D (2025). *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. R package version 4.2.1, https://cran.r-project.org/package=lidR.

Roussel, J., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R., Meador, A.S., Bourdon, J., de Boissieu, F., and Achim, A. (2020). “lidR: An R package for analysis of Airborne Laser Scanning (ALS) data.” *Remote Sensing of Environment*, 251, 112061. ISSN 0034-4257, DOI: https://doi.org/10.1016/j.rse.2020.112061.

Roussel, J., Goodbody, T.R., and Tompalski, P. (2024). The lidR package. *Digital Surface Model and Canopy Height Model.* Available at: https://r-lidar.github.io/lidRbook/dsm.html. Accessed: 06/08/2025.

Scottish Forestry. (2024). Native Woodland Survey of Scotland. *Data Explorer.* Available at: https://www.forestry.gov.scot/forests-environment/biodiversity/native-woodlands/native-woodland-survey-of-scotland-nwss. Accessed: 06/08/2025.

Simpson, G. (2024). *gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted using mgcv*. R package version 0.10.0, https://gavinsimpson.github.io/gratia/.

Wood, S.N. (2025). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.* R package version 1.9-3, https://cran.r-project.org/web/packages/mgcv/index.html**.**

Wu, Q. and Brown, A. (2022). *'whitebox': 'WhiteboxTools' R Frontend*. R package version 2.2.0, https://CRAN.R-project.org/package=whitebox.

Xiang Y, Gubian S, Suomela B, and Hoeng J. (2013). “Generalized Simulated Annealing for Efficient GlobalOptimization: the GenSA Package for R.” The R Journal, 5/1. https://journal.r-project.org.

Yang, Q., Su, Y., Jin, S., Kelly, M., Hu, T., Ma, Q., Li, Y., Song, S., Zhang, J., Xu, G., Wei, J. and Guo, Q. (2019). The Influence of Vegetation Characteristics on Individual Tree Segmentation Methods with Airborne LiDAR Data. *Remote Sensing*, 11(23). DOI: https://doi.org/10.3390/rs11232880.

Yu, X., Hyyppä, J., Vastaranta, M., Holopainen, M. and Viitala, R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS journal of photogrammetry and remote sensing*, 66(1), pp.28–37. DOI: https://doi.org/10.1016/j.isprsjprs.2010.08.003.

Appendices to the Technical Report

A – Scripts

Optimised Tree Detection (example for one site)

```
# Optimise individual tree detection for one site

# Load required libraries
library(SpaDES)
library(dplyr)
library(sf)
library(sp)
library(terra)
library(lidR)
library(GenSA)

# ------------------------------
# Step 1: Load and prepare data
```

```
# ----------------------------

# Read LAS file (height-normalised point cloud)
las <- readLAS("C:/Users/mclel/Documents/dissdata/outputs/cosh/norm_cosh.las")

# Read clipping extent and reference extent
extent <-
st_read("C:/Users/mclel/Documents/dissdata/outputs/groundtruth/cosh_ext_buff.gpkg")
ref_ext <-
st_read("C:/Users/mclel/Documents/dissdata/outputs/groundtruth/cosh_ref_ext.gpkg")

# Clip LAS to extent
las_clipped <- clip_roi(las, extent)

# Load field-mapped trees and filter to those above allometric threshold (6 m)
reference <-
st_read("C:/Users/mclel/Documents/dissdata/outputs/groundtruth/cosh_trees.gpkg") %>%
  mutate(
    x = st_coordinates(.)[, 1],
    y = st_coordinates(.)[, 2]
  ) %>%
  filter(height_m >= 6) %>%
  rename(Z = height_m)

# Ensure each reference tree has a unique ID
if (!"id" %in% colnames(reference)) {
  reference$id <- 1:nrow(reference)
}

cat("Number of reference trees after filtering:", nrow(reference), "\n")

# -------------------------------------
# Step 2: Define cost function for GenSA
# -------------------------------------

cost_function <- function(param) {
  # Extract CHM parameters
  res <- param[1]
  subc <- param[2]
  thresh <- c(0, param[3], param[4], param[5], param[6], param[7])  # pit-free
thresholds

  # Generate pit-free CHM
  chm <- rasterize_canopy(
    las = las_clipped,
    res = res,
    algorithm = pitfree(
      thresholds = thresh,
```

```
      max_edge = c(0, 1.5),
      subcircle = subc,
      highest = TRUE
    )
  )

  # Extract smoothing and LMF parameters
  w_size <- round(param[8])
  if (w_size %% 2 == 0) w_size <- w_size + 1  # ensure window size is odd

  gradient <- param[9]
  y_intercept <- param[10]

  cat("Running with parameters -",
      "Resolution:", res,
      "Subcircle:", subc,
      "Window size:", w_size,
      "Gradient:", gradient,
      "Y-intercept:", y_intercept, "\n")

  # Apply median filter to smooth CHM
  w <- matrix(1, w_size, w_size)
  smoothed_chm <- terra::focal(chm, w, fun = median, na.rm = TRUE)

  # Define variable window function for LMF
  vw <- function(x) {
    return(x * gradient + y_intercept)
  }

  # Locate tree tops using LMF
  ttops <- locate_trees(
    smoothed_chm,
    lmf(ws = vw)
  ) %>%
    filter(Z >= 6)  # filter trees under threshold

  # Accuracy assessment (see separate script)
  cost <- accuracy_assessment(test = ttops, reference = reference, extent = ref_ext)

  return(cost)
}

# ---------------------------------------------
# Step 3: Run initial GenSA to optimise settings
# ---------------------------------------------

starting_params <- c(
  0.235,  # CHM resolution
```

```
  0.07,    # Subcircle radius
  2, 5, 10, 15, 20,  # Pit-free thresholds (1-5)
  5,      # Smoothing window size
  0.07,    # LMF gradient
  0.19    # LMF y-intercept
)

# Run simulated annealing optimisation
sa_result <- GenSA(
  par = starting_params,
  fn = cost_function,
  lower = c(0.15, 0.01, 0, 5, 10, 15, 20, 5, 0.05, 0),
  upper = c(0.5,  0.5,  5, 10, 15, 20, 25, 9, 1,    3),
  control = list(
    verbose = TRUE,
    temperature = 200,
    maxit = 1000
  )
)

# Save output
sink("C:/Users/mclel/Documents/dissdata/outputs/cosh/site1_sa_result.txt")
print(sa_result)
sink()

# ---------------------------------------------
# Step 4: Fine-tune parameters with second GenSA
# ---------------------------------------------

starting_params <- sa_result$par  # use results of first run as new start

sa_result2 <- GenSA(
  par = starting_params,
  fn = cost_function,
  lower = c(0.15, 0.05, 3, 5, 10, 15, 20, 5, 0.01, 0),
  upper = c(0.3,  0.4,  5, 7,  15, 20, 23, 7, 0.1,  2),
  control = list(
    verbose = TRUE,
    temperature = 75,
    maxit = 1000
  )
)

# Save optimised result
saveRDS(sa_result2,
"C:/Users/mclel/Documents/dissdata/outputs/cosh/site1_sa_result2.rds")

# ------------------------------------------
```

```r
# Step 5: Apply optimal parameters to LAS data
# ------------------------------------------

# Reload LAS and optimal parameters
las <- readLAS("C:/Users/mclel/Documents/dissdata/outputs/cosh/norm_cosh.las")
sa_result2 <-
readRDS("C:/Users/mclel/Documents/dissdata/outputs/cosh/site1_sa_result2.rds")
param <- sa_result2$par

# Function to apply optimal settings and extract tree tops
optimal_ttops <- function(param) {
  res <- param[1]
  subc <- param[2]
  thresh <- c(0, param[3], param[4], param[5], param[6], param[7])

  chm <- rasterize_canopy(
    las = las,
    res = res,
    algorithm = pitfree(
      thresholds = thresh,
      max_edge = c(0, 1.5),
      subcircle = subc,
      highest = TRUE
    )
  )

  w_size <- round(param[8])
  if (w_size %% 2 == 0) w_size <- w_size + 1

  w <- matrix(1, w_size, w_size)
  smoothed_chm <- terra::focal(chm, w, fun = median, na.rm = TRUE)

  # Save CHM as raster for visualisation or export
  writeRaster(smoothed_chm,
"C:/Users/mclel/Documents/dissdata/outputs/cosh/smoothed_site1_pfchm.tif", overwrite =
TRUE)

  # Load raster to memory if needed
  smoothed_chm <-
rasterToMemory("C:/Users/mclel/Documents/dissdata/outputs/cosh/smoothed_site1_pfchm.tif"
)

  # Variable window function
  vw <- function(x) {
    return(x * param[9] + param[10])
  }

  # Locate tree tops
```

```
  ttops <- locate_trees(
    smoothed_chm,
    lmf(ws = vw)
  ) %>%
    filter(Z >= 6)

  return(ttops)
}

# Run final detection with optimal parameters
ttops <- optimal_ttops(param)

# Save detected trees as GeoPackage
st_write(ttops, "C:/Users/mclel/Documents/dissdata/outputs/trees/trees1_3D.gpkg")
```

Accuracy Assessment

```
# ITD accuracy assessment, as per Eysn et al. (2015): https://doi.org/10.3390/f6051721

accuracy_assessment <- function(ttops, reference_copy, extent) {

  # Extract XY coordinates from predicted tree tops
  ttops <- ttops %>%
    mutate(X = st_coordinates(.)[, 1], Y = st_coordinates(.)[, 2])

  # Convert both test and reference layers to data frames
  Test <- as.data.frame(ttops)
  Reference <- as.data.frame(reference_copy)
  Reference$matched <- FALSE  # track which reference trees are matched
  Test <- Test %>% arrange(desc(Z))  # prioritise tallest trees

  # -------------------------------------------------------
  # Helper function: search for potential matches
  # -------------------------------------------------------
  search_candidates <- function(test_tree, ref_unmatched, d2d_threshold = 5, h_threshold
= 5) {
    test_coords <- c(test_tree$X, test_tree$Y)
    dist2d <- spDistsN1(as.matrix(ref_unmatched[, c("x", "y")]), test_coords, longlat =
FALSE)

    potential <- ref_unmatched[
      dist2d <= d2d_threshold &
      abs(ref_unmatched$height_m - test_tree$Z) <= h_threshold,
    ]

    return(potential)
  }
```

```
  # -------------------------------------------------------
  # Helper function: choose the best candidate from search
  # -------------------------------------------------------
  vote_best_candidate <- function(test_tree, candidates) {
    if (nrow(candidates) == 0) {
      return(list(match_id = NA, delta_d2d = NA, delta_h = NA))
    }

    # Compute vertical and horizontal distance differences
    candidates <- candidates %>%
      mutate(
        delta_h = abs(height_m - test_tree$Z),
        delta_d2d = sqrt((x - test_tree$X)^2 + (y - test_tree$Y)^2)
      ) %>%
      filter(!is.na(delta_d2d), !is.na(delta_h), !is.na(id)) %>%
      arrange(delta_d2d)

    if (nrow(candidates) == 0) {
      return(list(match_id = NA, delta_d2d = NA, delta_h = NA))
    }

    # Initial best match
    best <- candidates[1, , drop = FALSE]

    # If more than one candidate, check for closer in height
    if (nrow(candidates) == 1) {
      return(list(match_id = best$id, delta_d2d = best$delta_d2d, delta_h =
best$delta_h))
    }

    for (i in 2:nrow(candidates)) {
      current <- candidates[i, , drop = FALSE]

      if (!is.na(current$delta_d2d) && !is.na(best$delta_d2d) &&
          current$delta_d2d <= (best$delta_d2d + 2.5)) {

        if (!is.na(current$delta_h) && !is.na(best$delta_h) &&
            current$delta_h < best$delta_h) {
          best <- current
        }
      } else {
        break
      }
    }

    return(list(match_id = best$id, delta_d2d = best$delta_d2d, delta_h = best$delta_h))
  }
```

```
# ------------------------------------------------------------------
# Helper function: check mutual nearest neighbour for valid matching
# ------------------------------------------------------------------
confirm_mutual_match <- function(test_idx, test_tree, reference_tree, test_df, d2d_threshold = 5, h_threshold = 5) {
  ref_id <- reference_tree$id
  ref_coords <- c(reference_tree$x, reference_tree$y)
  dist2d <- spDistsN1(as.matrix(test_df[, c("X", "Y")]), ref_coords, longlat = FALSE)

  nearby_tests <- test_df[
    dist2d <= d2d_threshold &
    abs(test_df$Z - reference_tree$height_m) <= h_threshold,
  ]

  if (nrow(nearby_tests) == 0) return(FALSE)

  nearby_tests <- nearby_tests %>%
    mutate(
      delta_h = abs(Z - reference_tree$height_m),
      delta_d2d = sqrt((X - reference_tree$x)^2 + (Y - reference_tree$y)^2)
    ) %>%
    filter(!is.na(delta_d2d) & !is.na(delta_h)) %>%
    arrange(delta_d2d, delta_h)

  best_test_id <- nearby_tests[1, "id"]
  return(test_tree$id == best_test_id)
}

# Ensure test trees have unique IDs
if (!"id" %in% colnames(Test)) {
  Test$id <- 1:nrow(Test)
}

# Initialise match tracking
matches <- vector("list", length = nrow(Test))

# ------------------------------------------------
# Main loop: attempt to match each test tree
# ------------------------------------------------
for (i in seq_len(nrow(Test))) {
  test_tree <- Test[i, ]
  ref_unmatched <- Reference[!Reference$matched, ]

  candidates <- search_candidates(test_tree, ref_unmatched)
  vote <- vote_best_candidate(test_tree, candidates)

  if (is.na(vote$match_id)) {
    matches[[i]] <- list(match_id = NA)
```

```
      next
    }

    ref_tree <- Reference[Reference$id == vote$match_id, ]
    mutual <- confirm_mutual_match(i, test_tree, ref_tree, Test)

    if (mutual) {
      Reference$matched[Reference$id == vote$match_id] <- TRUE
      matches[[i]] <- list(match_id = vote$match_id)
    } else {
      matches[[i]] <- list(match_id = NA)
    }
  }

  # ------------------------------------------------
  # Calculate accuracy metrics
  # ------------------------------------------------
  match_df <- do.call(rbind, lapply(matches, as.data.frame))

  TP <- sum(!is.na(match_df[, "match_id"]))  # True positives
  FP <- sum(is.na(match_df[, "match_id"]))   # False positives
  FN <- sum(!Reference$matched)              # False negatives

  Omission_error <- FN / (TP + FN)
  Commission_error <- FP / (TP + FP)
  Overall_Accuracy <- (TP / (TP + FP + FN)) * 100
  cost <- 100 - Overall_Accuracy

  print(list(
    TP = TP,
    FP = FP,
    FN = FN,
    Cost = cost,
    Overall_Accuracy = Overall_Accuracy
  ))

  return(cost)
}
```

Metrics

```
# all the metrics, calculated from the las files and mosaiced
library(lidR)
library(terra)
library(ggplot2)
library(patchwork)

# read in the las files
```

```
ctg <- readLAScatalog("C:/Users/mclel/Documents/dissdata/outputs/nlas")
opt_merge(ctg) <- TRUE

template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")


# rh99
rh99 <- pixel_metrics(ctg,
                      ~quantile(Z, probs = 0.99, na.rm = TRUE),
                      res = 1)

rh99.15 <- resample(rh99, template, method = "average")




# canopy cover (pts above 2m) - 1m resolution

cc <- pixel_metrics(
  ctg,
  ~sum(Z > 2) / length(Z) * 100,
  res = 1
)

cc_15 <- resample(cc, template, method = "average")

# sd of cc

cc_sd <- aggregate(cc, fact = 15, fun = function(x) {
  x <- x[!is.na(x)]
  if (length(x) >= 2) sd(x) else NA
})
cc_sd_15<- resample(cc_sd, template, method = "near")
cc_sd_15[is.na(cc_sd_15)] <- 0

writeRaster(cc, "C:/Users/mclel/Documents/dissdata/outputs/cc_1m.tif")

plot(cc)
plot(cc_15)
plot(cc_sd)
plot



# enl (1D and 2D)
enl1d.fn <- function(Z) {
  Z <- Z[Z>0.3]
```

```
  if (length(Z) == 0) return(1)

  h <- hist(Z,
            breaks = seq(0.3,
                         30,
                         by = 1
            ),
            plot = FALSE
  )

  p <- h$counts / sum(h$counts)

  p_nonzero <- p[p > 0]

  enl <- exp(-sum(p_nonzero * log(p_nonzero)))

  return(enl)
}

enl2d.fn <- function(Z) {

  Z <- Z[Z>0.3]

  if (length(Z) == 0) return(1)

  h <- hist(Z,
            breaks = seq(0.3,
                         30,
                         by = 1
            ),
            plot = FALSE
  )

  p <- h$counts / sum(h$counts)

  p_nonzero <- p[p > 0]

  enl <- 1 / sum(p_nonzero^2)

  return(enl)
}


enl1D <- pixel_metrics(ctg,
                       ~enl1d.fn(Z),
                       res = template)

enl2D <- pixel_metrics(ctg,
                       ~enl2d.fn(Z),
```

```
                       res = template)





# standard metrics from lidR

template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")

metrics <- pixel_metrics(
  ctg,
  .stdmetrics,
  template
)



ent <- pixel_metrics(
  ctg,
  ~entropy(Z[Z>0.5], by = 2, zmax = 30),
  res = template
)

ent[is.na(ent)] <- 0

plot(ent)






# dtm for the whole estate
ctg <- readLAScatalog("C:/Users/mclel/Documents/dissdata/las")
opt_merge(ctg) <- TRUE
projection(ctg) <- "EPSG:27700"

dtm <- rasterize_terrain(
  ctg,
  res = 15,
  algorithm = knnidw(
    k = 6,
    p = 2,
    rmax = 30
  )
)

dsm <- rasterize_canopy(ctg, res = 15, algorithm = pitfree())
```

```
aspect <- terrain(dtm, v = "aspect", unit = "degrees", neighbors = 8)

slope <- terrain(dtm, v = "slope", unit = "degrees")

template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")

dtm.15 <- resample(dtm, template, method = "near")
aspect.15 <- resample(aspect, template, method = "near")
slope.15 <- resample(slope, template, method = "near")


# slope

dtm <- rast("C:/Users/mclel/Documents/dissdata/outputs/dtm.tif")
template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")

slope_1m <- terrain(dtm, v = "slope", unit = "degrees")

slope_15m <- resample(slope_1m, template, method = "average")

sd_dtm_15m <- aggregate(dtm, fact = 15, fun = sd)
sd_slope_15m   <- aggregate(slope_1m, fact = 15, fun = sd)

sd_dtm <- resample(sd_dtm_15m, template, method = "near")
sd_slope <- resample(sd_slope_15m, template, method = "near")



# solar

# dsm

dsm <- rasterize_canopy(ctg, res = 15, algorithm = pitfree())

solar_raw <-
rast("C:/Users/mclel/Documents/dissdata/outputs/sunlight_dsm/solar_dsm.tif")

template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")

solar_resamp <- resample(solar_raw, template, method = "near")
```

```
# first returns > 2m

# canopy cover (>2m) across the whole estate
ccfr.2m <- pixel_metrics(
  ctg,
  ~sum(Z > 2 & ReturnNumber == 1) / sum(ReturnNumber == 1) * 100,
  res = template
)

# canopy cover (>6m) across the whole estate
ccfr.6m <- pixel_metrics(
  ctg,
  ~sum(Z > 6 & ReturnNumber == 1) / sum(ReturnNumber == 1) * 100,
  res = template
)


# % points below 2m
pts.2m <- pixel_metrics(
  ctg,
  ~sum(Z > 0.3 & Z <= 2) / length(Z) * 100,
  res = template
)



# % points below 6m
pts.6m <- pixel_metrics(
  ctg,
  ~sum(Z > 0.3 & Z <= 6) / length(Z) * 100,
  res = template
)

# % points below 4m
pts.4m <- pixel_metrics(
  ctg,
  ~sum(Z > 0.3 & Z <= 4) / length(Z) * 100,
  res = template
)



# wind exposure raster
library(ExposR)
library(terra)
```

```
# check NA
dem <- rast("C:/Users/mclel/Documents/dissdata/outputs/sunlight_dsm/dsm.tif")
sum(values(dem) == 0, na.rm = TRUE)
sum(is.na(values(dem)))

# change na to 0

values(dem)[is.na(values(dem))] <- 0
writeRaster(dem, "C:/Users/mclel/Documents/dissdata/outputs/dsm_15_nona.tif",
overwrite=TRUE)

# create file
file.copy("C:/Users/mclel/Documents/dissdata/outputs/dsm_15_nona.tif",
          "C:/Users/mclel/Documents/dissdata/outputs/expos/dem/dem.tif",
          overwrite = TRUE)

# check crs
dem <- rast("C:/Users/mclel/Documents/dissdata/outputs/expos/dem/dem.tif")
crs(dem)

expos_set_path("C:/Users/mclel/Documents/dissdata/outputs/expos")

expos_model(
  wind_direction = 225,
  inflection_angle = 6,
  lat_long = FALSE,
  orient = 0,
  save = TRUE
)

expos_plot("expos-225-06", vector = FALSE)

# resample
template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")
exposure_dsm <- rast("C:/Users/mclel/Documents/dissdata/outputs/expos/exposure/expos-
225-06.tif")

expos_dsm <- resample(exposure_dsm, template, method = "near")




# distance from Taynish

taynish <- rast("C:/Users/mclel/Documents/dissdata/outputs/taynish_dist.tif")
template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")


tay_resamp <- resample(taynish, template, method = "near")
```

```r
# twi
library(whitebox)
wbt_init(exe_path = "C:/whitebox/whitebox_tools.exe")
library(terra)

# Fill depressions (sinks)
wbt_fill_depressions(
  dem = "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/dtm.tif",
  output = "C:/Users/mclel/Documents/dissdata/outputs/whtbx/filled.dem.tif"
)

wbt_slope(
  dem = "C:/Users/mclel/Documents/dissdata/outputs/whtbx/filled.dem.tif",
  output = "C:/Users/mclel/Documents/dissdata/outputs/whtbx/slope_15m.tif",
  units = "radians"
)

wbt_fd8_flow_accumulation(
  dem = "C:/Users/mclel/Documents/dissdata/outputs/whtbx/filled.dem.tif",
  output = "C:/Users/mclel/Documents/dissdata/outputs/whtbx/flowacc_fd8_15m.tif",
  out_type = "Specific Contributing Area"
)

slope <- rast("C:/Users/mclel/Documents/dissdata/outputs/whtbx/slope_15m.tif")
flowacc <- rast("C:/Users/mclel/Documents/dissdata/outputs/whtbx/flowacc_fd8_15m.tif")

twi <- log(flowacc / tan(slope))

template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")
twi_resamp <- resample(twi, template, method = "near")


#tpi
library(MultiscaleDTM)

dtm <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/dtm.tif")

tpi30 <- TPI(dtm, w = 500, shape = "circle", unit = "map", na.rm = TRUE) # 90 selected based on measurement in arc

tpi_resamp <- resample(tpi30, template, method = "near")

plot(tpi30)

writeRaster(tpi_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/tpi500.tif", overwrite=TRUE)
```

```
# distance to water

lakes <- rast("C:/Users/mclel/Documents/dissdata/outputs/water/surface_dist.tif")
coast <- rast("C:/Users/mclel/Documents/dissdata/outputs/water/foreshore_dist.tif")
combo <- rast("C:/Users/mclel/Documents/dissdata/outputs/water/combo.tif")
template <- rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif")

lakes_resamp <- resample(lakes, template, method = "near")
coast_resamp <- resample(coast, template, method = "near")
combo_resamp <- resample(combo, template, method = "near")

writeRaster(lakes_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/lake_dist.tif", overwrite=TRUE)
writeRaster(coast_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/sea_dist.tif", overwrite=TRUE)
writeRaster(coast_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/water_dist.tif",
overwrite=TRUE)

# save rasters

writeRaster(canopy_cover.2m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc2m.tif", overwrite = T)
writeRaster(canopy_cover.6m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc6m.tif", overwrite = T)
writeRaster(rh99.15, "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/rh99.tif",
overwrite = T)
writeRaster(enl1D, "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/enl1d.tif",
overwrite = T)
writeRaster(enl2D, "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/enl2d.tif",
overwrite = T)
writeRaster(metrics,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/lidR_metrics.tif", overwrite =
T)
writeRaster(dtm, "C:/Users/mclel/Documents/dissdata/outputs/dtm.tif", overwrite = T)
writeRaster(dtm.15, "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/dtm.tif",
overwrite = T)
writeRaster(aspect, "C:/Users/mclel/Documents/dissdata/outputs/aspect.tif", overwrite =
T)
writeRaster(aspect.15,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/aspect.15.tif", overwrite = T)
writeRaster(ent, "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/entropy.tif",
overwrite = T)
writeRaster(ccfr.2m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/ccfr_2m.tif", overwrite = T)
writeRaster(ccfr.6m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/ccfr_6m.tif", overwrite = T)
```

```
writeRaster(pts.2m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/pts_below2m.tif", overwrite =
T)
writeRaster(pts.6m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/pts_below6m.tif", overwrite =
T)
writeRaster(pts.4m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/pts_below4m.tif", overwrite =
T)
writeRaster(pts.rh50,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/pts_below50rh.tif", overwrite =
T)
writeRaster(sd_dtm,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/sd_dtm.tif", overwrite = T)
writeRaster(sd_slope,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/sd_slope.tif", overwrite = T)
writeRaster(slope.15,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/slope.15.tif", overwrite = T)
writeRaster(solar_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/solar_dsm.tif", overwrite = T)
writeRaster(resampled_sl_15m_sd,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/sunlight_sd.tif", overwrite =
T)
writeRaster(dsm, "C:/Users/mclel/Documents/dissdata/outputs/sunlight_dsm/dsm.tif",
overwrite = T)
writeRaster(dsm_sl_15m,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/dsm_sunlight_15.tif", overwrite
= T)
writeRaster(resampled_dsm_sl_15m_sd,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/dsm_sunlight_sd.tif", overwrite
= T)
writeRaster(exposure_dsm, "C:/Users/mclel/Documents/dissdata/outputs/expos_dsm.tif",
overwrite = T)
writeRaster(expos_dsm,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/exposure_dsm.tif", overwrite =
T)
writeRaster(exposure_dsm_15,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/expos_dsm.tif", overwrite = T)
writeRaster(exposure_dtm_15,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/expos_dtm.tif", overwrite = T)
writeRaster(exposure_dsm_sd,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/expos_dsm_sd.tif", overwrite =
T)
writeRaster(exposure_dtm_sd,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/expos_dtm_sd.tif", overwrite =
T)
writeRaster(cc_15, "C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc_avg.tif",
overwrite = T)
```

```r
writeRaster(cc_sd_15,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/cc_sd.tif", overwrite = T)
writeRaster(tay_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/taynish_dist.tif", overwrite =
T)
writeRaster(twi_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/twi.tif", overwrite = T)
writeRaster(tpi_resamp,
"C:/Users/mclel/Documents/dissdata/outputs/final_metrics/tpi30.tif", overwrite=TRUE)

# save the rasters to the data frame
library(terra)
library(dplyr)
library(tools)
library(ggplot2)

# read vector points and centroids
rra_scores <- vect("C:/Users/mclel/Documents/dissdata/field/site_scores.gpkg")
rra_scores <- centroids(rra_scores)

# list all tif files except "lidr_metrics.tif" as it is a multi-layer raster
rfiles <- list.files(
  path = "C:/Users/mclel/Documents/dissdata/outputs/final_metrics",
  pattern = "\\.tif$",
  full.names = TRUE
)
rfiles <- rfiles[!basename(rfiles) %in% "lidR_metrics.tif"]

# turn them into spatrasters
rasters <- lapply(rfiles, rast)

names(rasters) <- file_path_sans_ext(basename(rfiles))

# load lidr_metrics
lidr_metrics <-
rast("C:/Users/mclel/Documents/dissdata/outputs/final_metrics/lidr_metrics.tif")

# split lidr_metrics into a list of single-layer rasters
lidr_list <- lapply(1:nlyr(lidr_metrics), function(i) lidr_metrics[[i]])

# name these layers
names(lidr_list) <- names(lidr_metrics)

# combine all single-layer rasters into one list
rlist <- c(rasters, lidr_list)

# normalise each raster
normalise_raster <- function(r) {
```

```
  (r - minmax(r)[1]) / (minmax(r)[2] - minmax(r)[1])
}

norm_list <- lapply(rlist, normalise_raster)

# create a SpatRaster stack from the list
rstack <- rast(norm_list)

# extract values
extracted <- as.data.frame(extract(rstack, rra_scores, bind = TRUE))


coords <- crds(rra_scores)
extracted$x <- coords[, 1]
extracted$y <- coords[, 2]

head(extracted)

base::saveRDS(extracted,
"C:/Users/mclel/Documents/dissdata/outputs/relationships/norm_dataframe.Rda")

# non-normalised spatraster stack

rstack <- rast(rlist)

# extract values
extracted <- as.data.frame(extract(rstack, rra_scores, bind = TRUE))

# add coordinates of sites to data frame
coords <- crds(rra_scores)
extracted$x <- coords[, 1]
extracted$y <- coords[, 2]

head(extracted)

base::saveRDS(extracted,
"C:/Users/mclel/Documents/dissdata/outputs/relationships/dataframe.Rda")
```

Fieldwork data digitisation

```
# read in the GPX sample sites, combine into one gpkg and filter out the extra sites

library(sf)
library(dplyr)

first <- st_read("C:/Users/mclel/Documents/dissdata/field/Waypoints_08-JUN-25.gpx")

second <- st_read("C:/Users/mclel/Documents/dissdata/field/Waypoints_09-JUN-25.gpx")
```

```
first.df <- as.data.frame(first)

second.df <- as.data.frame(second)

View(first.df)
View(second.df)

combined <- rbind(first.df, second.df)

View(combined)

# filter out accidental points

filtered <- filter(combined, !name %in% c("038", "039", "046", "049", "050", "055"))

View(filtered)

st_write(filtered, "C:/Users/mclel/Documents/dissdata/field/sites.gpkg")



# clear env and read in the sites

sites <- st_read("C:/Users/mclel/Documents/dissdata/field/sites.gpkg")
View(sites)







# now create the 4 polygons for each site (the sites just created are the centroids)

library(terra)

centroids_84 <- vect("C:/Users/mclel/Documents/dissdata/field/sites.gpkg")

centroids <- project(centroids_84, "EPSG:27700")

# create a square poly
create_square <- function(pt, dx, dy) {
  x <- crds(pt)[1,1]
  y <- crds(pt)[1,2]

  coords <- matrix(c(
    x,      y,
    x+dx,   y,
```

```
    x+dx,  y+dy,
    x,     y+dy,
    x,     y
  ), ncol = 2, byrow = TRUE)

  vect(list(coords), type = "polygons", crs = crs(pt))
}

# create all squares for each point
create_all_squares <- function(pt) {
  sq_NW <- create_square(pt, -15,  15)  # Centroid = bottom-right corner
  sq_SW <- create_square(pt, -15, -15)  # Centroid = top-right corner
  sq_SE <- create_square(pt,  15, -15)  # Centroid = top-left corner
  sq_NE <- create_square(pt,  15,  15)  # Centroid = bottom-left corner
  rbind(sq_NW, sq_SW, sq_SE, sq_NE)
}

# loop over each centroid (row) and make squares
all_squares <- lapply(1:nrow(centroids), function(i) {
  pt <- centroids[i]
  create_all_squares(pt)
})

# Combine all into one SpatVector
squares <- do.call(rbind, all_squares)

squares$id <- 1:length(squares)

writeVector(squares, "C:/Users/mclel/Documents/dissdata/field/site_polys.gpkg",
overwrite = T)









# read in the new gpkg and the attributes, add the attributes and save as new gpkg

library(sf)

sites <- st_read("C:/Users/mclel/Documents/dissdata/field/site_polys.gpkg")

attributes <- read.csv("C:/Users/mclel/Documents/dissdata/field/site_attributes.csv")

merged <- merge(sites, attributes, by = "id")

site_ids <- rep(1:(nrow(merged)/4), each=4)
```

```
merged$site_id <- site_ids

merged <- merged[, c("id", "site_id", "GPS_site", "score", "indicators", "lb_scores",
"geom")]

st_write(merged, "C:/Users/mclel/Documents/dissdata/field/site_scores.gpkg", delete_dsn
= TRUE)


# clear env, read in gpkg and convert to raster for later prediction/classification
library(sf)
library(terra)


site_polys <- st_read("C:/Users/mclel/Documents/dissdata/field/site_scores.gpkg")

polygons_list <- split(site_polys, site_polys$site_id)

site_rasters <- list()

for(i in seq_along(polygons_list)) {

  site_poly <- polygons_list[[i]]

  r_template <- rast(ext(site_poly), resolution=15, crs = "EPSG:27700")

  site_rasters[[i]] <- rasterize(vect(site_poly), r_template, field = "score")
}

for (i in seq_along(site_rasters)) {
  writeRaster(site_rasters[[i]],
              filename =
paste0("C:/Users/mclel/Documents/dissdata/outputs/rra_scores/site_", i, "_RRA.tif"),
              overwrite = TRUE)
}







# the sites to be classified into the rra bands

library(stringr)
library(terra)
```

```
rras <- list.files(path = "C:/Users/mclel/Documents/dissdata/outputs/rra_scores",
pattern = "\\.tif$", full.names = TRUE)

site_nums <- as.numeric(str_extract(basename(rras), "(?<=site_)\\d+"))

rras_sorted <- rras[order(site_nums)]

rra_scores <- lapply(rras_sorted, rast)

rra.df <- as.data.frame(rra_scores)

View(rra.df)

# after viewing the data, the bands decided are 5=poor, 5-14=fair and =>15=good

classify_rra <- function(x){

  classified <- classify(x, rcl = matrix(c(
    -10, 5, 1,
    5, 15, 2,
    15, Inf, 3
  ), ncol = 3, byrow = TRUE)
  )

  return(classified)

}

rra_classes <- list()

for(i in seq_along(rra_scores)){

  x <- rra_scores[[i]]

  rra_class <- classify_rra(x)

  rra_classes[[i]] <- rra_class

}


output_folder <- "C:/Users/mclel/Documents/dissdata/outputs/rra_classes"
dir.create(output_folder, showWarnings = FALSE)

for (i in seq_along(rra_classes)) {
  writeRaster(rra_classes[[i]],
              filename = file.path(output_folder, paste0("site_", i, "_class.tif")),
              overwrite = TRUE)
}
```

```
rras <- list.files(path = "C:/Users/mclel/Documents/dissdata/outputs/rra_classes",
pattern = "\\.tif$", full.names = TRUE)

site_nums <- as.numeric(str_extract(basename(rras), "(?<=site_)\\d+"))

rras_sorted <- rras[order(site_nums)]

rra_scores <- lapply(rras_sorted, rast)

rra.df <- as.data.frame(rra_scores)

View(rra.df)
table(unlist(rra.df))

mosaic_rra <- do.call(mosaic, rra_scores)

plot(mosaic_rra)

writeRaster(mosaic_rra,
"C:/Users/mclel/Documents/dissdata/outputs/rra_classes/rra_classes_mosaic.tif",
overwrite = T)
```

For scatter plots, correlation matrix and other plots

```
# correlation matrix
df <- readRDS("C:/Users/mclel/Documents/dissdata/outputs/relationships/dataframe.Rda")

names(df)

metrics <- c(
    "zmax",
    "zmean",
    "pts_below4m",
    "cc2m",
    "cc6m",
    "cc_sd",
    "zsd",
    "enl1d",
    "enl2d",
    "slope.15",
    "aspect.15",
```

```
    "dtm",
    "twi",
    "tpi500",
    "tpi90",
    "tpi30",
    "exposure_dsm",
    "solar",
    "solar_dsm",
    "taynish_dist",
    "lake_dist",
    "sea_dist",
    "water_dist"
)

# highlight correlated variables in the df
df_subset <- df[, metrics]

# calculate correlation matrix
cor_matrix <- cor(df_subset, use = "pairwise.complete.obs")

# plot
library(corrplot)


png("C:/Users/mclel/Documents/dissdata/outputs/relationships/a_plots/full_cor_mtrx.png",
height = 1500, width = 1500)

corrplot(cor_matrix, method = "color", type = "upper",
         order = "original",
         tl.col = "black", tl.srt = 90, tl.cex = 2,
         addCoef.col = "black",
         number.cex = 1.75,
         cl.cex = 2,
         diag = FALSE,
         na.label = " "
)

dev.off()


# Scatter plot for variables
df <- readRDS("C:/Users/mclel/Documents/dissdata/outputs/relationships/dataframe.Rda")
names(df)

metrics <- c(
  "zmax",
  "zmean",
  "cc2m",
  "cc6m",
```

```
  "cc_sd",
  "zsd",
  "enl1d",
  "enl2d",
  "slope.15",
  "aspect.15",
  "dtm",
  "twi",
  "tpi500",
  "tpi90",
  "tpi30",
  "exposure_dsm",
  "solar",
  "solar_dsm",
  "taynish_dist",
  "lake_dist",
  "sea_dist",
  "water_dist"
)

png("C:/Users/mclel/Documents/dissdata/outputs/relationships/a_plots/scatter/rra/spatial
/water_dist.png", width = 1000, height = 650)

ggplot(df, aes(x = water_dist, y = score)) +
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "gam", se = TRUE, colour = "red") +
  labs(title = "Distance from Water and RRA Score",
       x = "Distance (m)",
       y = "RRA Score") +
  theme_minimal()

dev.off()

"kWh/m²"

cov <- df[df$cc2m >= 70 & df$cc2m <= 88, ]

result <- cor.test(df$score, df$indicators, method = "spearman")

print(result)
print(result$p.value)

names(cov)

cov_sf <- st_as_sf(cov, coords = c("x", "y"), crs = 27700)

st_write(cov_sf, "C:/Users/mclel/Documents/dissdata/outputs/cov70_88.gpkg", append =
FALSE)
```

GAM

```
library(mgcv)
library(Metrics)
library(ggplot2)
library(gratia)
library(patchwork)

# function to make the title text, called later
make_title_text <- function(formula_str) {
  terms <- trimws(unlist(strsplit(formula_str, "\\+")))
  interactions <- terms[grepl(":", terms)]
  main_terms <- terms[!grepl(":", terms)]
  pretty_interactions <- function(int_terms) {
    int_terms <- gsub(":", " and ", int_terms)
    if (length(int_terms) == 1) {
      paste("the interaction between", int_terms)
    } else {
      paste0("the interactions between ", paste(int_terms, collapse = ", "))
    }
  }
  main_text <- paste(main_terms, collapse = ", ")
  if (length(interactions) == 0) {
    paste("Model Accuracy Using", main_text)
  } else {
    paste("Model Accuracy Using", main_text, "and", pretty_interactions(interactions))
  }
}

# leave-one-site-out cross-validation function
loso_gam_cv <- function(
    data,
    site_col = "site_id",
    response_var,
    formula_terms,
    family,
    method = "REML"
) {
  sites <- unique(data[[site_col]])
  rmses <- c()
  maes <- c()
  r2s <- c()
  preds <- numeric(nrow(data))
  data$.row <- seq_len(nrow(data))
  for (test_site in sites) {
    train_data <- subset(data, data[[site_col]] != test_site)
    test_data  <- subset(data, data[[site_col]] == test_site)
    formula_str <- paste(response_var, "~", formula_terms)
    gam_formula <- as.formula(formula_str)
```

```
    model <- gam(gam_formula, data = train_data, family = family, method = method)
    test_preds <- predict(model, newdata = test_data, type = "response")
    preds[test_data$.row] <- test_preds
    fold_rmse <- rmse(test_data[[response_var]], test_preds)
    fold_mae  <- mae(test_data[[response_var]], test_preds)
    fold_r2   <- cor(test_data[[response_var]], test_preds, use = "complete.obs")^2
    rmses <- c(rmses, fold_rmse)
    maes <- c(maes, fold_mae)
    r2s  <- c(r2s, fold_r2)
  }
  overall_rmse <- rmse(data[[response_var]], preds)
  overall_mae  <- mae(data[[response_var]], preds)
  overall_r2   <- cor(data[[response_var]], preds, use = "complete.obs")^2
  list(
    per_site_rmse = rmses,
    mean_site_rmse = mean(rmses),
    overall_rmse = overall_rmse,
    per_site_mae = maes,
    mean_site_mae = mean(maes),
    overall_mae = overall_mae,
    per_site_r2 = r2s,
    mean_site_r2 = mean(r2s, na.rm = TRUE),
    overall_r2 = overall_r2,
    predictions = preds
  )
}

# bring in the data
df <- read.csv("C:/Users/mclel/Documents/dissdata/outputs/relationships/df.csv")

names(df)
df$exposure_dsm <- as.factor(df$exposure_dsm)
df$tree_class <- as.factor(df$tree_class)
tab <- table(df$trees)
rare_types <- names(tab)[tab <= 4]
df$trees_collapsed <- as.character(df$trees)
df$trees_collapsed[df$trees_collapsed %in% rare_types | is.na(df$trees_collapsed)] <-
"other"
df$trees_collapsed <- as.factor(df$trees_collapsed)

# terms used for both quality indices
vars <- paste0("s(taynish_dist) + s(cc2m)")

# formula for richness
gam_formula <- as.formula(paste("indicators ~", vars))

# leave-one-site-out cross-validation for richness
loso_out_richness <- loso_gam_cv(
```

```
  data = df,
  site_col = "site_id",
  response_var = "indicators",
  formula_terms = vars,
  family = poisson(link = "log")
)

# applying the gam model using to the full data set
gam_model <- gam(gam_formula,
                 data = df,
                 method = "REML",
                 family = poisson(link = "log")
)

aic_val <- AIC(gam_model)
fitted_values <- predict(gam_model, newdata = df, type = "response")
observed_values <- df$indicators

vals <- data.frame(
  Richness = observed_values,
  LOSO_Model = loso_out_richness$predictions,
  Training_Model = fitted_values
)
vals$residuals <- vals$Richness - vals$Training_Model
r2_val <- cor(vals$Richness, vals$Training_Model)^2

formula <- paste(deparse(gam_formula[[3]]), collapse = "")
file_suffix <- formula
file_suffix <- gsub("[[:space:]]+", "", file_suffix)
file_suffix <- gsub("[\\(\\),]", "_", file_suffix)
file_suffix <- gsub("__+", "_", file_suffix)
file_suffix <- gsub("_+$", "", file_suffix)

abbreviations <- c(
  "zmean" = "zm",
  "enl2d" = "e2",
  "cc2m" = "cc",
  "cc_sd" = "ccs",
  "pts_below4m" = "b4",
  "aspect.15" = "a",
  "dtm.15" = "d",
  "sd_dtm" = "sd",
  "slope.15" = "s",
  "sd_slope" = "ss",
  "solar_dsm" = "slr_dsm",
  "expos_dsm" = "ex_ds",
  "expos_dtm" = "ex_dt",
  "taynish_dist" = "t_dst",
```

```r
  "sea_dist" = "s_dst",
  "lake_dist" = "l_dst",
  "water_dst" = "w_dst"
)

for (pattern in names(abbreviations)) {
  file_suffix <- gsub(pattern, abbreviations[[pattern]], file_suffix)
}

title_text <- make_title_text(formula)

base_dir <- "C:/Users/mclel/Documents/dissdata/outputs/relationships/a_plots/new_gam/richness/tech_rep"
output_dir <- file.path(base_dir, file_suffix)
dir.create(output_dir, recursive = TRUE, showWarnings = FALSE)

# Plot LOSO predictions vs. observed
plot_path <- file.path(output_dir, paste0(file_suffix, "_LOSO.png"))
png(plot_path, width = 2000, height = 1300)
ggplot(vals, aes(x = LOSO_Model, y = Richness)) +
  geom_point(color = "blue", size = 6) +
  geom_smooth(method = "lm", se = TRUE, colour = "red") +
  labs(title = paste0(title_text, " (LOSO predictions)"),
       x = "LOSO Predicted Richness",
       y = "Observed Richness") +
  annotate("text", x = min(vals$LOSO_Model, na.rm=TRUE), y = max(vals$Richness, na.rm=TRUE),
           label = paste0("LOSO R² = ", round(loso_out_richness$mean_site_r2, 3)),
           hjust = 0, vjust = 1, size = 10, color = "black") +
  theme_minimal(base_size = 30)
dev.off()

pretty_titles <- list(
  "s(sea_dist, k=3)" = "Distance to Sea",
  "s(sea_dist)" = "Distance to Sea",
  "s(cc2m, k = 20)" = "Canopy Cover",
  "s(cc2m)" = "Canopy Cover",
  "s(enl2d)" = "ENL",
  "s(zmax)" = "Maximum Height",
  "s(taynish_dist)" = "Distance to Taynish",
  "s(x, y)" = "Spatial Smooth",
  "s(tpi500)" = "TPI 500",
  "s(lake_dist)" = "Distance to Standing Water"
)

pretty_xlabs <- list(
  "s(sea_dist, k=3)" = "Distance to Sea (m)",
```

```
  "s(sea_dist)" = "Distance to Sea (m)",
  "s(cc2m, k = 20)" = "Canopy Cover (%)",
  "s(cc2m)" = "Canopy Cover (%)",
  "s(enl2d)" = "ENL",
  "s(zmax)" = "Maximum Height (m)",
  "s(taynish_dist)" = "Distance to Taynish (m)",
  "s(x, y)" = "Spatial Position",
  "s(tpi500)" = "TPI 500",
  "s(lake_dist)" = "Distance to Standing Water"
)
resp_name <- "Richness"

all_smooths <- gratia::smooths(gam_model)
for(sm_name in all_smooths) {
  var_title <- pretty_titles[[sm_name]]; if (is.null(var_title)) var_title <- sm_name
  xlab <- pretty_xlabs[[sm_name]]; if (is.null(xlab)) xlab <- sm_name
  resp_name <- "Richness"
  main_title <- paste0(var_title, " and ", resp_name)
  safe_name <- gsub("[\\(\\), ]", "_", sm_name)
  plot_path_i <- file.path(output_dir, paste0("PE_", file_suffix, "_", safe_name,
".png"))
  p <- draw(gam_model, select = sm_name, unconditional = TRUE) +
    labs(
      title = main_title,
      x = xlab,
      y = "Partial Effect"
    ) +
    theme_minimal(base_size = 30) +
    theme(
      plot.title = element_text(size = 36),
      axis.title.x = element_text(size = 28),
      axis.title.y = element_text(size = 28),
      axis.text = element_text(size = 24)
    )
  ggsave(plot_path_i, plot = p, width = 2000/150, height = 1300/150, dpi = 150)
}


write.csv(vals, file.path(output_dir, paste0("residuals_", file_suffix, ".csv")),
row.names = FALSE)

loso_rmse <- loso_out_richness$mean_site_rmse
loso_mae <- loso_out_richness$mean_site_mae
loso_r2 <- loso_out_richness$mean_site_r2

summary_gam <- summary(gam_model)
if(!is.null(summary_gam$p.table)) {
  param_p_tab <- summary_gam$p.table
  param_p_tab <- param_p_tab[rownames(param_p_tab) != "(Intercept)", , drop = FALSE]
```

```
  param_p <- as.list(param_p_tab[,4])
  names(param_p) <- rownames(param_p_tab)
} else {
  param_p <- list()
}
smooth_p <- if(!is.null(summary_gam$s.table)) as.list(summary_gam$s.table[,"p-value"])
else list()
names(smooth_p) <- rownames(summary_gam$s.table)
all_p <- c(param_p, smooth_p)
format_pval <- function(x) {
  if (is.na(x)) return(NA)
  if (x < 0.001) "<0.001" else format(round(x, 3), nsmall = 3)
}
pvals <- paste(names(all_p), sapply(unlist(all_p), format_pval), collapse = "; ")

if (!is.null(summary_gam$s.table)) {
  edf_vals <- as.list(summary_gam$s.table[, "edf"])
  names(edf_vals) <- rownames(summary_gam$s.table)
  edfs <- paste(names(edf_vals), format(unlist(edf_vals), digits = 2), collapse = "; ")
} else {
  edfs <- NA
}

mod_terms <- as.character(gam_formula)[3]

results_row <- data.frame(
  Terms = mod_terms,
  P_values = pvals,
  EDFs = edfs,
  Adj_R2 = round(summary_gam$r.sq, 3),
  Deviance_Explained = round(summary_gam$dev.expl * 100, 1),
  AIC = round(AIC(gam_model), 1),
  CV_RMSE = round(loso_rmse, 2),
  CV_MAE = round(loso_mae, 2),
  CV_R2 = round(loso_r2, 3)
)

csv_path <-
"C:/Users/mclel/Documents/dissdata/outputs/relationships/a_plots/new_gam/richness/tech_r
ep/rich_results_summary.csv"
if (file.exists(csv_path)) {
  write.table(results_row, file = csv_path, sep = ",", row.names = FALSE, col.names =
FALSE, append = TRUE)
} else {
  write.table(results_row, file = csv_path, sep = ",", row.names = FALSE, col.names =
TRUE)
}
print("Model results saved/updated!")
```

```
##################################################################################

# for RRA

# formula for RRA
gam_formula <- as.formula(paste("score ~", vars))

loso_out_rra <- loso_gam_cv(
  data = df,
  site_col = "site_id",
  response_var = "score",
  formula_terms = vars,
  family = gaussian()
)

gam_model <- gam(gam_formula,
                 data = df,
                 method = "REML",
)

aic_val <- AIC(gam_model)
fitted_values <- predict(gam_model, newdata = df, type = "response")
observed_values <- df$score

vals <- data.frame(
  Score = observed_values,
  LOSO_Model = loso_out_rra$predictions,
  Training_Model = fitted_values
)
vals$residuals <- vals$Score - vals$Training_Model
r2_val <- cor(vals$Score, vals$Training_Model)^2

formula <- paste(deparse(gam_formula[[3]]), collapse = "")
file_suffix <- formula
file_suffix <- gsub("[[:space:]]+", "", file_suffix)
file_suffix <- gsub("[\\(\\),]", "_", file_suffix)
file_suffix <- gsub("__+", "_", file_suffix)
file_suffix <- gsub("_+$", "", file_suffix)

for (pattern in names(abbreviations)) {
  file_suffix <- gsub(pattern, abbreviations[[pattern]], file_suffix)
}

title_text <- make_title_text(formula)

base_dir <-
"C:/Users/mclel/Documents/dissdata/outputs/relationships/a_plots/new_gam/rra/tech_rep"
```

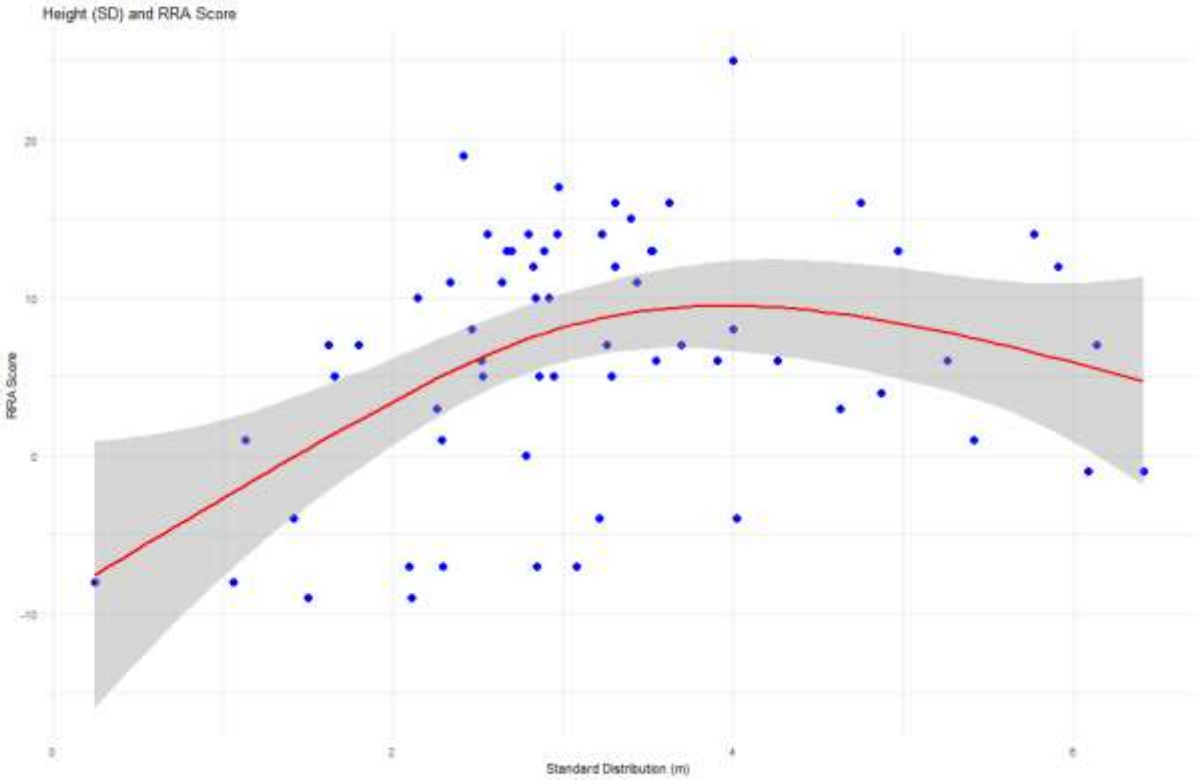
```
output_dir <- file.path(base_dir, file_suffix)
dir.create(output_dir, recursive = TRUE, showWarnings = FALSE)

# Plot LOSO predictions vs. observed
plot_path <- file.path(output_dir, paste0(file_suffix, "_LOSO.png"))
png(plot_path, width = 2000, height = 1300)
ggplot(vals, aes(x = LOSO_Model, y = Score)) +
  geom_point(color = "blue", size = 6) +
  geom_smooth(method = "lm", se = TRUE, colour = "red") +
  labs(title = paste0(title_text, " (LOSO predictions)"),
       x = "LOSO Predicted RRA Score",
       y = "Observed RRA Score") +
  annotate("text", x = min(vals$LOSO_Model, na.rm=TRUE), y = max(vals$Score,
na.rm=TRUE),
           label = paste0("LOSO R² = ", round(loso_out_rra$mean_site_r2, 3)),
           hjust = 0, vjust = 1, size = 10, color = "black") +
  theme_minimal(base_size = 30)
dev.off()

pretty_titles <- list(
  "s(sea_dist, k=3)" = "Distance to Sea",
  "s(sea_dist)" = "Distance to Sea",
  "s(cc2m, k = 20)" = "Canopy Cover",
  "s(cc2m)" = "Canopy Cover",
  "s(enl2d)" = "ENL",
  "s(zmax)" = "Maximum Height",
  "s(taynish_dist)" = "Distance to Taynish",
  "s(x, y)" = "Spatial Smooth",
  "s(tpi500)" = "TPI 500",
  "s(lake_dist)" = "Distance to Standing Water"
)

pretty_xlabs <- list(
  "s(sea_dist, k=3)" = "Distance to Sea (m)",
  "s(sea_dist)" = "Distance to Sea (m)",
  "s(cc2m, k = 20)" = "Canopy Cover (%)",
  "s(cc2m)" = "Canopy Cover (%)",
  "s(enl2d)" = "ENL",
  "s(zmax)" = "Maximum Height (m)",
  "s(taynish_dist)" = "Distance to Taynish (m)",
  "s(x, y)" = "Spatial Position",
  "s(tpi500)" = "TPI 500",
  "s(lake_dist)" = "Distance to Standing Water"
)
resp_name <- "RRA Score"

all_smooths <- gratia::smooths(gam_model)
for(sm_name in all_smooths) {
```

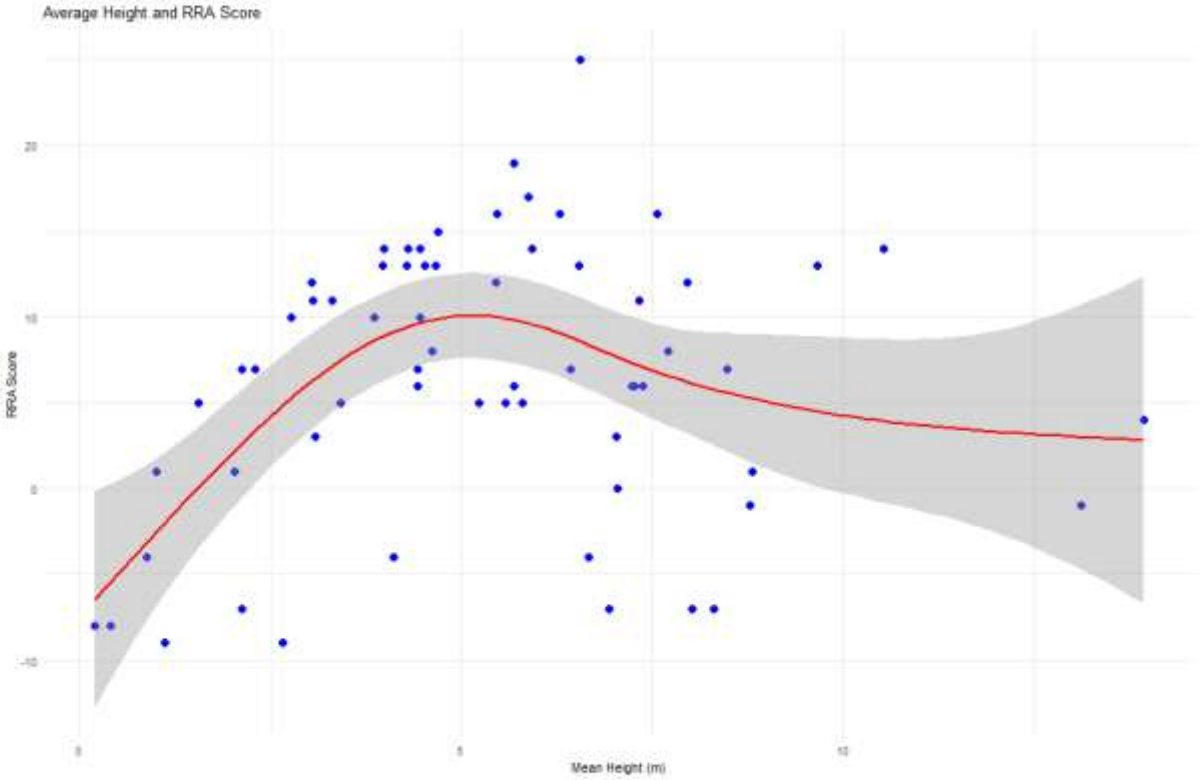
```
  var_title <- pretty_titles[[sm_name]]; if (is.null(var_title)) var_title <- sm_name
  xlab <- pretty_xlabs[[sm_name]]; if (is.null(xlab)) xlab <- sm_name
  resp_name <- "RRA Score"
  main_title <- paste0(var_title, " and ", resp_name)
  safe_name <- gsub("[\\(\\), ]", "_", sm_name)
  plot_path_i <- file.path(output_dir, paste0("PE_", file_suffix, "_", safe_name,
".png"))
  p <- draw(gam_model, select = sm_name, unconditional = TRUE) +
    labs(
      title = main_title,
      x = xlab,
      y = "Partial Effect"
    ) +
    theme_minimal(base_size = 30) +
    theme(
      plot.title = element_text(size = 36),
      axis.title.x = element_text(size = 28),
      axis.title.y = element_text(size = 28),
      axis.text = element_text(size = 24)
    )
  ggsave(plot_path_i, plot = p, width = 2000/150, height = 1300/150, dpi = 150)
}

write.csv(vals, file.path(output_dir, paste0("residuals_", file_suffix, ".csv")),
row.names = FALSE)

loso_rmse <- loso_out_rra$mean_site_rmse
loso_mae <- loso_out_rra$mean_site_mae
loso_r2 <- loso_out_rra$mean_site_r2

summary_gam <- summary(gam_model)
if(!is.null(summary_gam$p.table)) {
  param_p_tab <- summary_gam$p.table
  param_p_tab <- param_p_tab[rownames(param_p_tab) != "(Intercept)", , drop = FALSE]
  param_p <- as.list(param_p_tab[,4])
  names(param_p) <- rownames(param_p_tab)
} else {
  param_p <- list()
}
smooth_p <- if(!is.null(summary_gam$s.table)) as.list(summary_gam$s.table[,"p-value"])
else list()
names(smooth_p) <- rownames(summary_gam$s.table)
all_p <- c(param_p, smooth_p)
format_pval <- function(x) {
  if (is.na(x)) return(NA)
  if (x < 0.001) "<0.001" else format(round(x, 3), nsmall = 3)
}
pvals <- paste(names(all_p), sapply(unlist(all_p), format_pval), collapse = "; ")
```

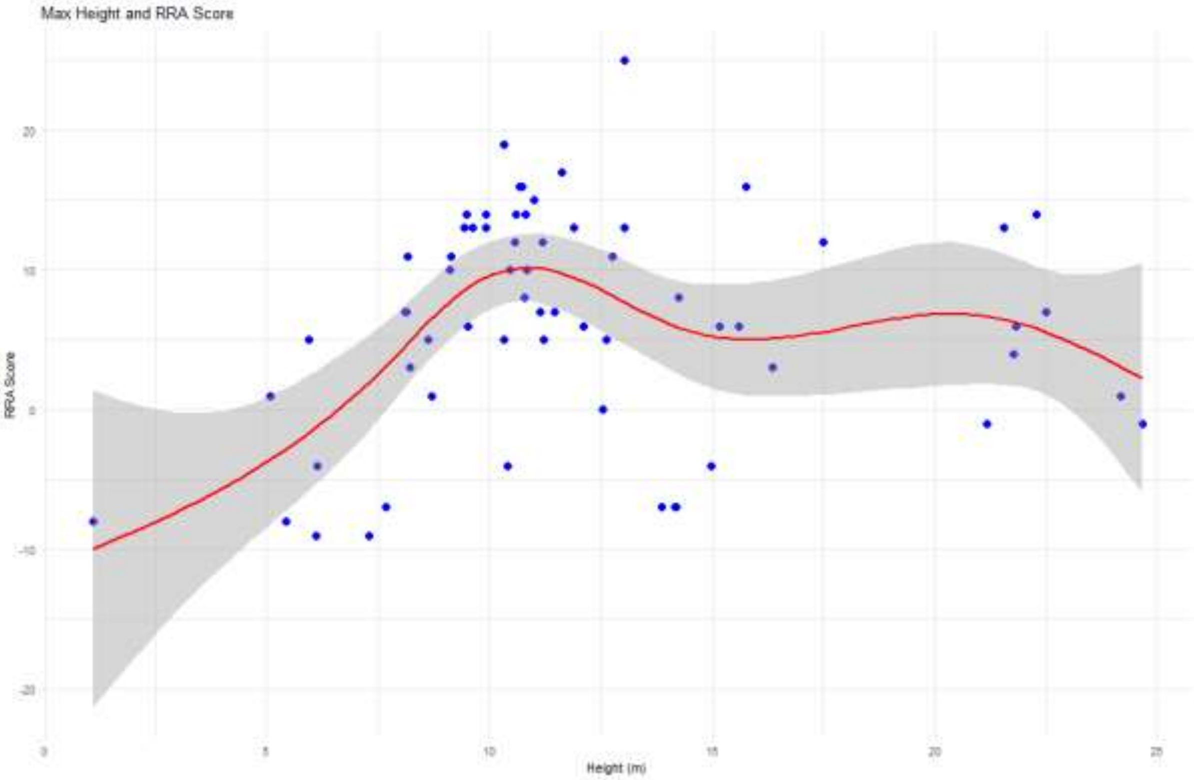
```
if (!is.null(summary_gam$s.table)) {
  edf_vals <- as.list(summary_gam$s.table[, "edf"])
  names(edf_vals) <- rownames(summary_gam$s.table)
  edfs <- paste(names(edf_vals), format(unlist(edf_vals), digits = 2), collapse = "; ")
} else {
  edfs <- NA
}

mod_terms <- as.character(gam_formula)[3]

results_row <- data.frame(
  Terms = mod_terms,
  P_values = pvals,
  EDFs = edfs,
  Adj_R2 = round(summary_gam$r.sq, 3),
  Deviance_Explained = round(summary_gam$dev.expl * 100, 1),
  AIC = round(AIC(gam_model), 1),
  CV_RMSE = round(loso_rmse, 2),
  CV_MAE = round(loso_mae, 2),
  CV_R2 = round(loso_r2, 3)
)

csv_path <- "C:/Users/mclel/Documents/dissdata/outputs/relationships/a_plots/new_gam/rra/tech_rep/rra_results_summary.csv"
if (file.exists(csv_path)) {
  write.table(results_row, file = csv_path, sep = ",", row.names = FALSE, col.names = FALSE, append = TRUE)
} else {
  write.table(results_row, file = csv_path, sep = ",", row.names = FALSE, col.names = TRUE)
}
print("Model results saved/updated!")
```

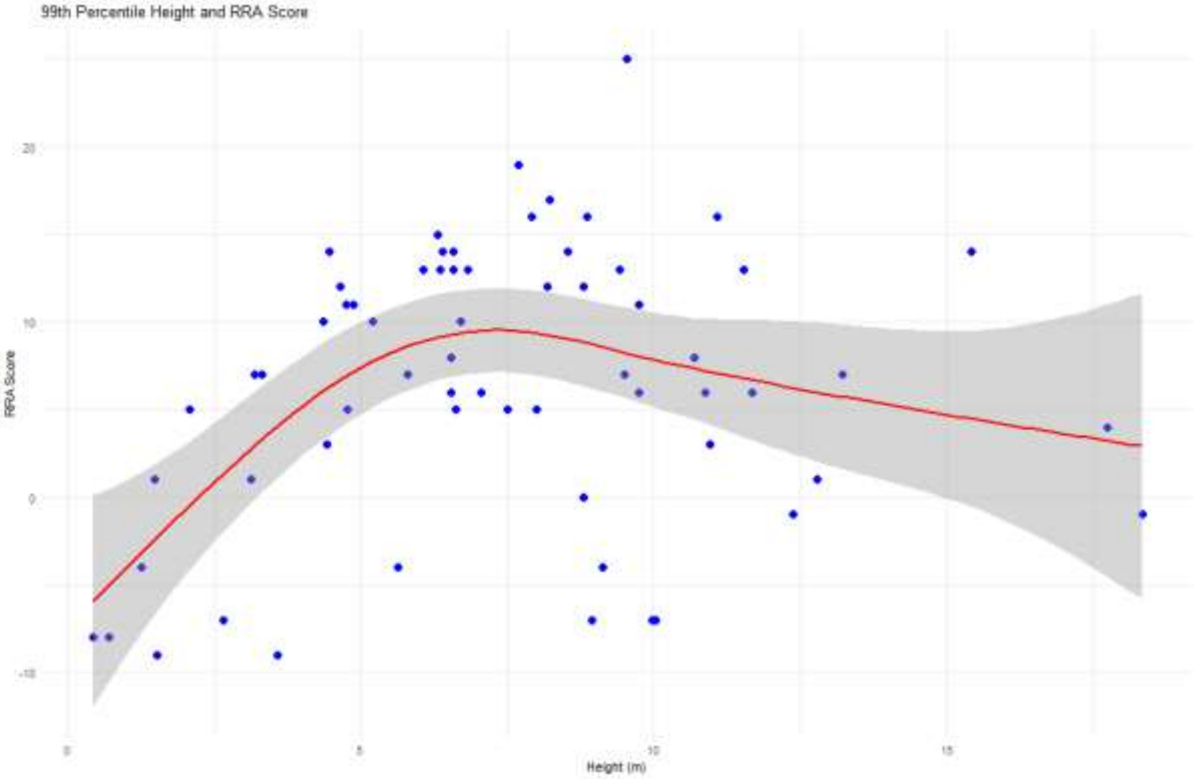
B – Scatter Plots

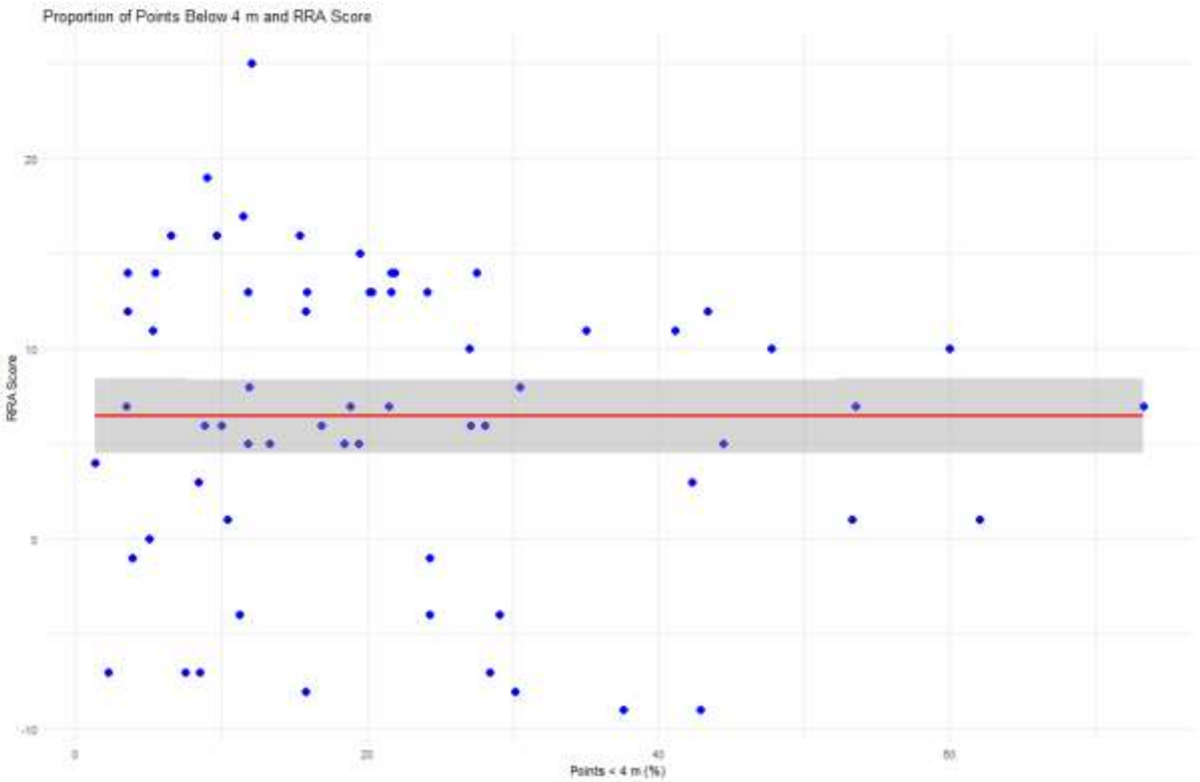
Scatter plots show the influence of each metric before accounting for tree type.



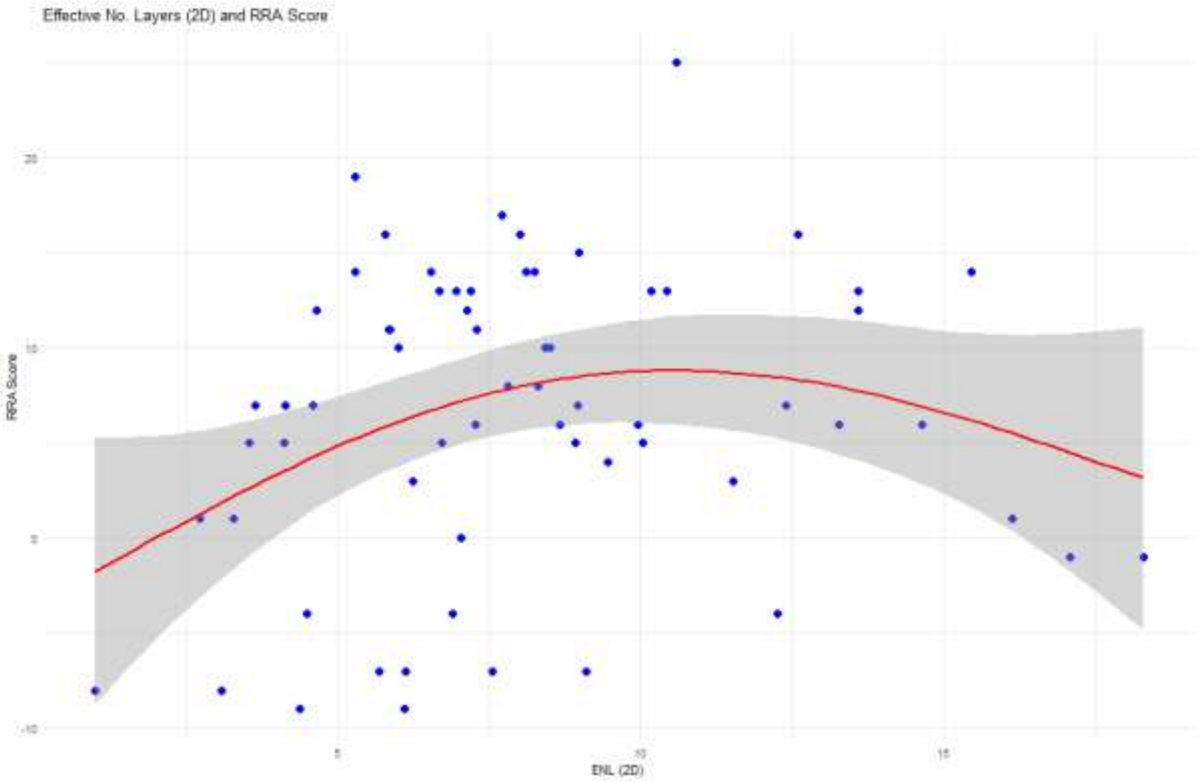



Max Height and RRA Score
RRA Score
20
10
0
-10
-20
0
5
10
15
20
25
Height (m)


99th Percentile Height and RRA Score
RRA Score
20
10
0
-10
0
5
10
15
Height (m)

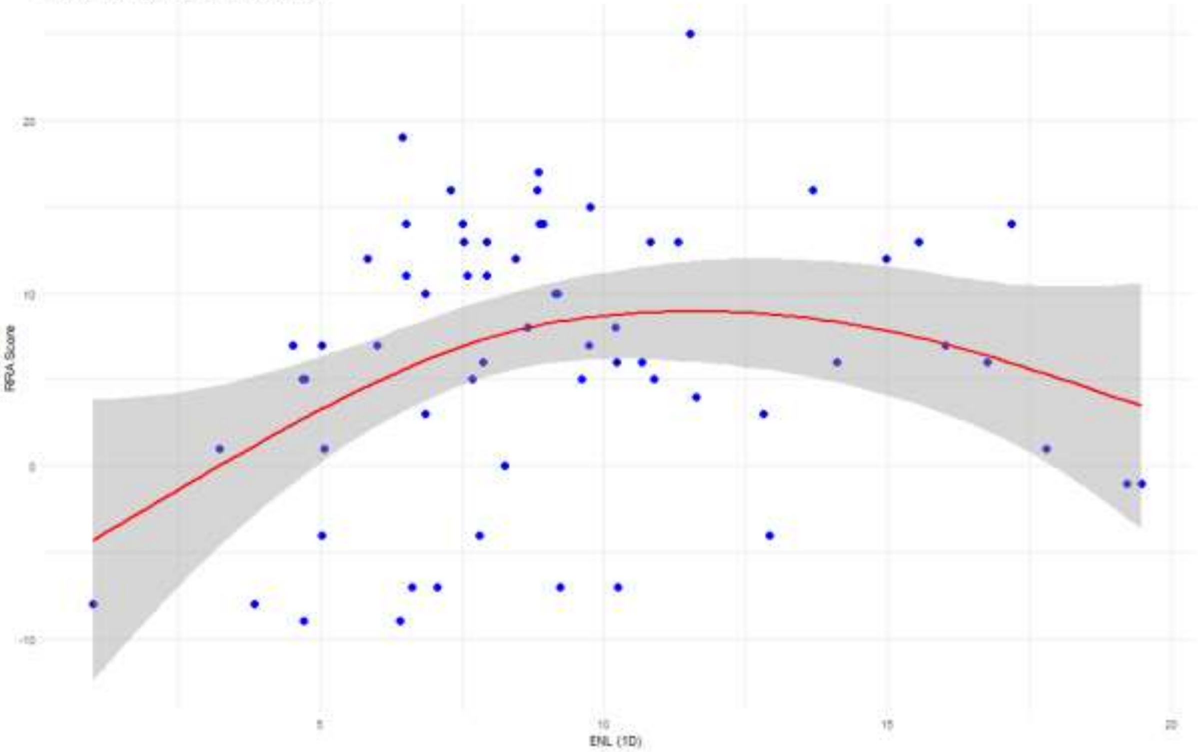


Proportion of Points Below 4 m and RRA Score

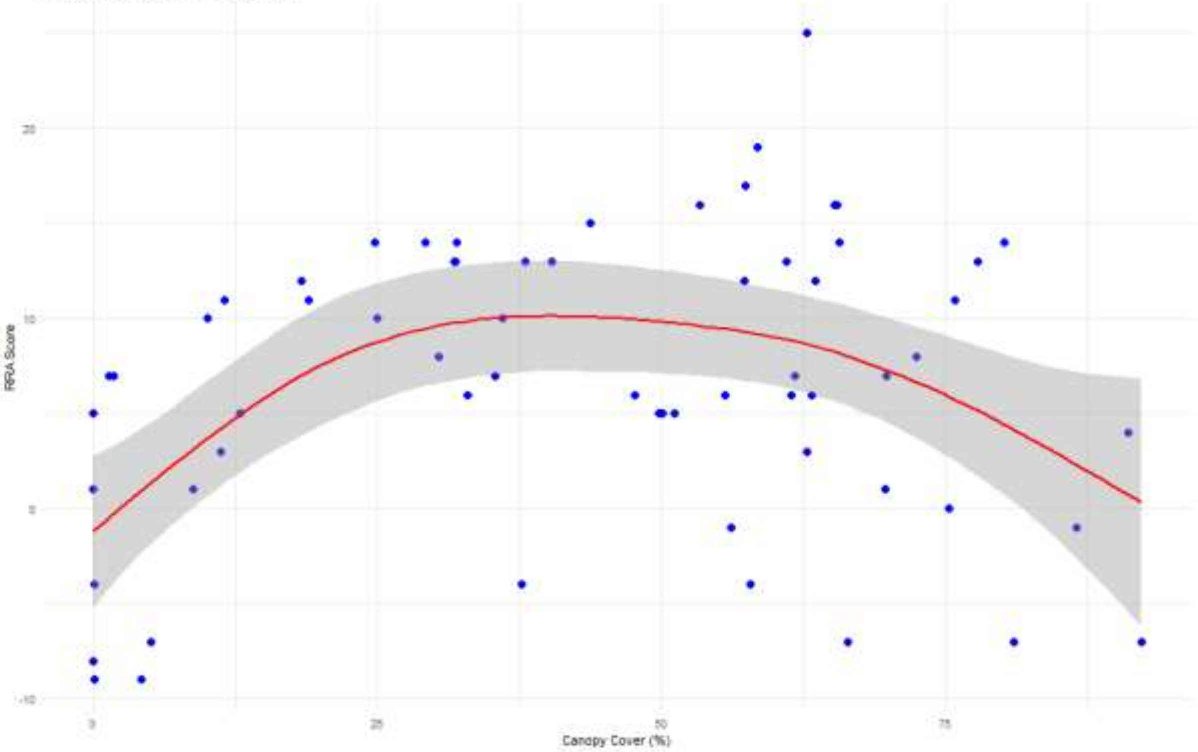


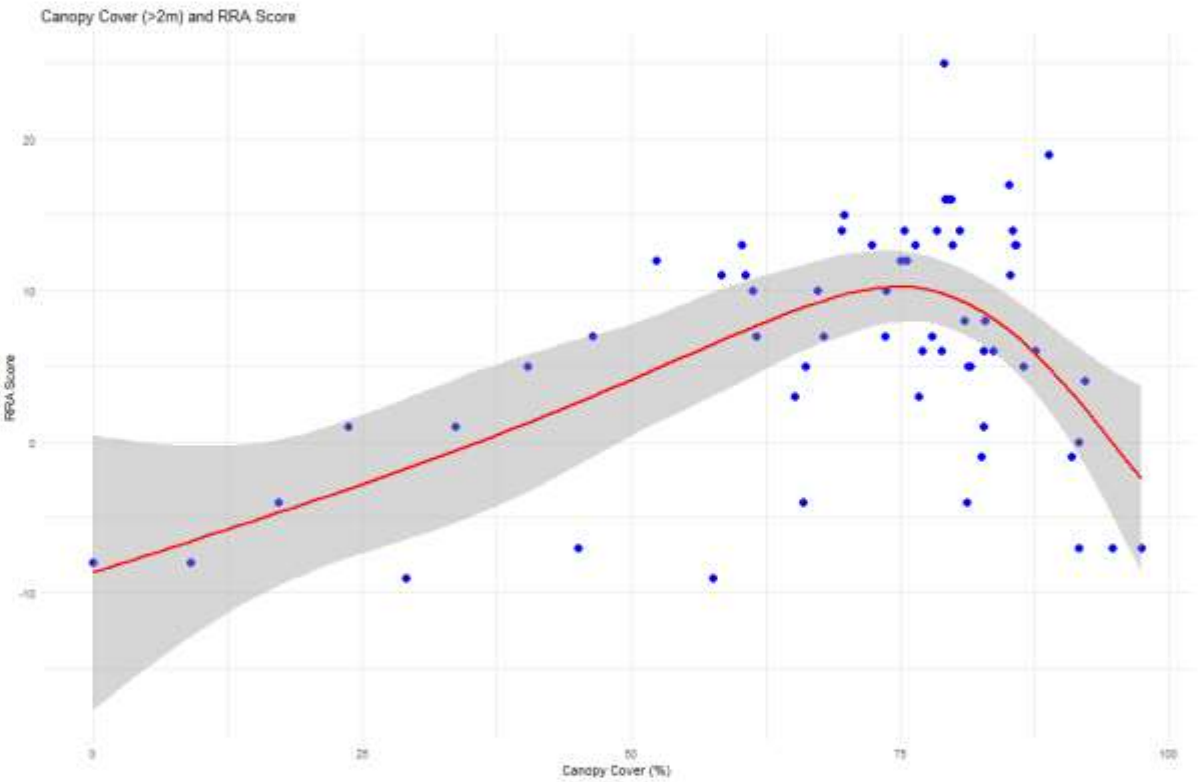
Effective No. Layers (2D) and RRA Score

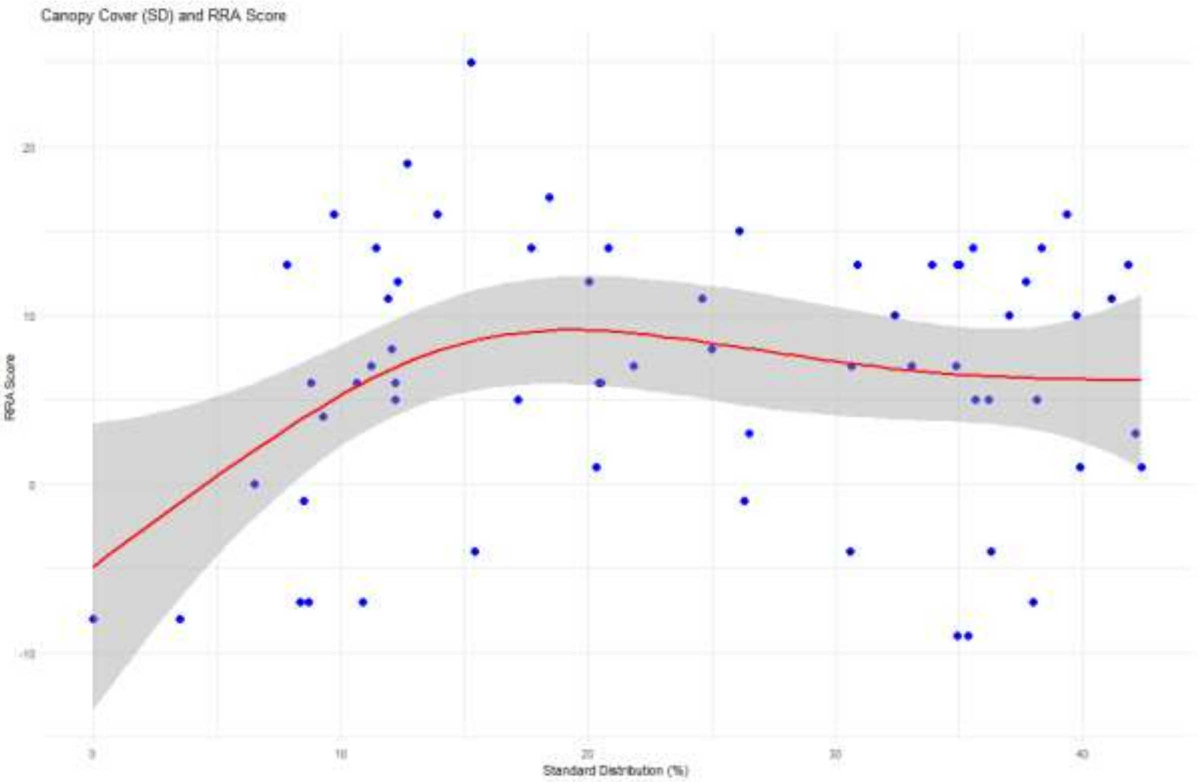
Effective No. Layers (1D) and RRA Score

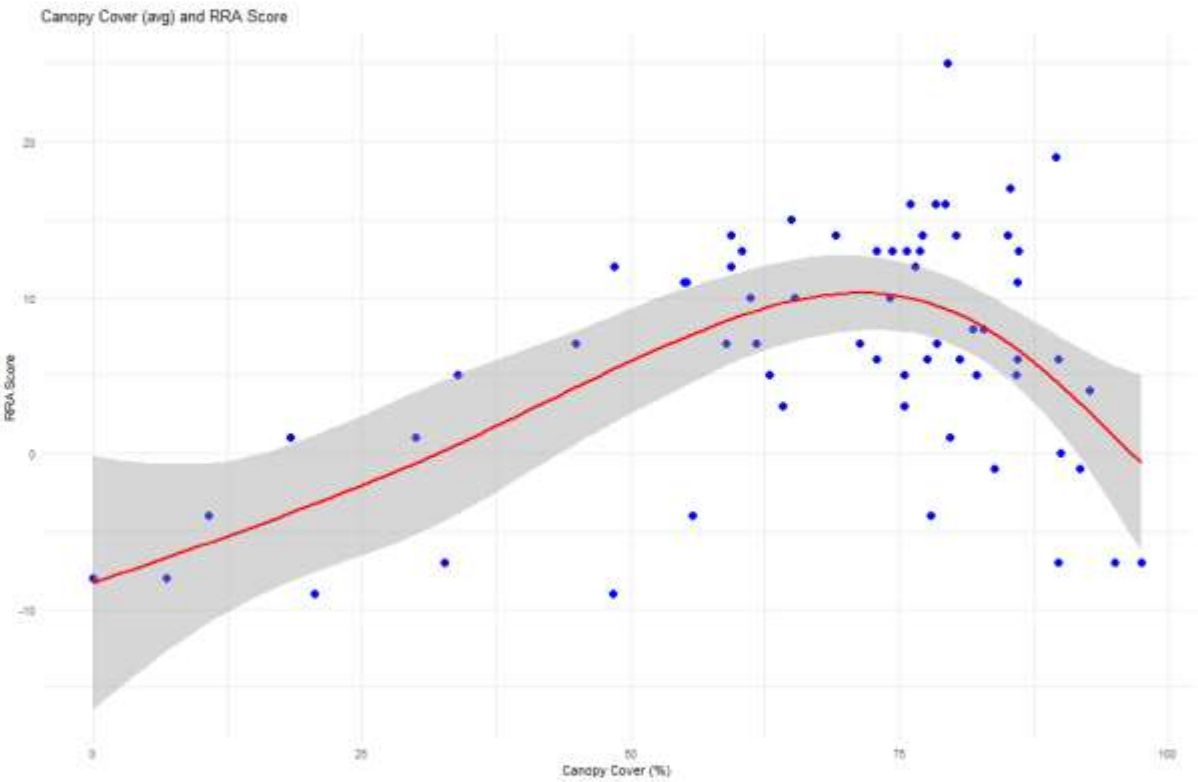


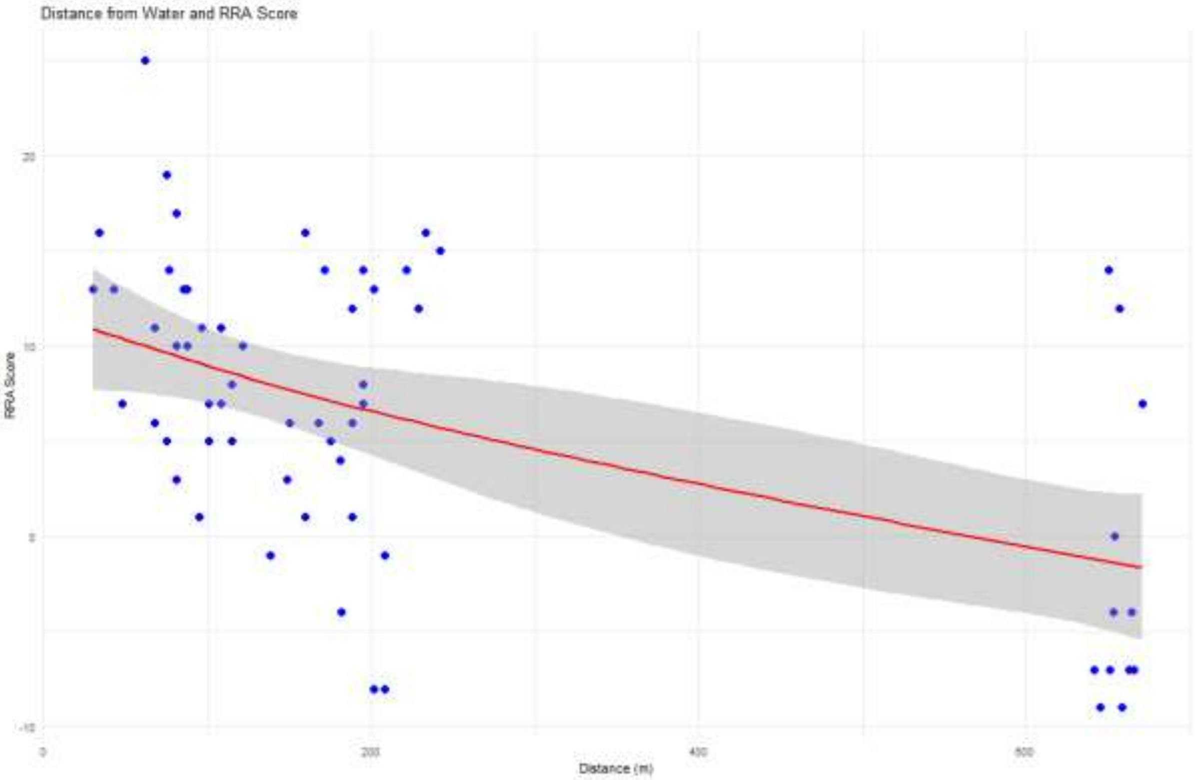
Canopy Cover (>6m) and RRA Score

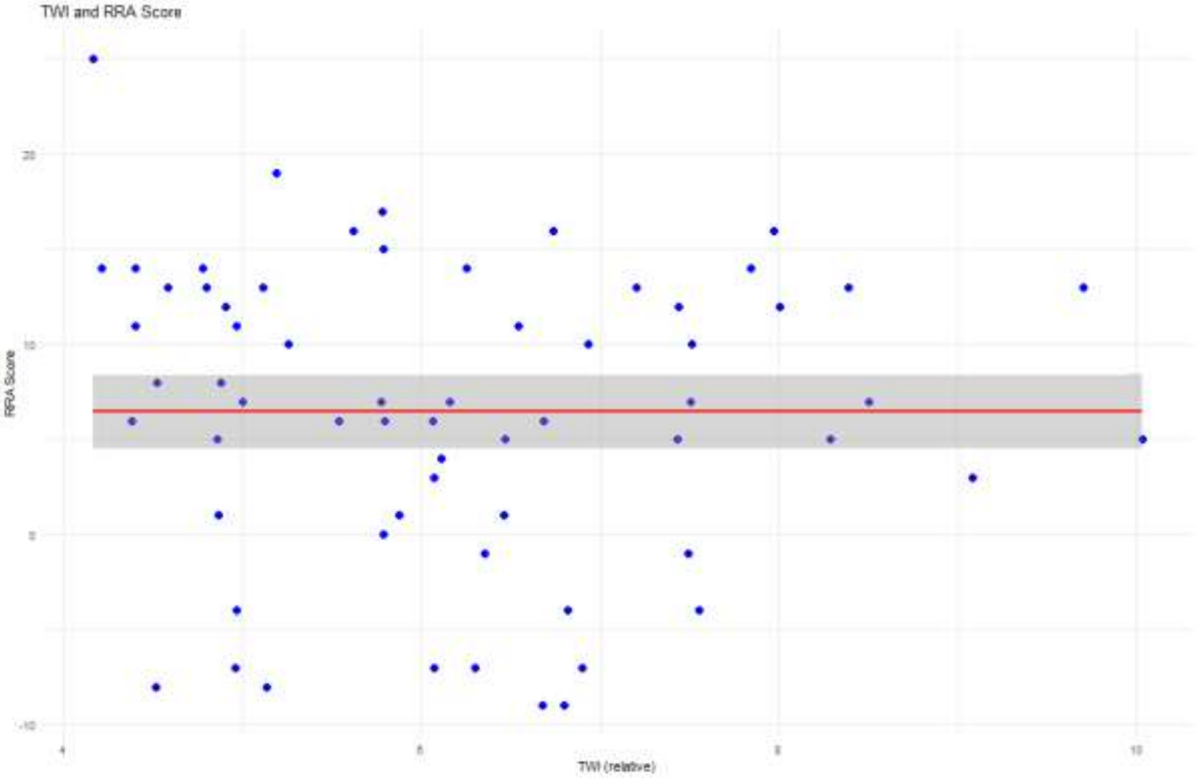



Canopy Cover (>2m) and RRA Score
RRA Score
20
10
0
-10
0
25
50
75
100
Canopy Cover (%)

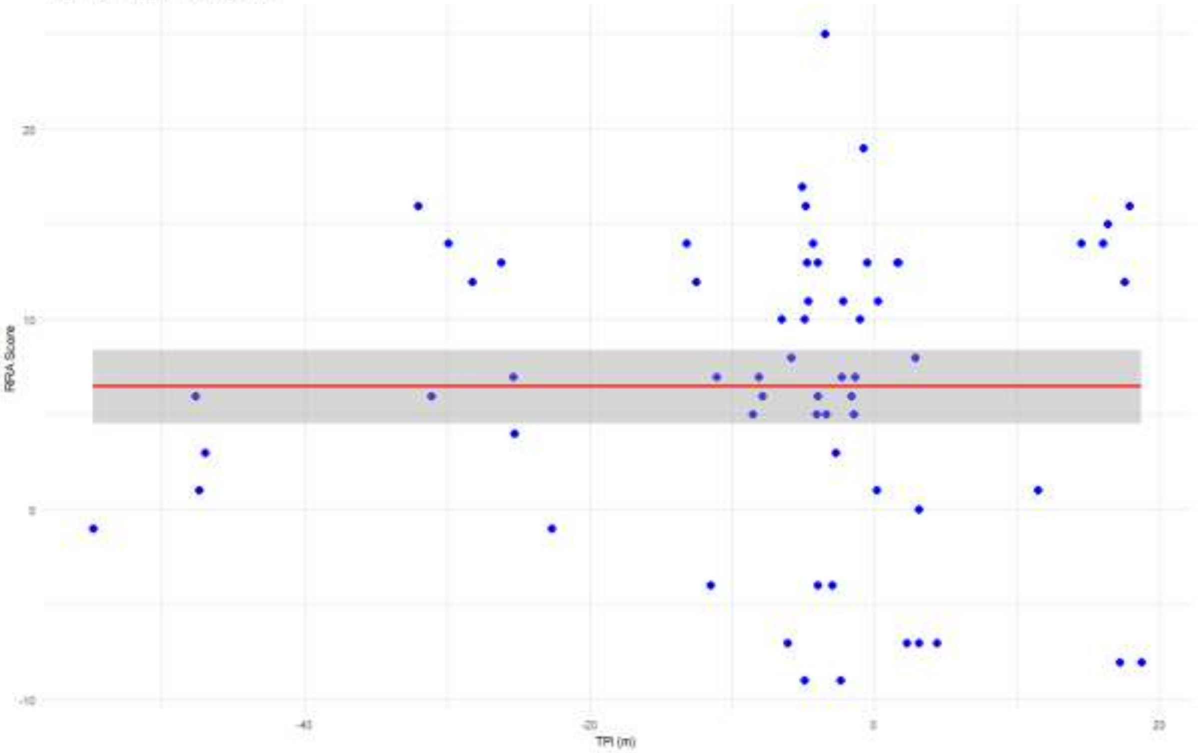

Canopy Cover (SD) and RRA Score
RRA Score
20
10
0
-10
0
10
20
30
40
Standard Distribution (%)


Canopy Cover (avg) and RRA Score
RRA Score
20
10
0
-10
0
25
50
75
100
Canopy Cover (%)

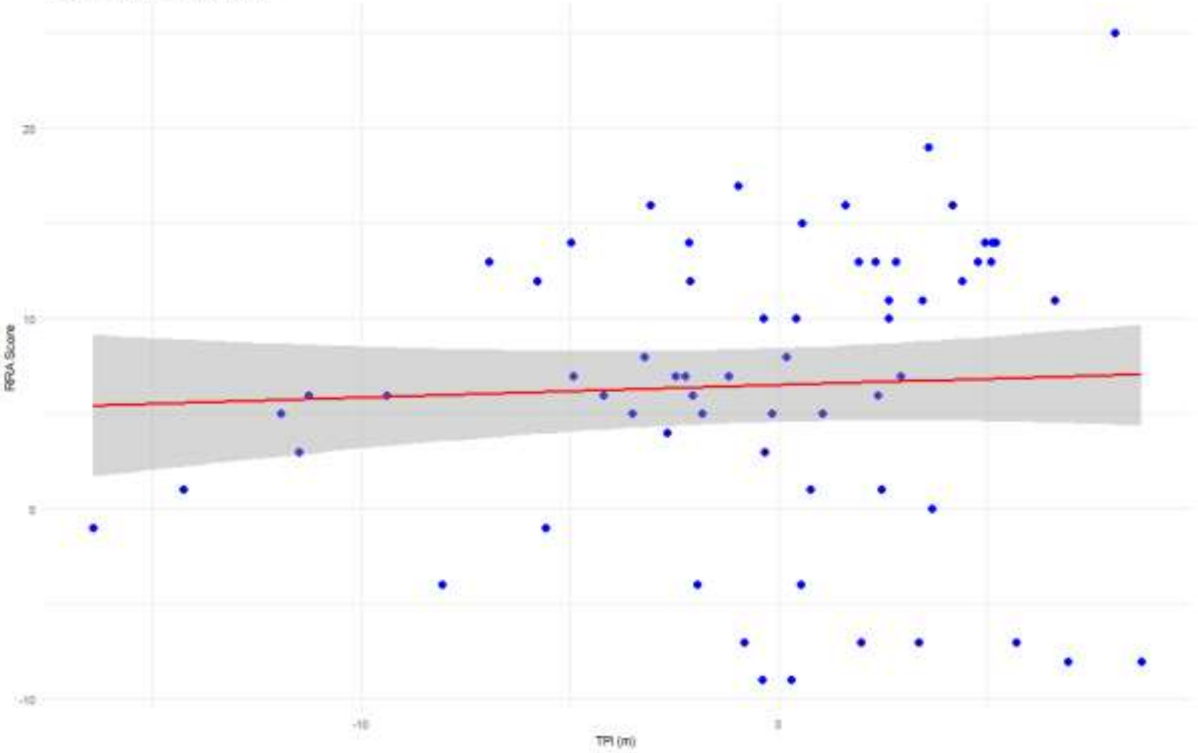

Distance from Water and RRA Score
RRA Score
20
10
0
-10
0
200
400
600
Distance (m)


TWI and RRA Score
RRA Score
20
10
0
-10
4
6
8
10
TWI (relative)

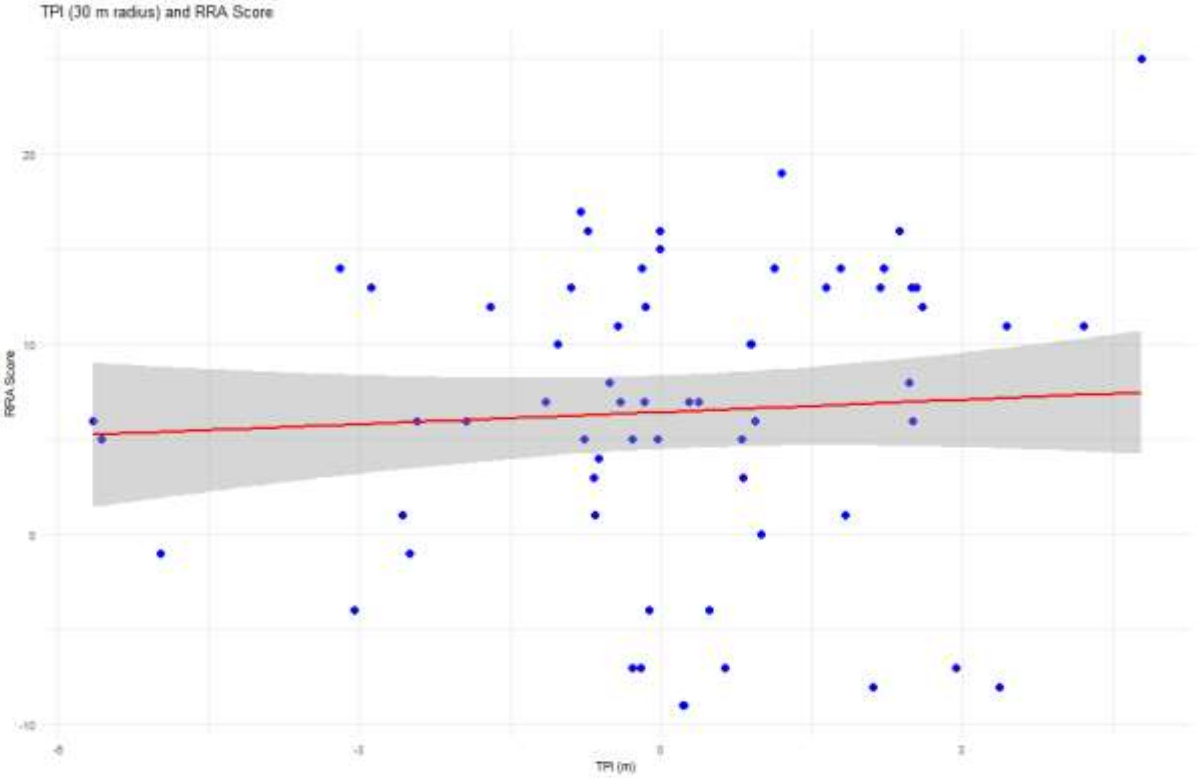
TPI (500 m radius) and RRA Score



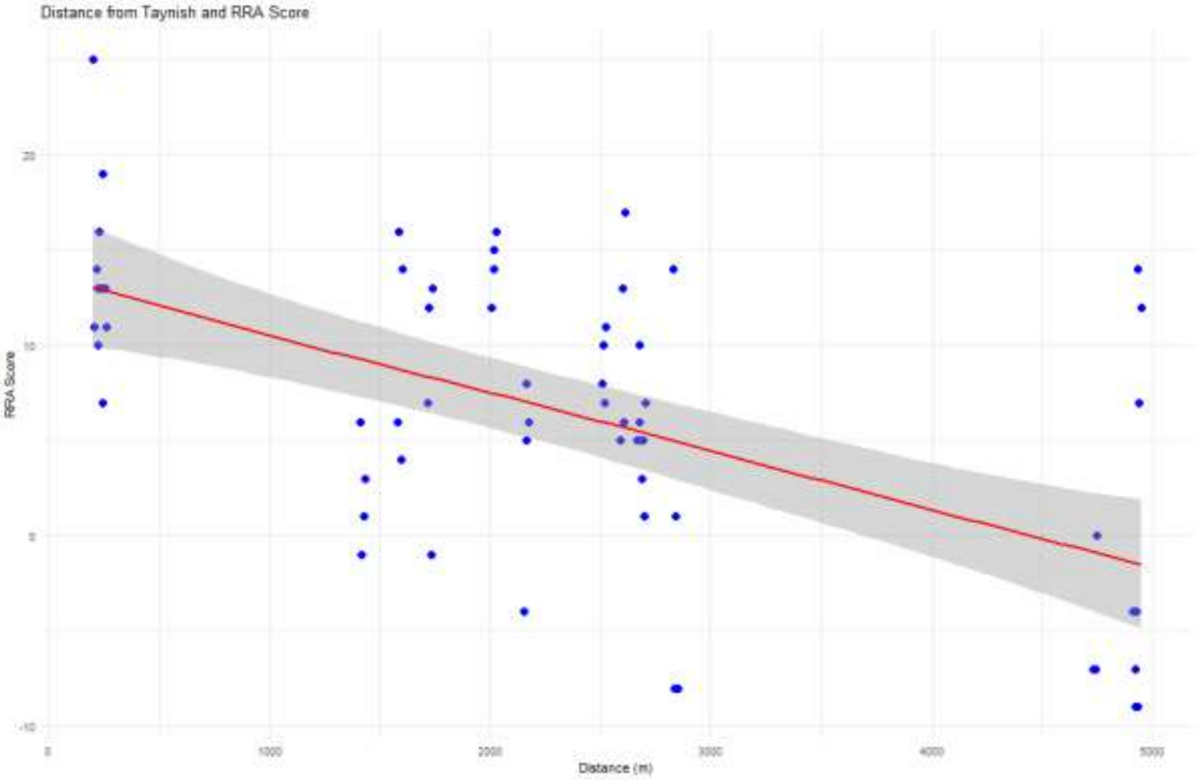
TPI (90 m radius) and RRA Score

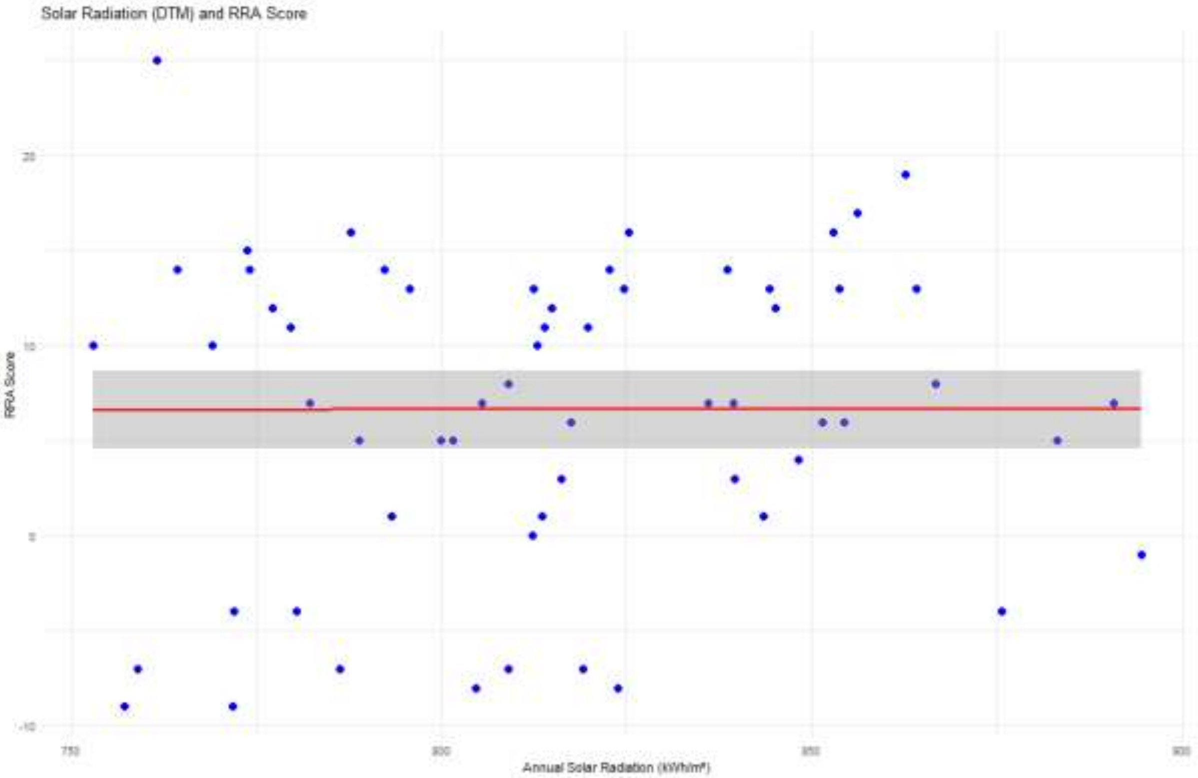


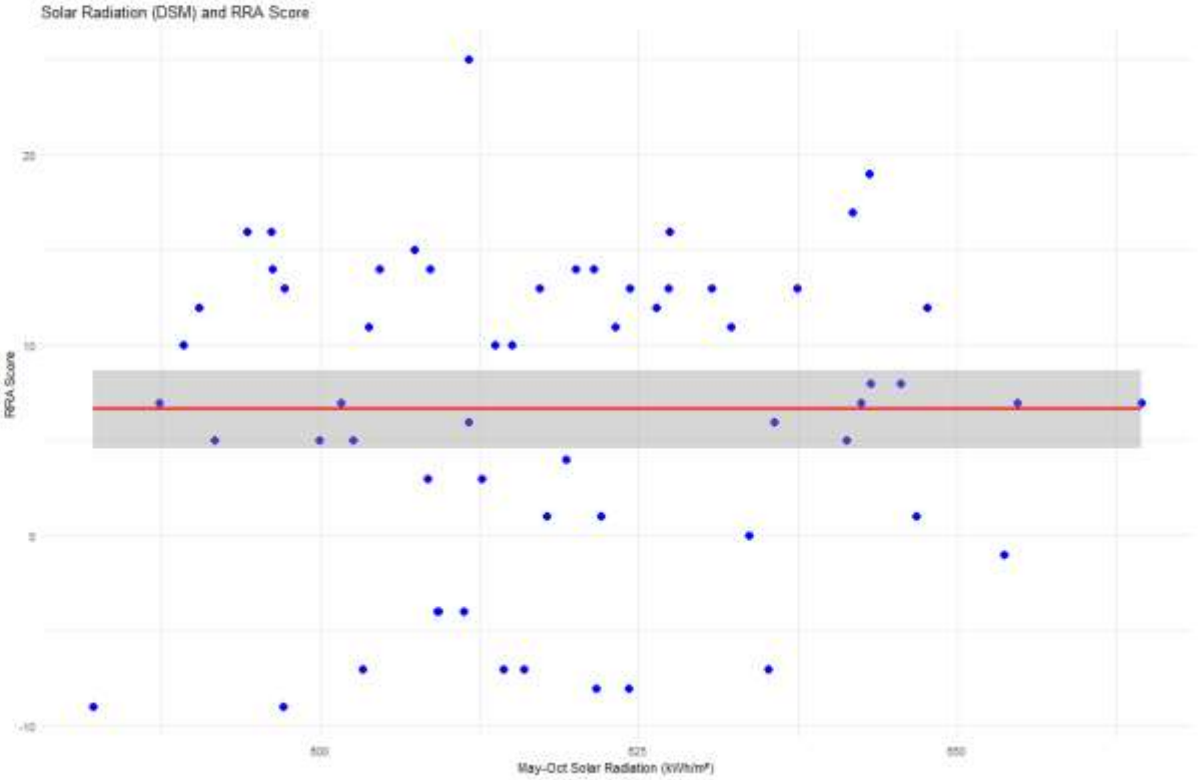
TPI (30 m radius) and RRA Score

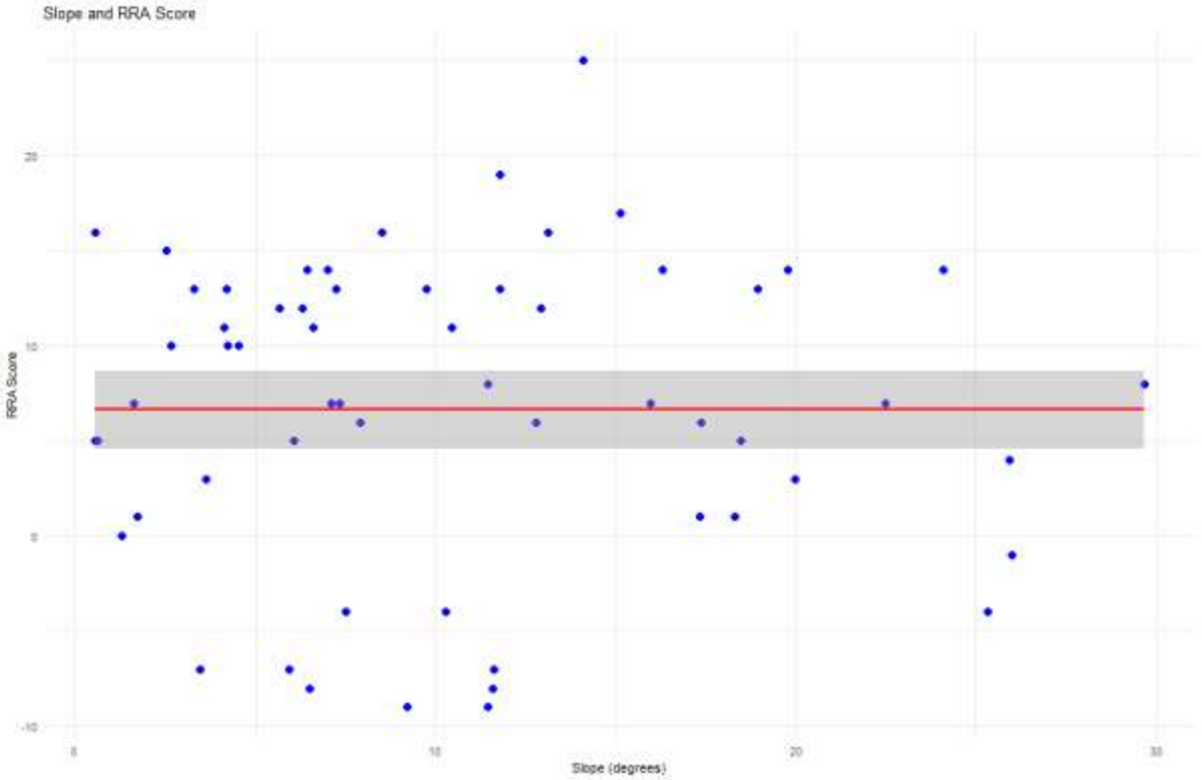


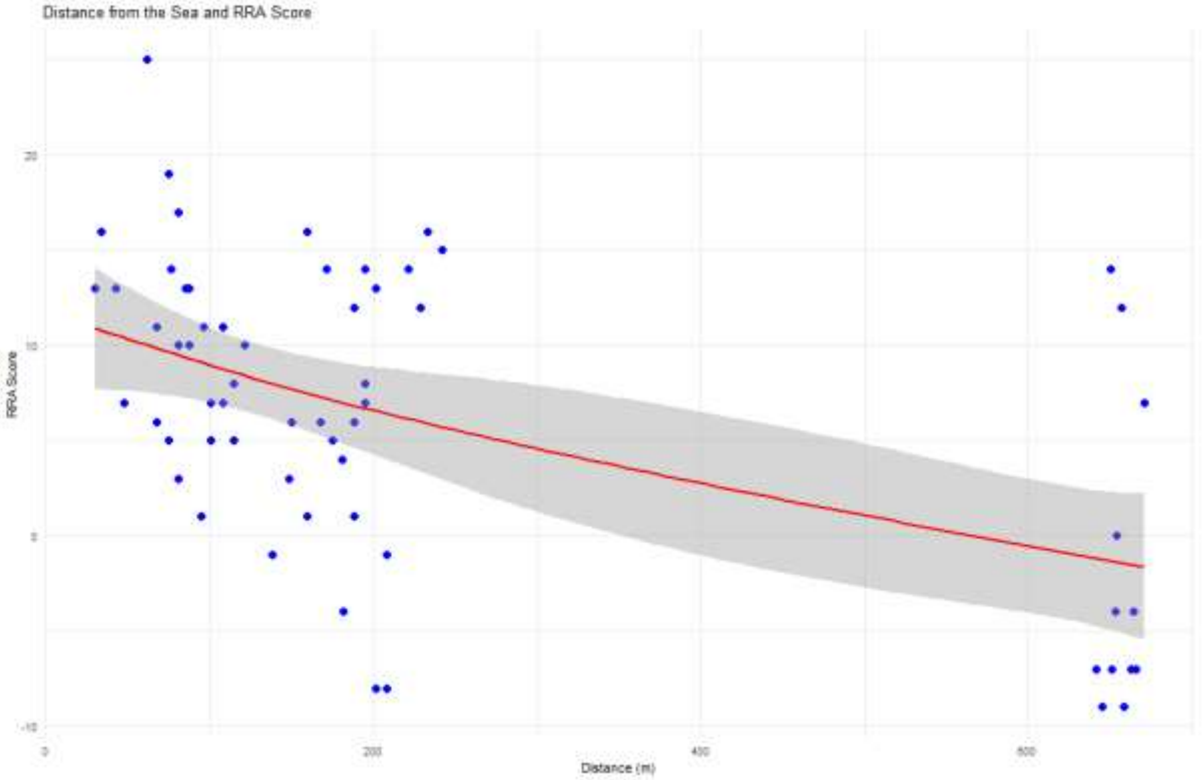
Distance from Taynish and RRA Score

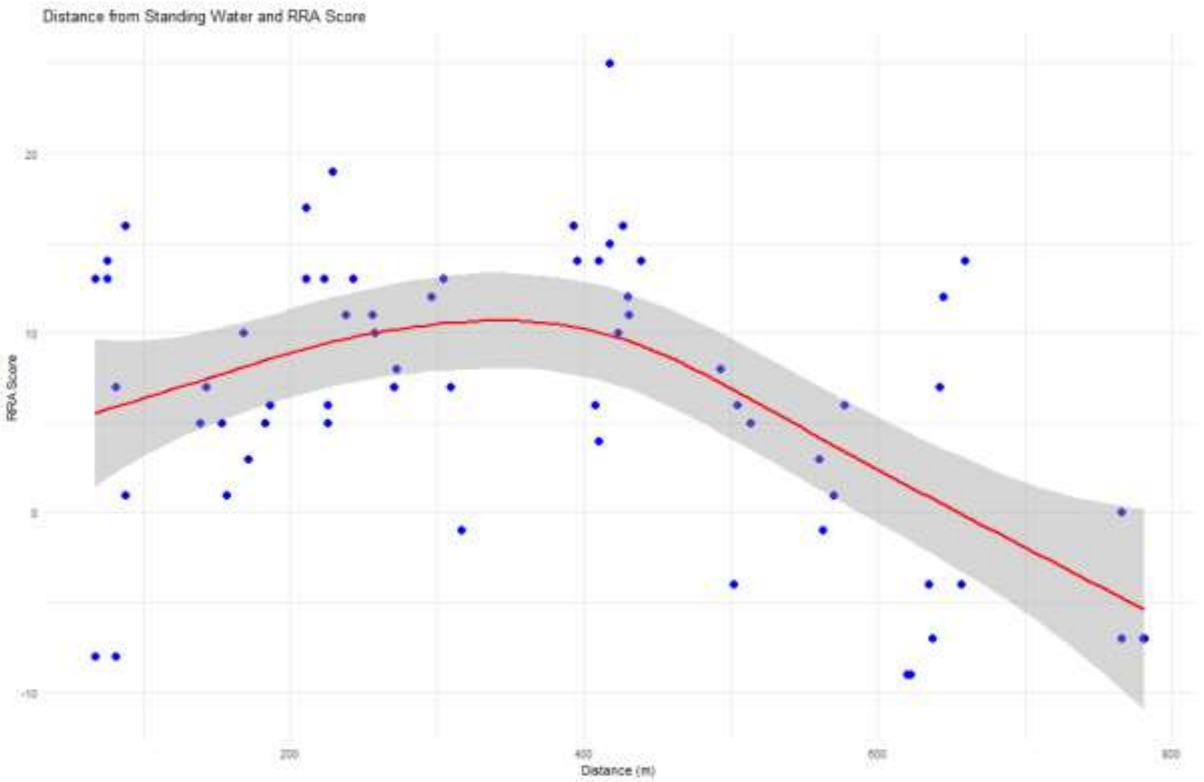


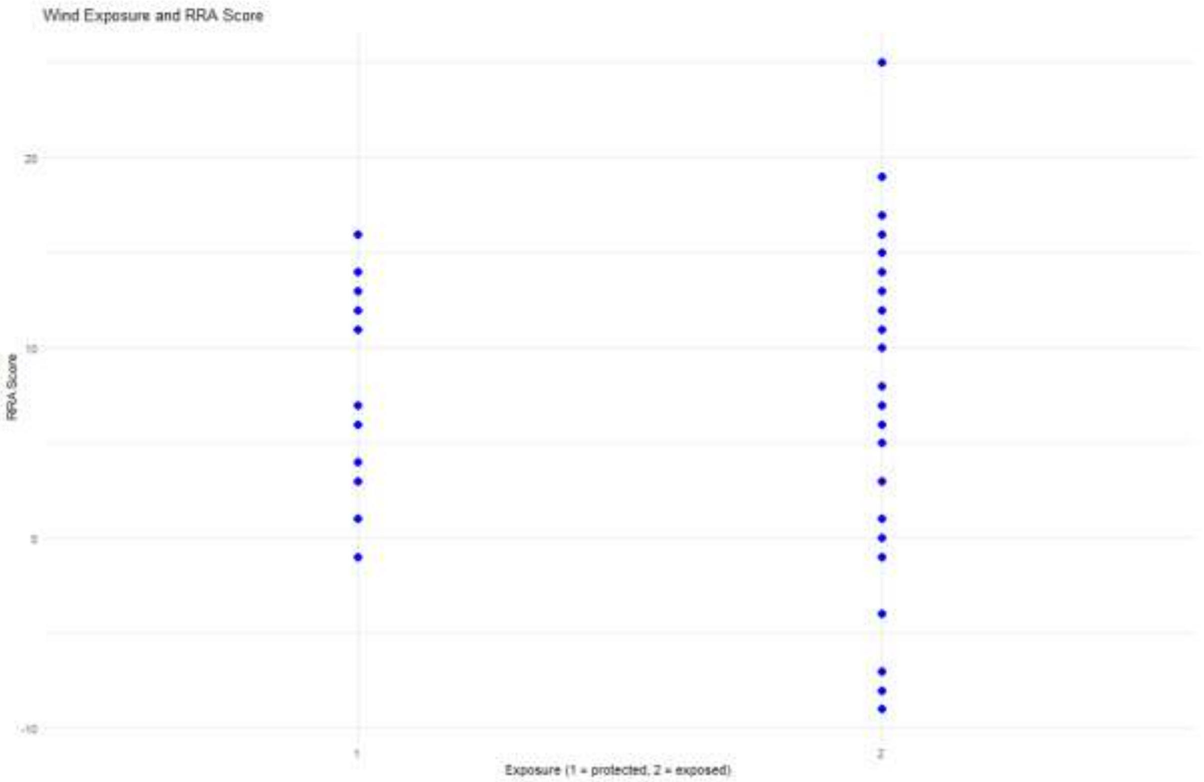

Solar Radiation (DTM) and RRA Score

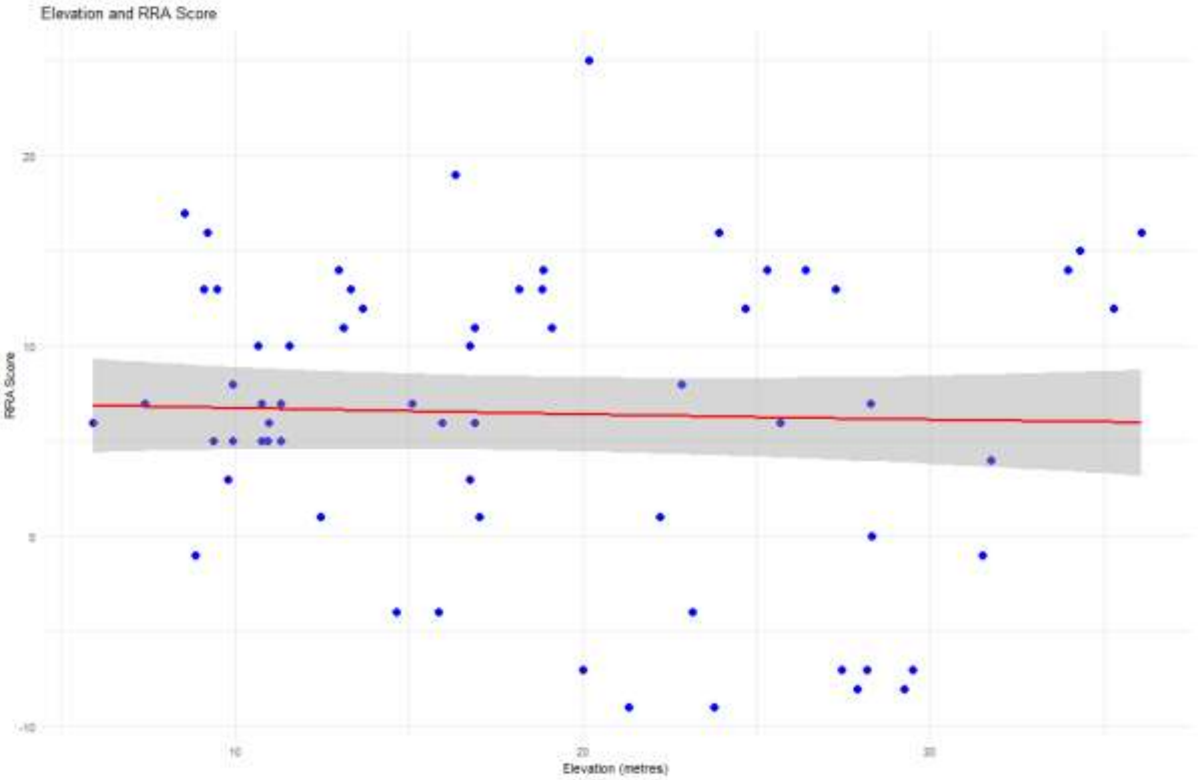

Solar Radiation (DSM) and RRA Score

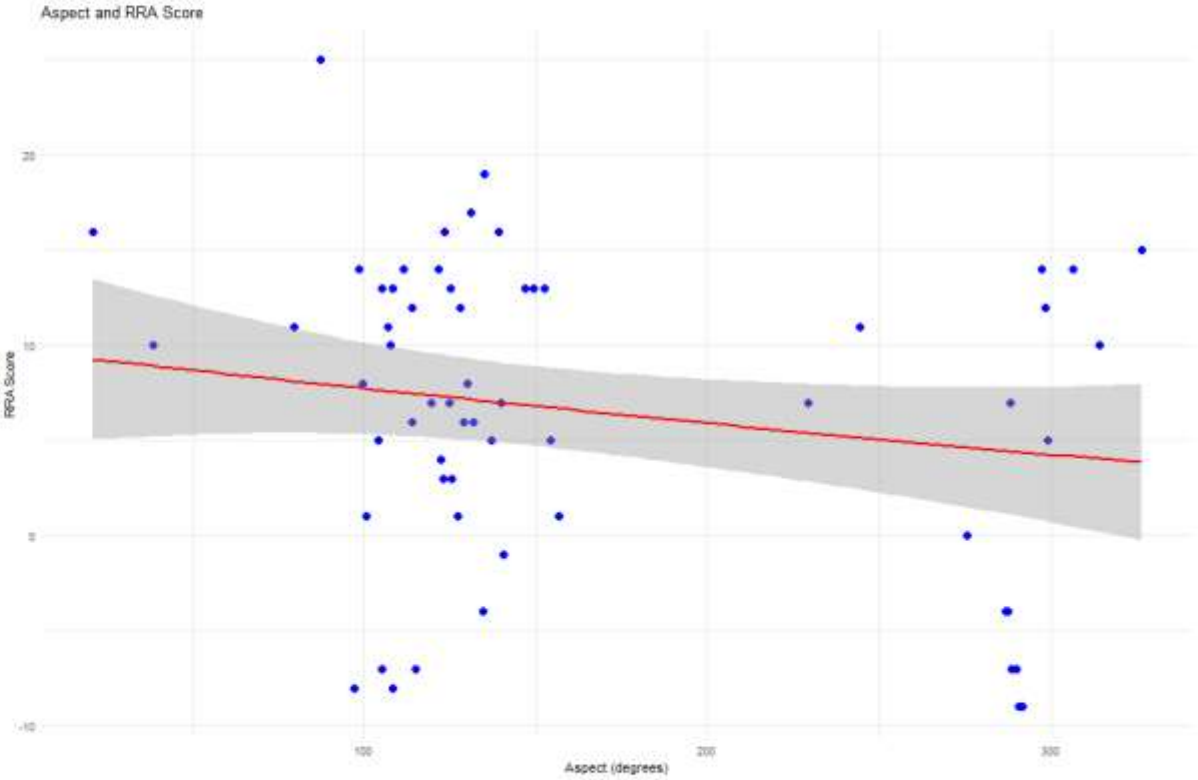

Slope and RRA Score
RRA Score
Slope (degrees)
-10
0
10
20
0
10
20
30

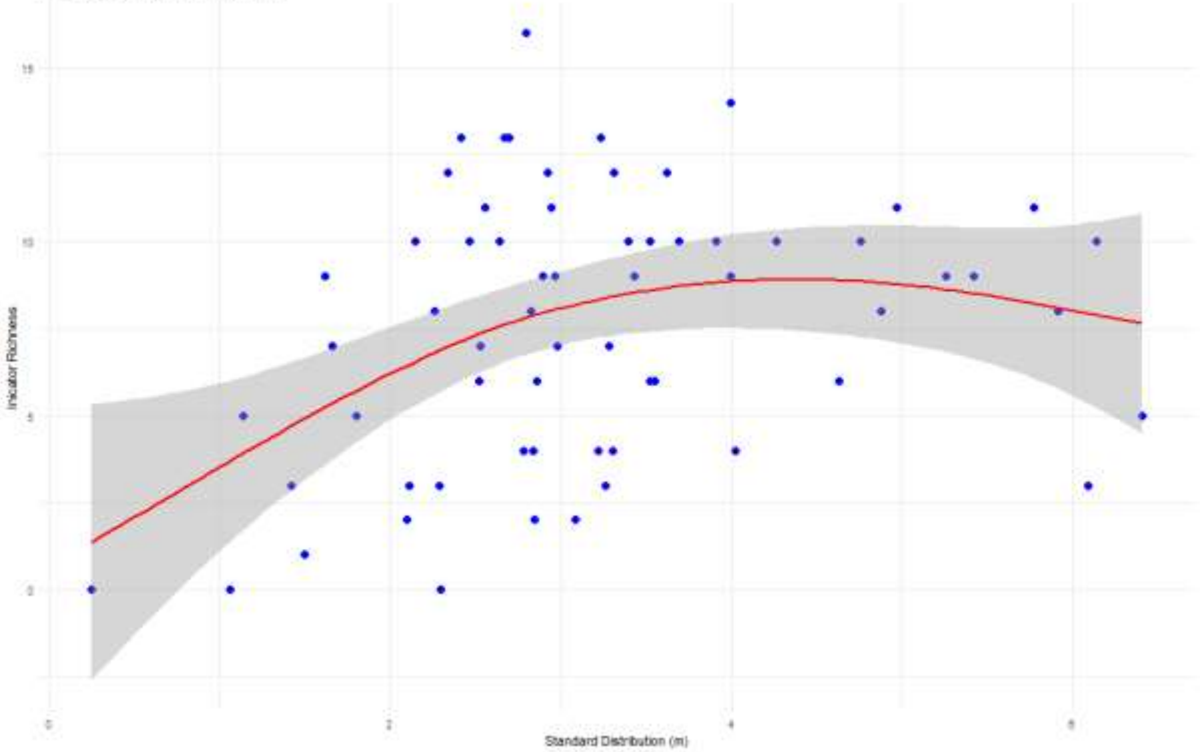

Distance from the Sea and RRA Score
RRA Score
Distance (m)
-10
0
10
20
0
200
400
600

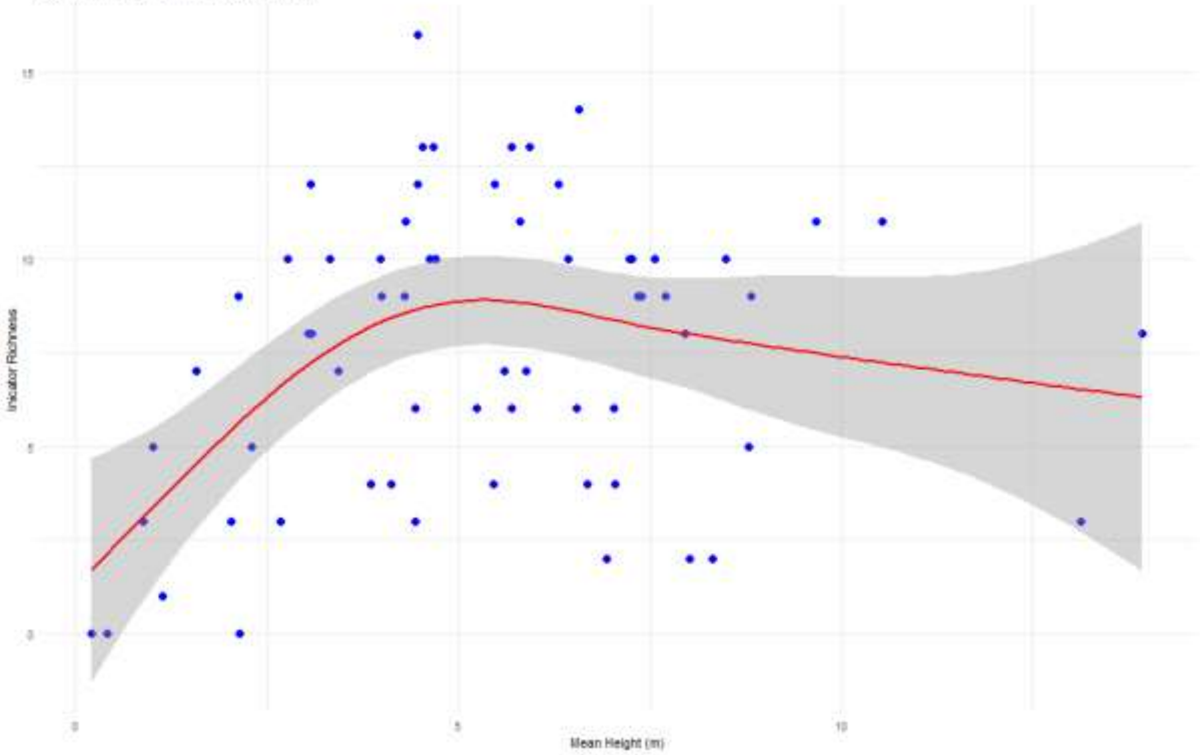

Distance from Standing Water and RRA Score
RRA Score
20
10
0
-10
200
400
600
800
Distance (m)

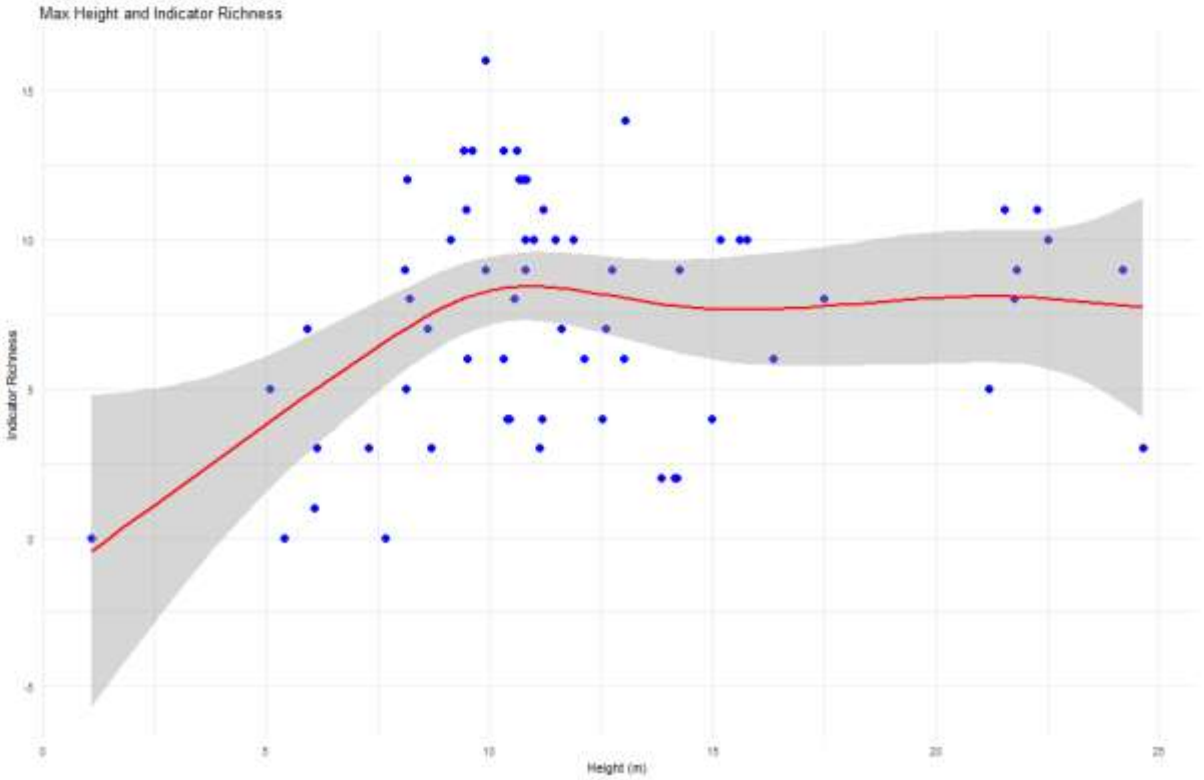

Wind Exposure and RRA Score
RRA Score
20
10
0
-10
1
2
Exposure (1 = protected, 2 = exposed)

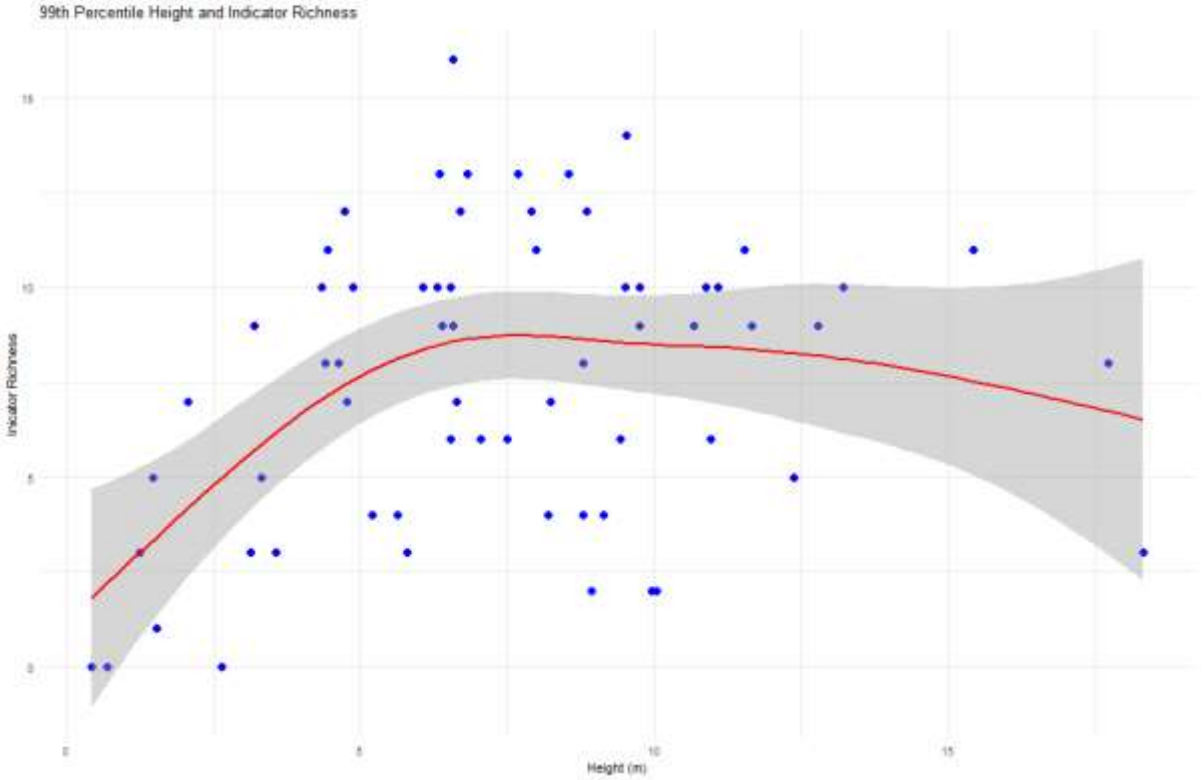

Elevation and RRA Score
RRA Score
20
10
0
-10
10
20
30
Elevation (metres)

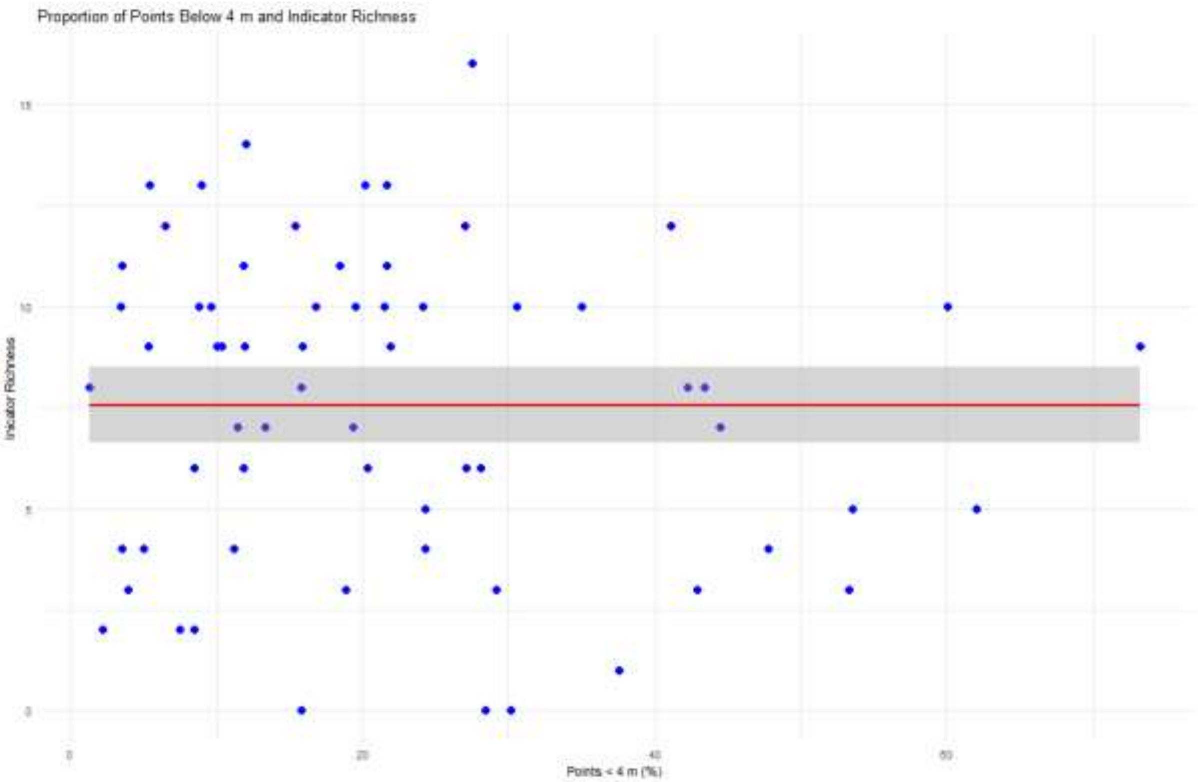

Aspect and RRA Score
RRA Score
20
10
0
-10
100
200
300
Aspect (degrees)


Height (SD) and Indicator Richness

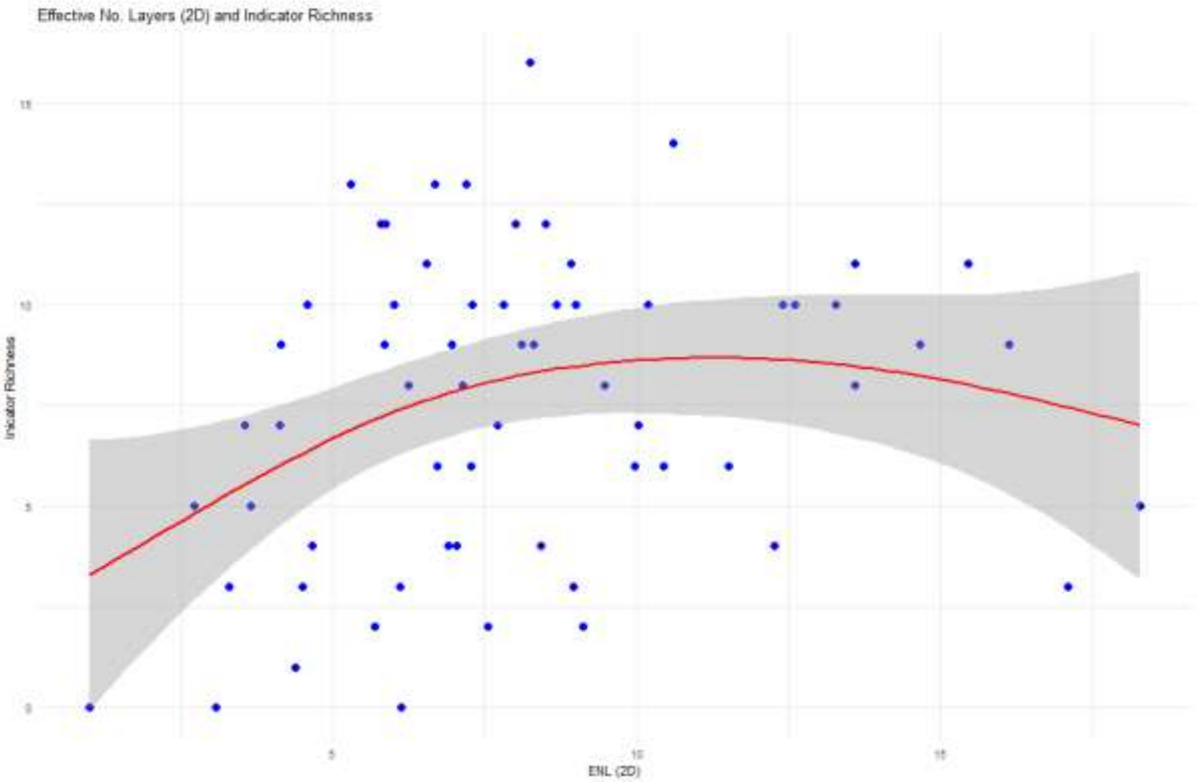

Average Height and Indicator Richness


Max Height and Indicator Richness
Indicator Richness
15
10
5
0
-5
0
5
10
15
20
25
Height (m)

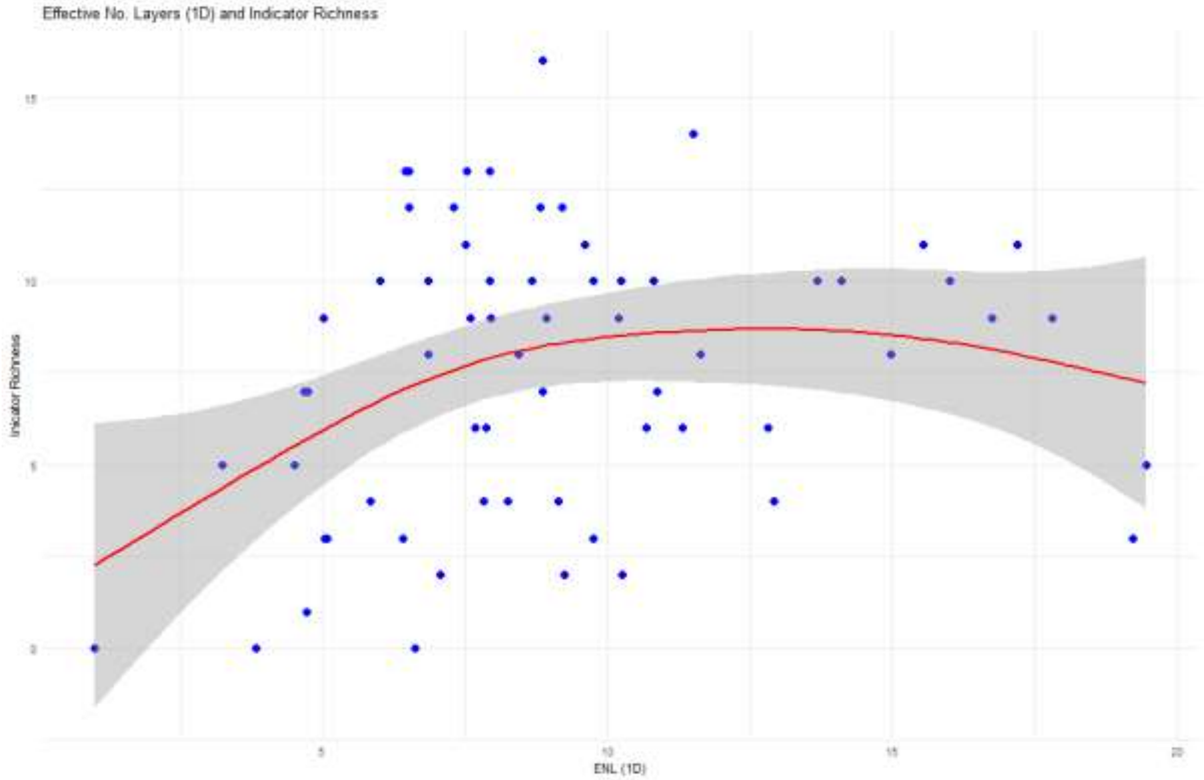

99th Percentile Height and Indicator Richness
Inicator Richness
15
10
5
0
0
5
10
15
Height (m)

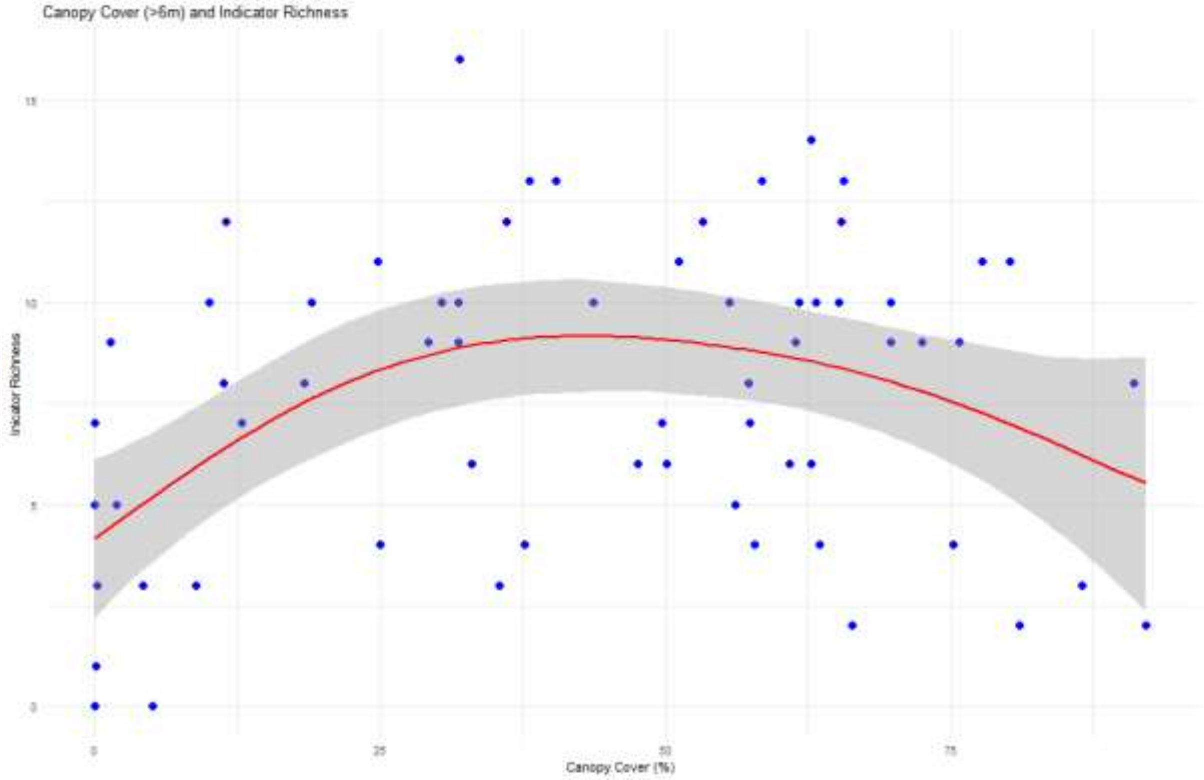


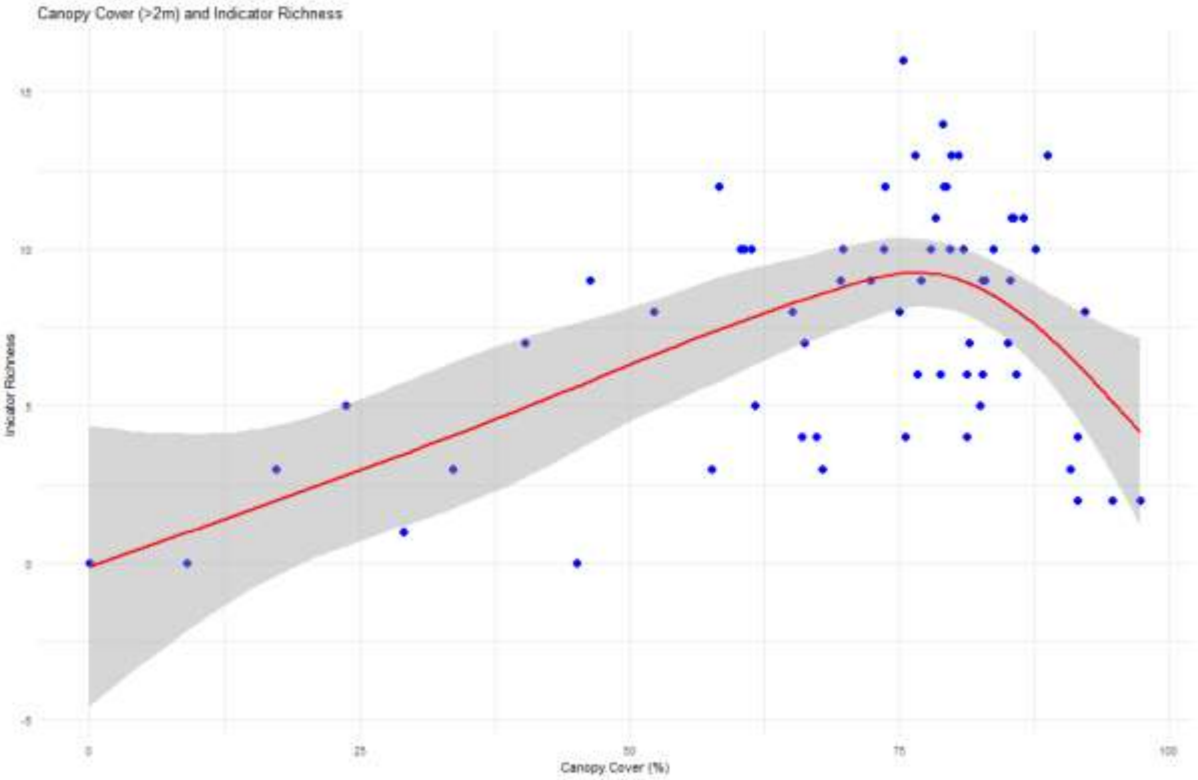
Proportion of Points Below 4 m and Indicator Richness

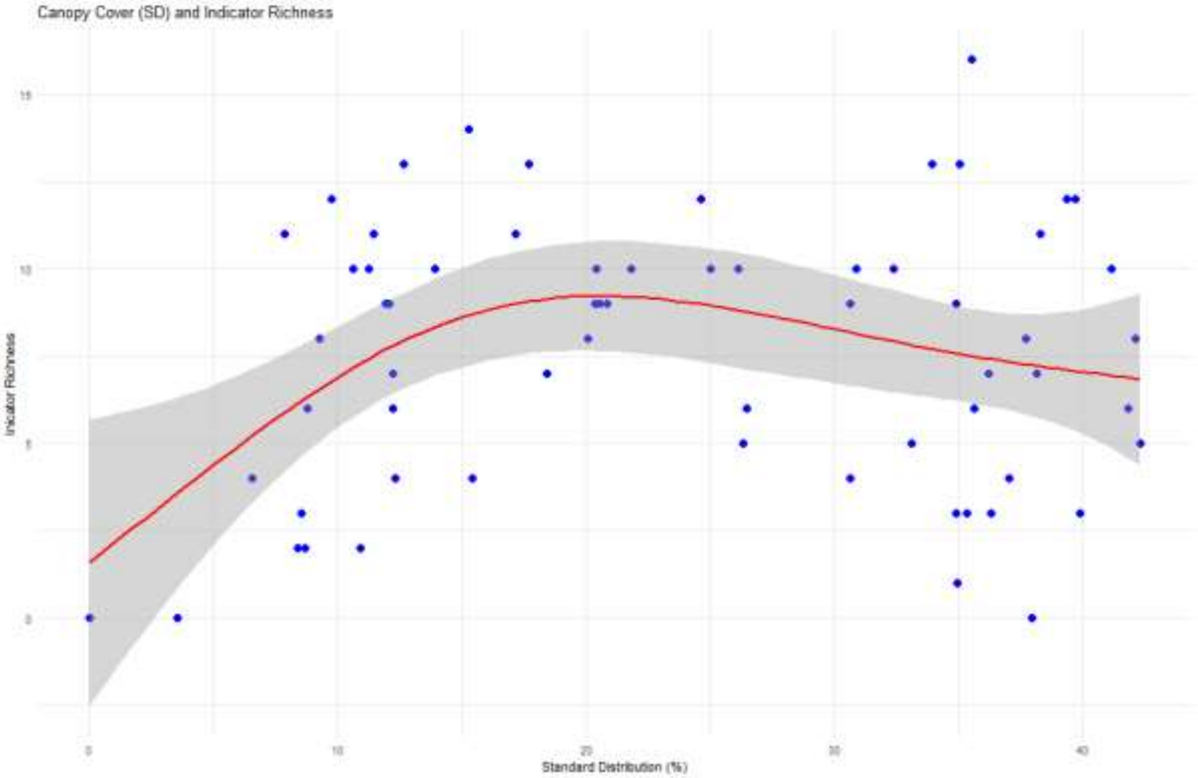


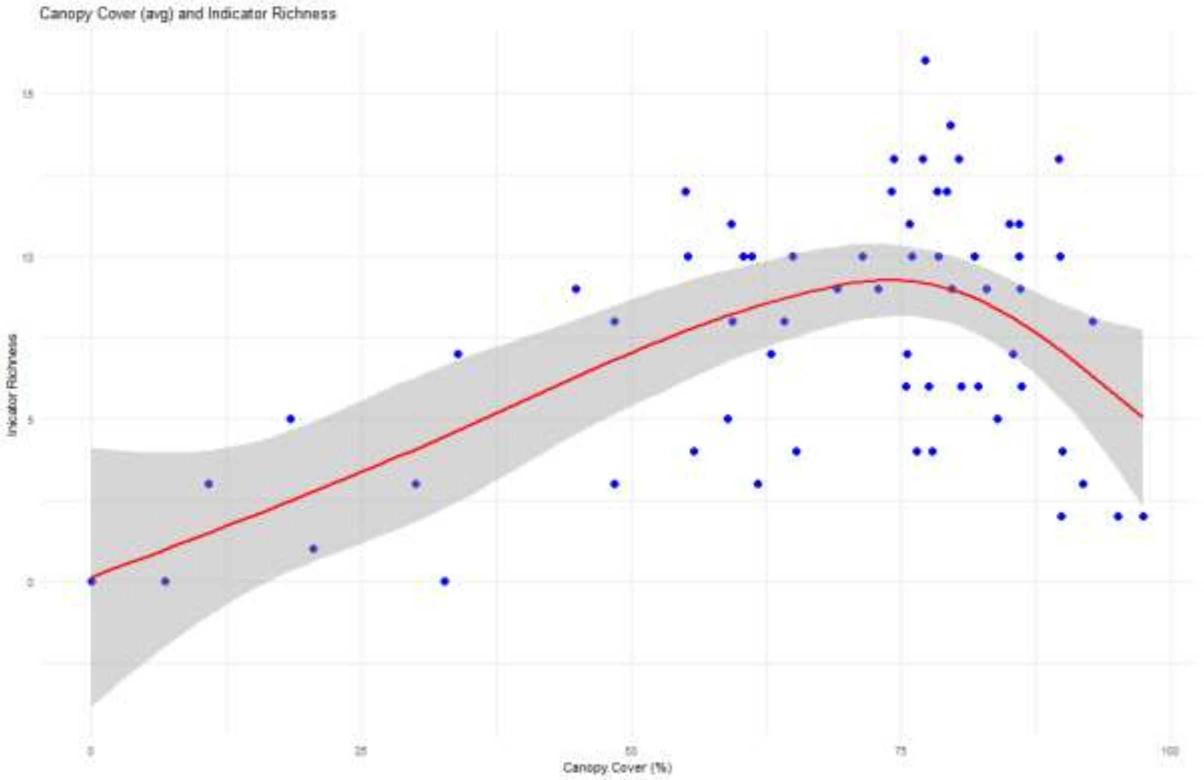
Effective No. Layers (2D) and Indicator Richness


Effective No. Layers (1D) and Indicator Richness

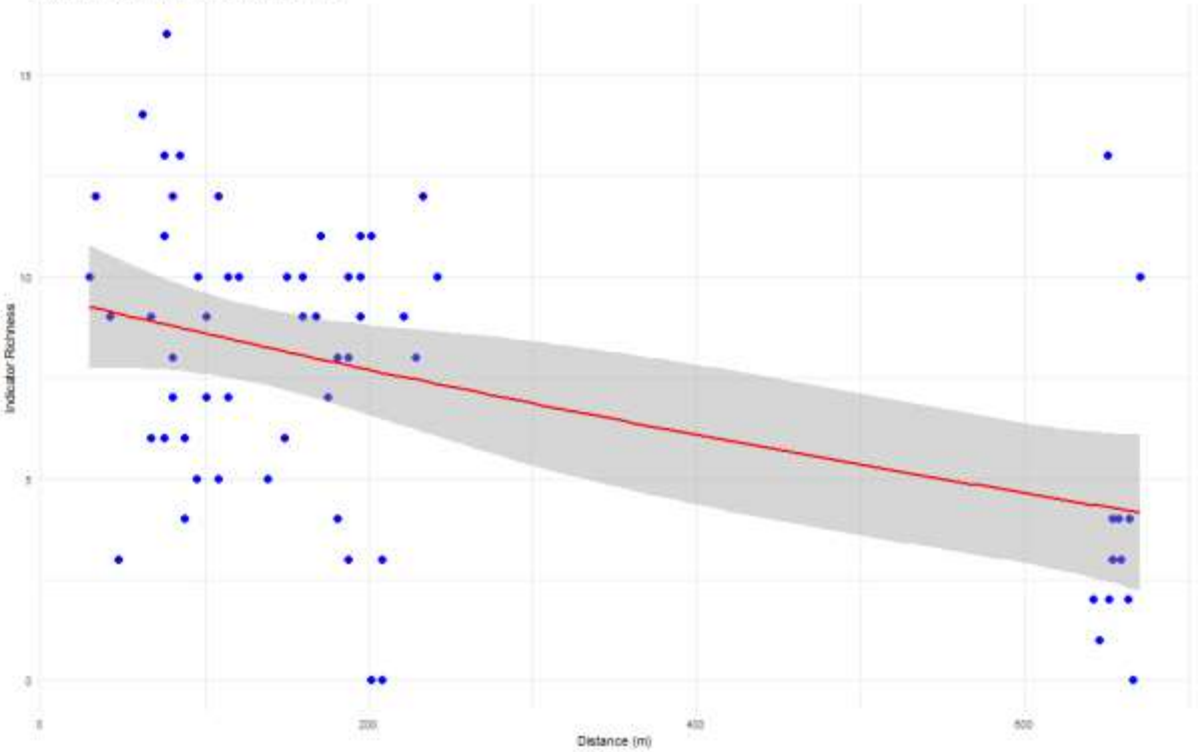

Canopy Cover (>6m) and Indicator Richness


Canopy Cover (>2m) and Indicator Richness
Inicator Richness
15
10
5
0
-5
0
25
50
75
100
Canopy Cover (%)

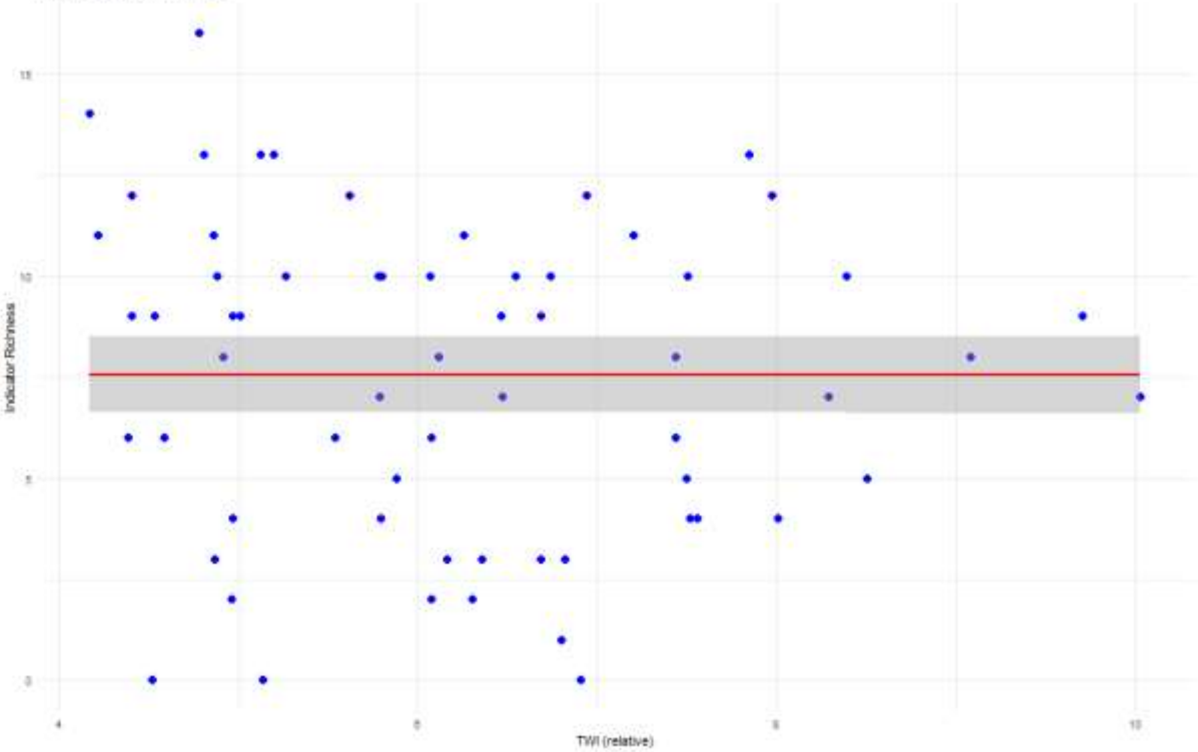

Canopy Cover (SD) and Indicator Richness
Inicator Richness
15
10
5
0
0
10
20
30
40
Standard Distribution (%)

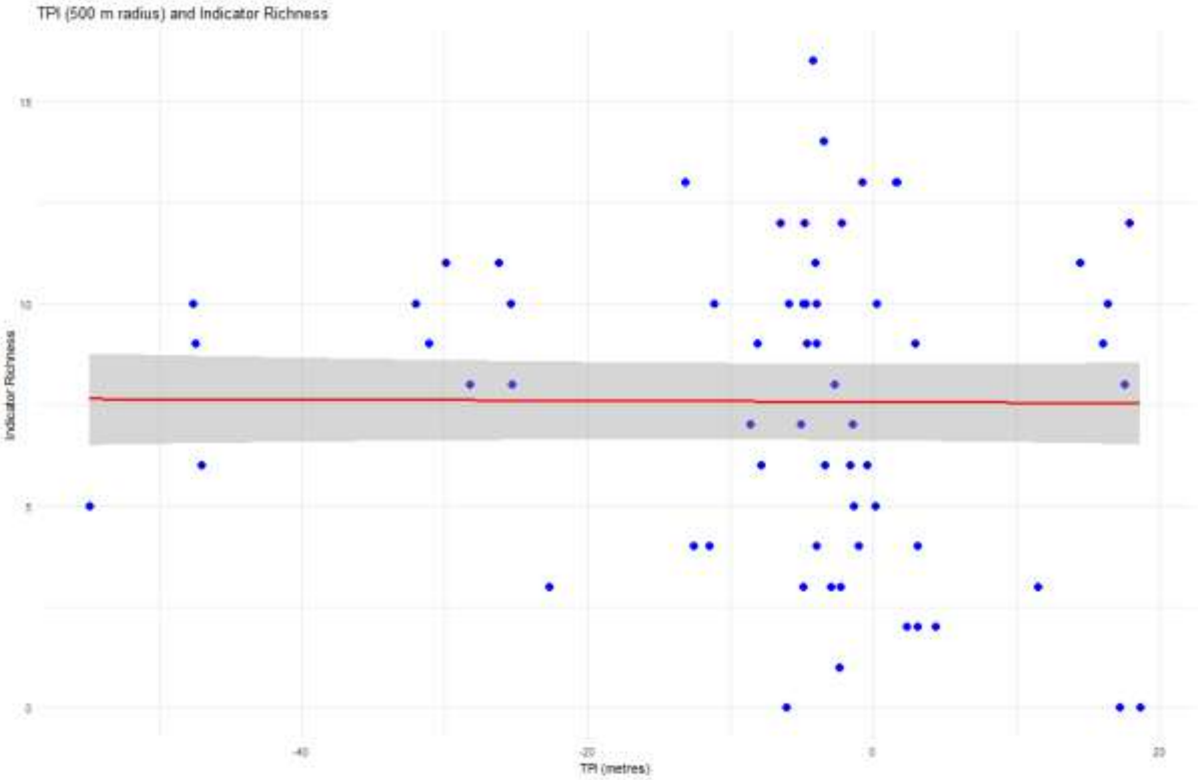

Canopy Cover (avg) and Indicator Richness
Inicator Richness
15
10
5
0
0
25
50
75
100
Canopy Cover (%)

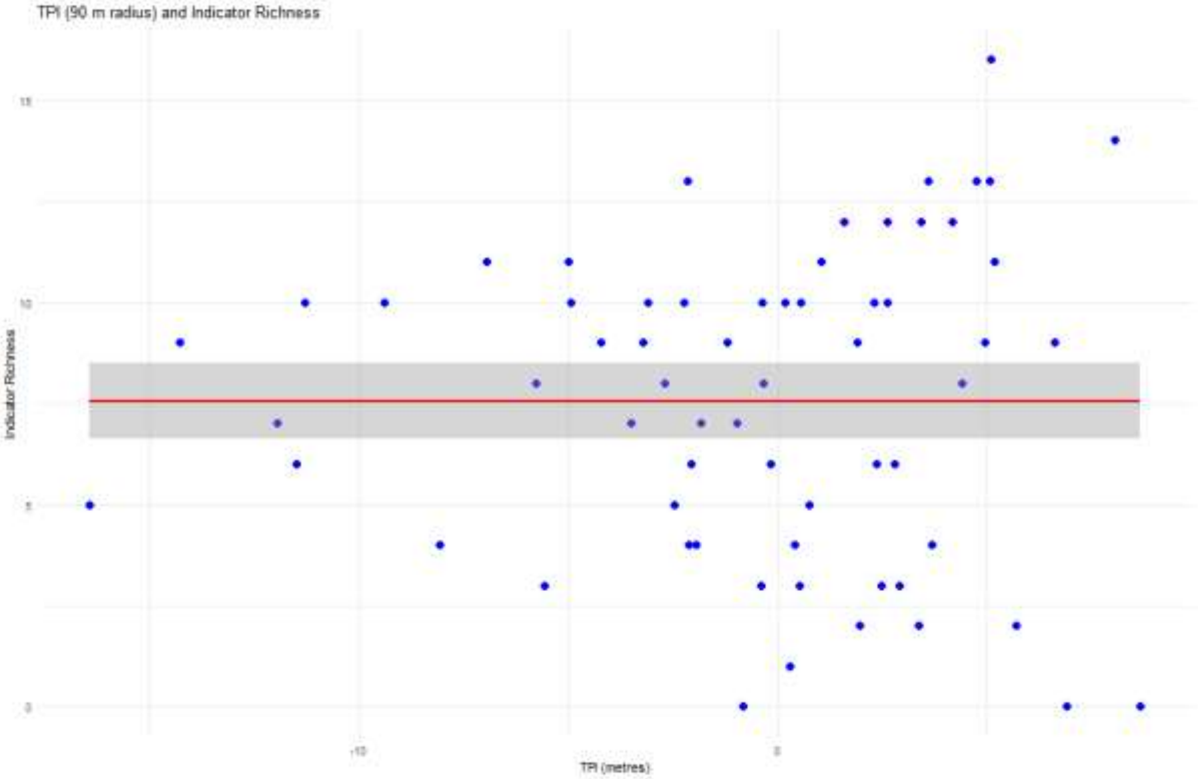
Distance from Water and Indicator Richness

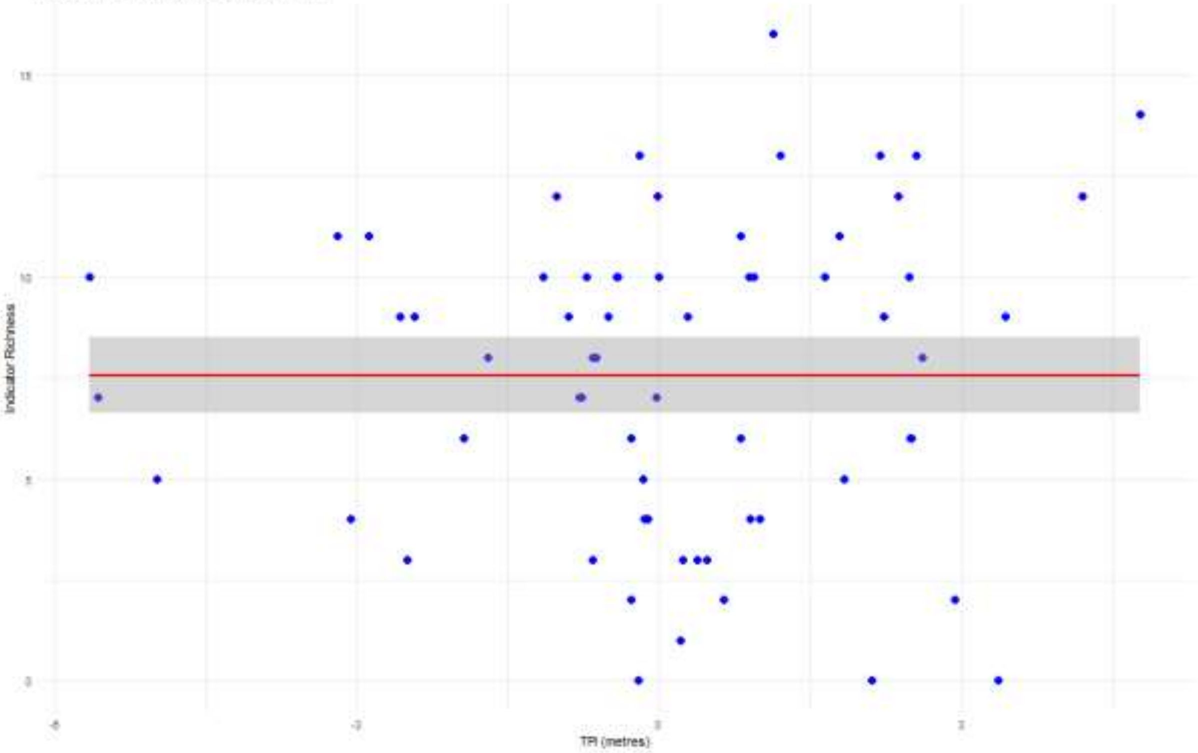


TWI and Indicator Richness

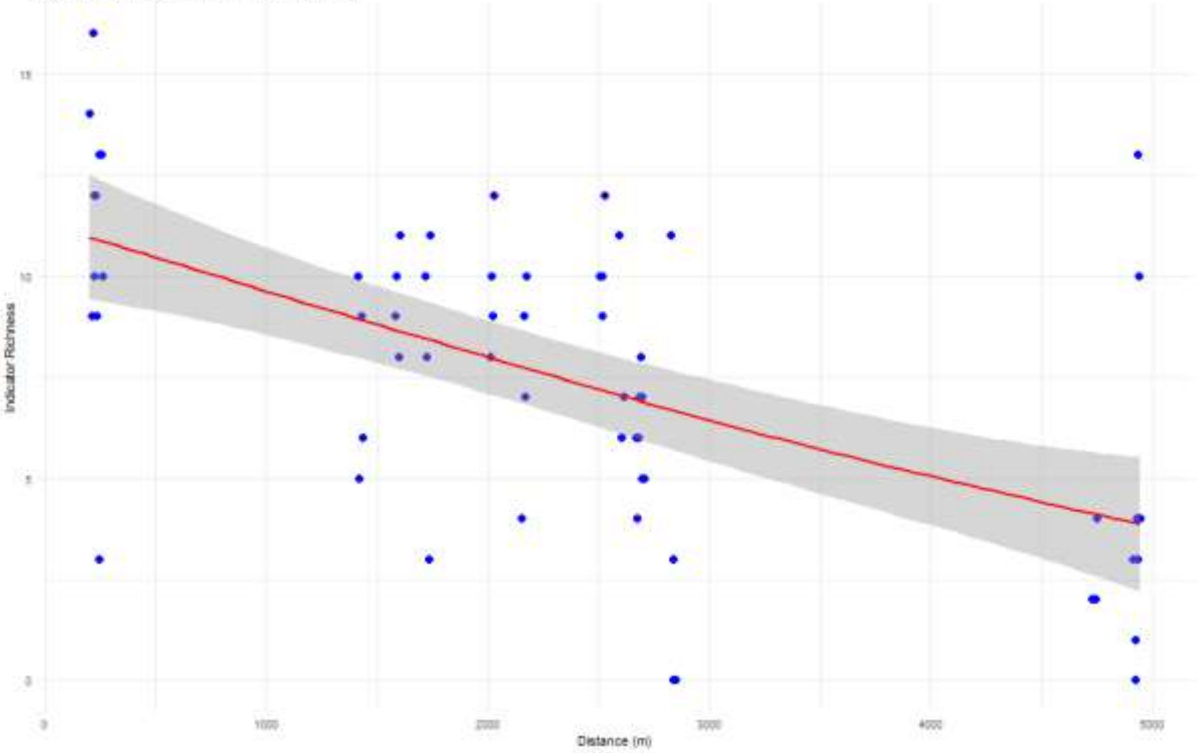



TPI (500 m radius) and Indicator Richness
Indicator Richness
15
10
5
0
-40
-20
0
20
TPI (metres)

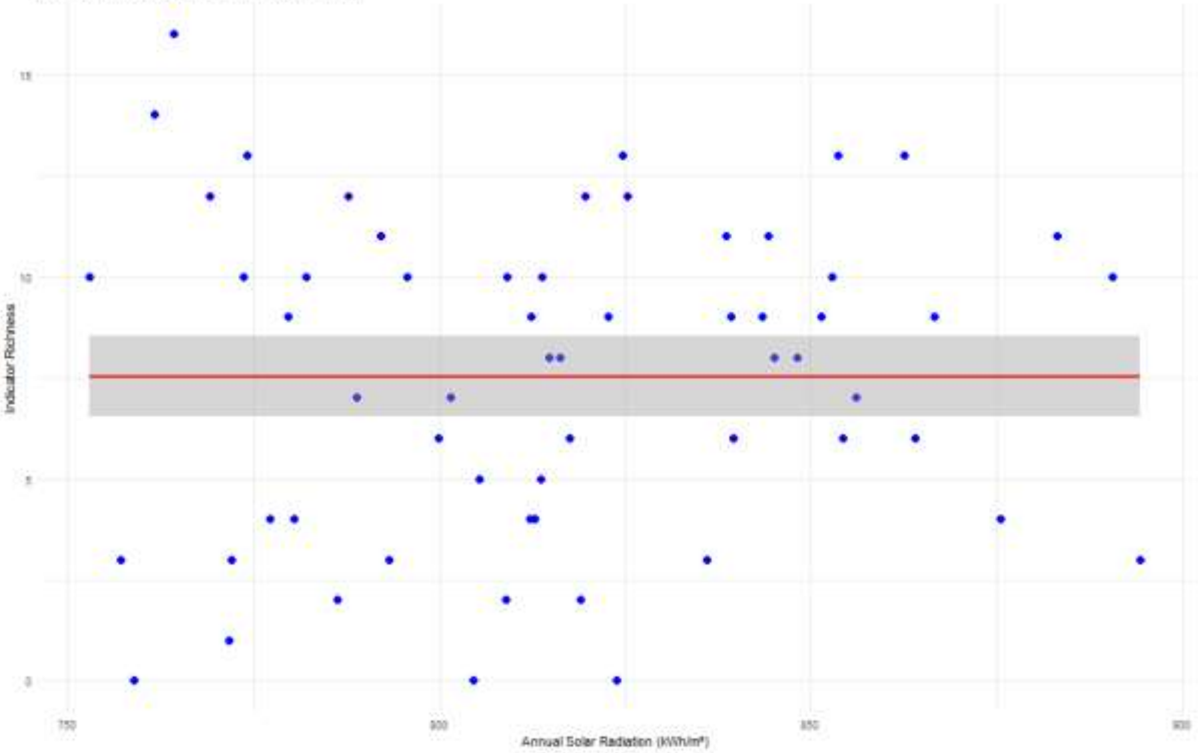

TPI (90 m radius) and Indicator Richness
Indicator Richness
15
10
5
0
-10
0
TPI (metres)



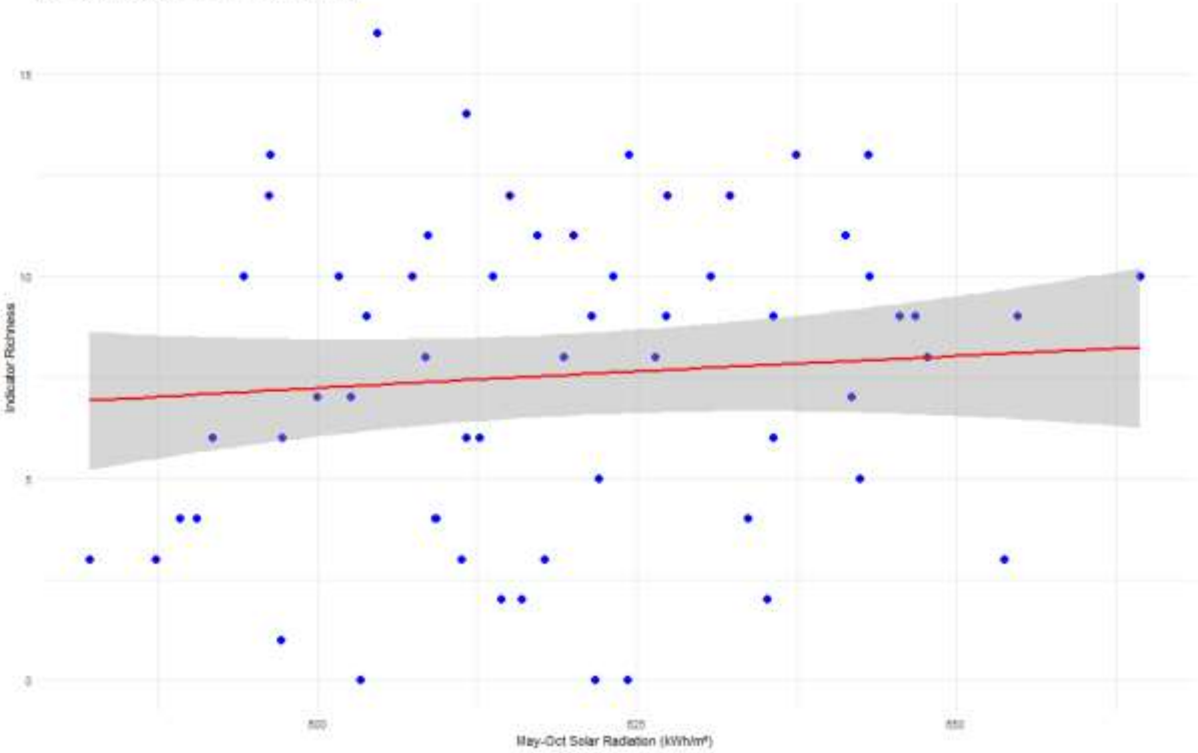
TPI (30 m radius) and Indicator Richness



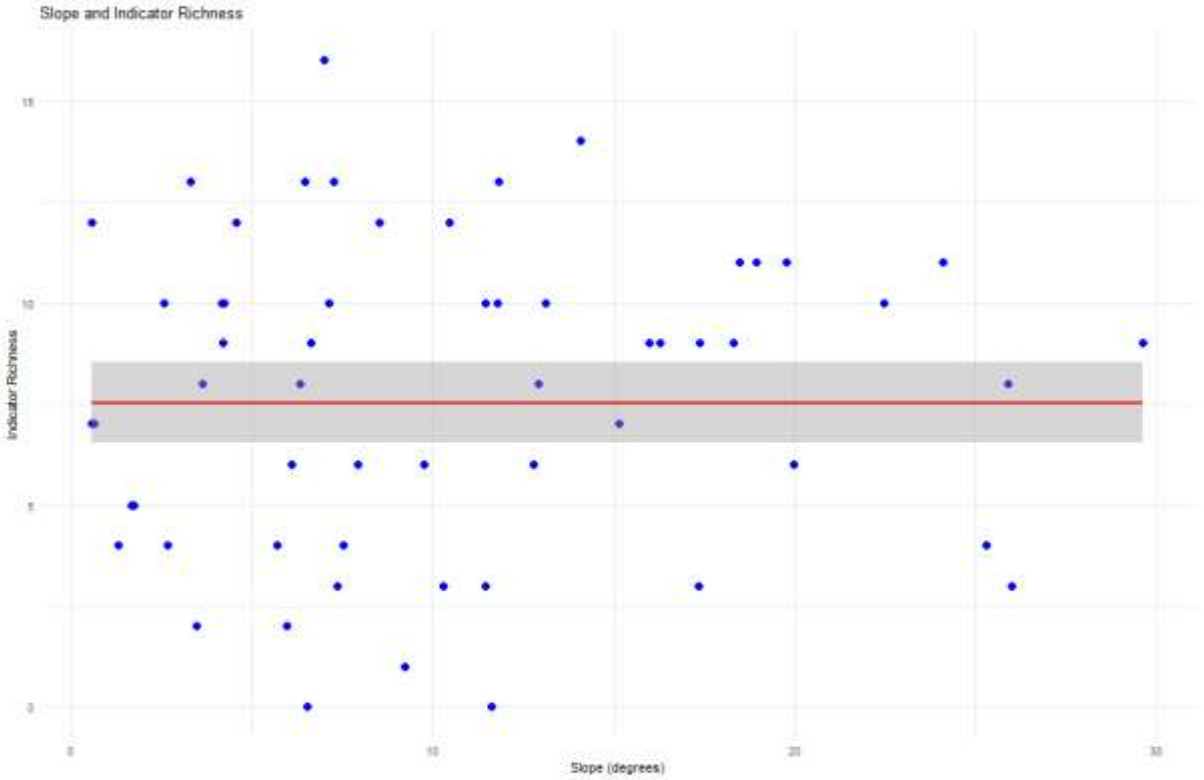
Distance from Taynish and Indicator Richness

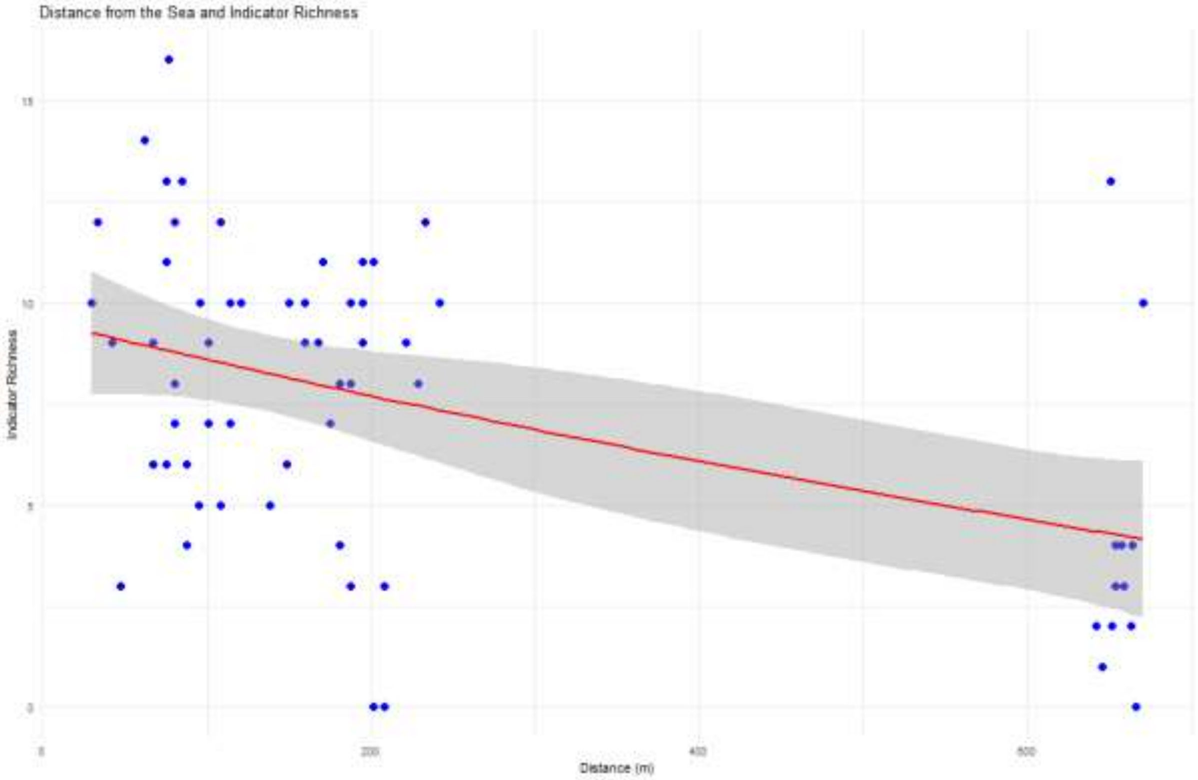


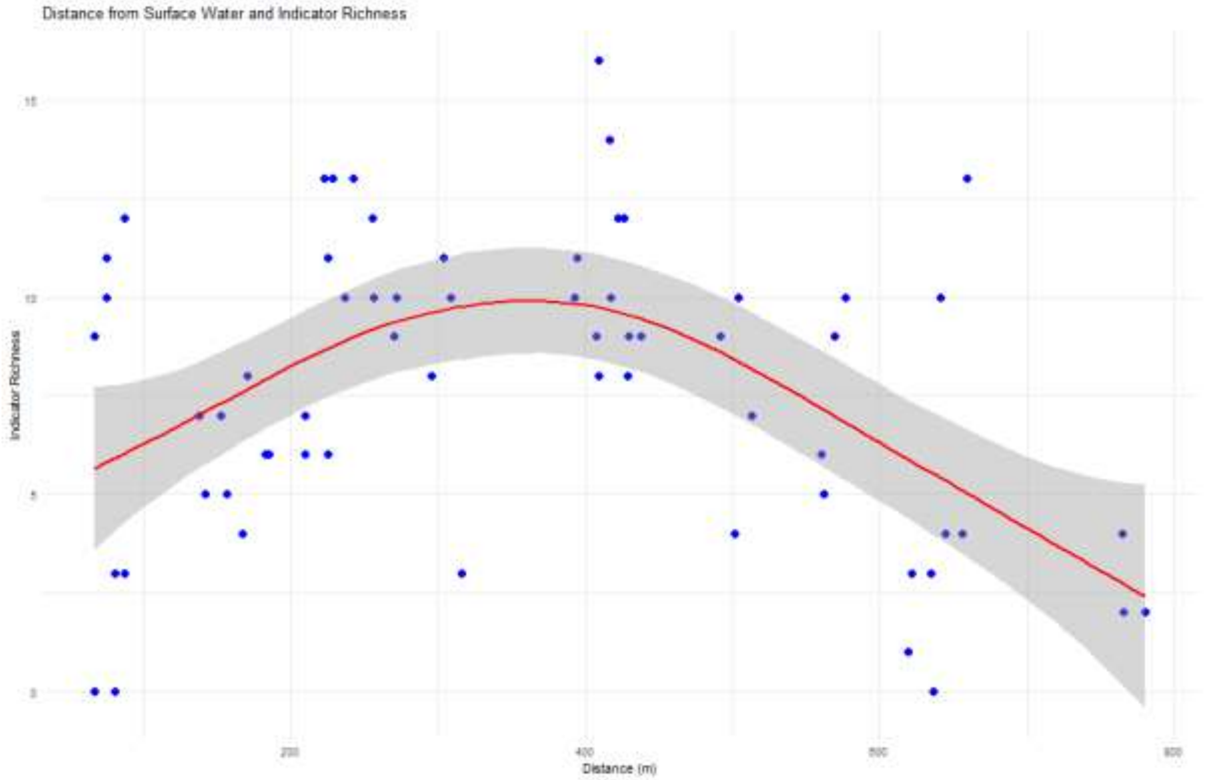
Solar Radiation (DTM) and Indicator Richness

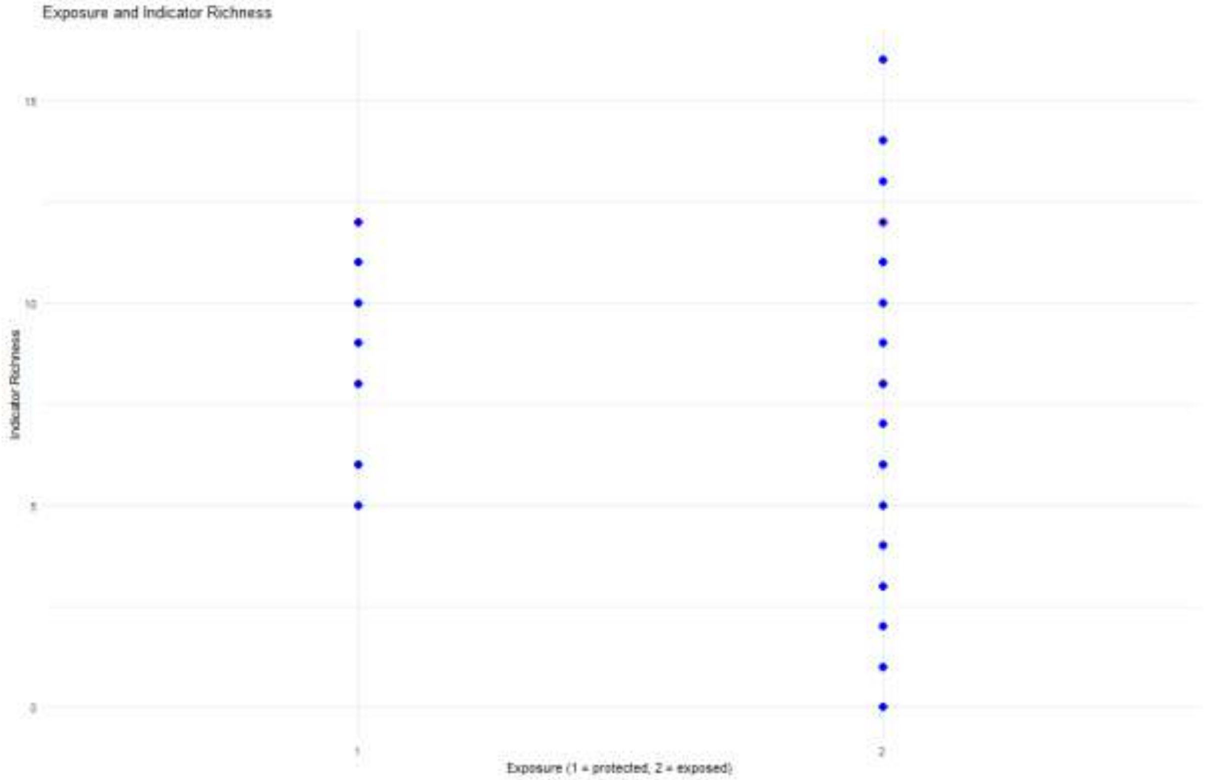


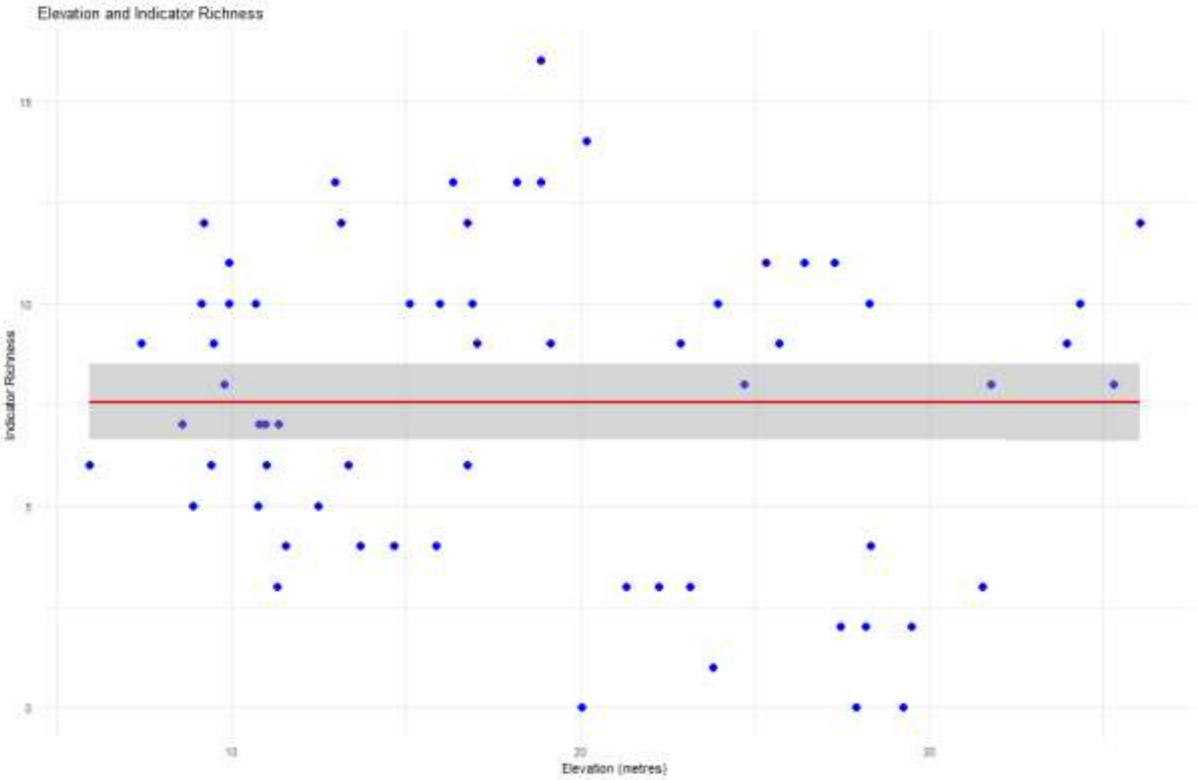
Solar Radiation (DSM) and Indicator Richness

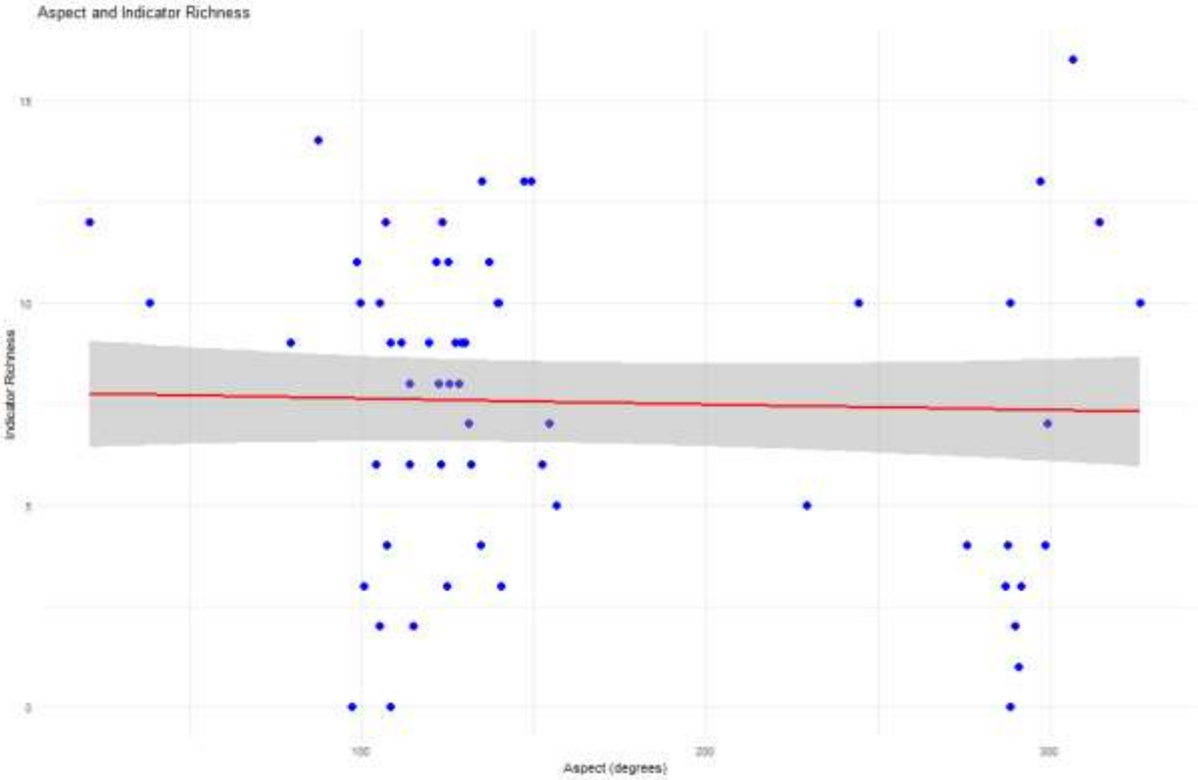

Slope and Indicator Richness
Indicator Richness
0
5
10
15
0
10
20
30
Slope (degrees)


Distance from the Sea and Indicator Richness
Indicator Richness
0
5
10
15
0
200
400
600
Distance (m)


Distance from Surface Water and Indicator Richness
Indicator Richness
15
10
5
0
200
400
600
800
Distance (m)


Exposure and Indicator Richness
Indicator Richness
15
10
5
0
1
2
Exposure (1 = protected, 2 = exposed)


Elevation and Indicator Richness
Indicator Richness
15
10
5
0
10
20
30
Elevation (metres)


Aspect and Indicator Richness
Indicator Richness
15
10
5
0
100
200
300
Aspect (degrees)

C – Data Availability and Structure

A Note on Data Availability

Due to terms in the signed data agreement, the ALS data, orthophoto (used for visualisation), and other data received from Highlands Rewilding for this study are not accessible in the M drive. Consequently, should a demonstration be required, please contact s2748011@ed.ac.uk.

Other data, including open-source data and data created throughout the project, are available and the locations of which can be found below.

Folder Structure and Contents

The table below highlights the folder structure found at M:\dissfinal.

Table 5: dissfinal folder structure.

diss_final:	data:	dataframe
		metrics
		site_notes
		vectors
	scripts:	data_management
		field_info
		tree_detection
		metrics
		dump
	gams:	rra_score
		richness
	plots	
	forms	

The 'data' folder contains all data not included in the data agreement, such as the Taynish polygon and the metrics.

The 'scripts' folder contains all R scripts used throughout the projects, including ones containing bugs and for performing small tasks.

The 'gams' folder contains all GAM outputs, including all partial effect graphs for both rra_score and richness. Explanatory power and performance results can be seen in the corresponding CSV files. Partial effect graphs can be seen in the individual folders, which are named after the model used.

The 'plots' folder contains the scatter plots and the correlation matrix.

The ‘forms’ folder contains the ethics forms, risk assessments, lichen guides, RRA survey form, and the data agreement.