Drivers of vascular plant species area relationships in British broadleaved woodlands and their effects on the species area curve

Petra Guy[1*], Simon Smart[2], Colin Prentice[1],

[1]Imperial College, Silwood Park, Buckhurst Rd, Ascot, SL5 7QN, UK. [2]Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster University, Library Avenue, Bailrigg, Lancaster, LA1 4YQ, UK

*Correspondence author: petra.guy.uk@gmail.com

Data is available from Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster University, Library Avenue, Bailrigg, Lancaster, LA1 4YQ, UK

https://catalogue.ceh.ac.uk/documents/ddff0f17-c95d-4415-80cb-aa9487edcb06

Summary

The loss of plant biodiversity in Great Britain is a major concern, with a fifth of species endangered or vulnerable according to the latest IUCN Red List. The Government's 25 Year Plan for the environment aims to halt this loss and build new habitats, including new woodlands. To ensure that biodiversity loss is halted in existing woodlands and gain is maximised in new ones, we need to better understand which drivers have been most influential in controlling biodiversity. Here we focus on vascular plant species' richness.

Previous attempts to explain plant species richness have mainly focussed on alpha diversity in a consistent, fixed unit area. Here, we additionally undertake a novel analysis of the effects of environmental heterogeneity and abiotic factors on species-area relationships derived from 16 randomly placed quadrats in each of 103 semi-natural, broad-leaved woodlands across Britain. Species-area relationships were examined at two scales ($4m^2$ to $200m^2$ and $200m^2$ to $3200m^2$) to explore the relationship between the drivers of species richness and the exponent z, of the canonical species-area curve, $S = cA^z$. We also explore the use of a new metric ζ_r, based on zeta diversity to quantify heterogeneity. Zeta diversity quantifies the number of species shared between multiple combinations of plots.

Habitat heterogeneity increased species richness, as did the proximity of the woodlands to surrounding natural habitats. Higher levels of soil organic matter and the progression of woodlands to later successional stages, decreased species richness. Richness was also seen to have a unimodal response to soil acidity with a peak around pH 6. At the smaller scale, heterogeneity elements in the woodland such as riparian zones or coppicing led to an increase in the value of the exponent of the species area curve. At the larger scale, species turnover led to an increase in the exponent of the curves while succession led to a decrease. At both scales, soil organic matter content had a negative effect. ζ_r was found to be a significant and important variable and to affect both species richness and the slope of the species accumulation curves at larger scales.

Synthesis. Habitat heterogeneity measures included the presence of coppicing, open areas such as rides and riparian zones and the difference between species assemblages in different plots in the woodland. Results suggest that to maximize vascular plant diversity, woodlands should be managed for heterogeneity. In addition, the increase in richness with exposure to surrounding natural habitats suggests that woodlands benefit from being embedded in more benign habitats and further, that land management surrounding woodlands has a clear role to play in supporting biodiversity within woodlands. This is an area were Agri-environment schemes have an important role.

Key words: biodiversity, species area curves, species richness, woodlands, zeta diversity.

Introduction

Biodiversity worldwide continues to decline (Butchart et al., 2010, Tittensor et al., 2014) and Britain is no exception to this trend (Hayhow et al., 2016). The IUCN Red List for the UK shows 22% of plant taxa as critically endangered, endangered or vulnerable (Stroh et al., 2014). Biodiversity is important for ecosystem function (Gaston & Spicer, 2004), with woodlands being especially important in supporting ecosystem functioning and supply of ecosystem services. These include improvement of water quality, flood control (McCulloch and Robinson, 1993) carbon sequestration (Brainard, Bateman and Lovett, 2009; Ostle *et al.*, 2009)nectar provision within farmed landscapes (Baude et al 2016) and the provision of habitats for other taxa (Amar et al., 2006, Fuller & Warren, 1993).

The aims of the new 25-Year Environment Plan (Defra, 2018) include achieving "thriving plants and wildlife", planting 180,000 hectares of new woodland and ensuring that existing woodlands are better managed. However, more woodland is not necessarily the answer to biodiversity loss if we cannot ascertain why our current woodlands are not thriving.

This study was aimed at finding the main drivers of vascular plant species richness in British broad-leaved semi-natural woodlands to provide research-based evidence on how best to manage, maintain and increase their plant biodiversity. We analyse diversity as the slope of the species area relationship expressed across multiple samples at the national scale.

SUMMARY OF FACTORS THAT HAVE BEEN SHOWN TO AFFECT SPECIES RICHNESS

Many factors have been shown to affect vascular plant species richness in woodlands, including climate, edaphic properties, eutrophication, shading, presence of invasive or dominant species, habitat fragmentation and isolation, disturbance and previous land use, to name but a few. (Dzwonko and Loster, 1988; Dumortier *et al.*, 2002; Dzwonko and Gawroński, 2002; Petit *et al.*, 2004; Kirby *et al.*, 2005; Brudvig and Damschen, 2011; Hill and Preston, 2015). Rather than carry out an in-depth review of all these effects, we review those factors whose effects we can represent from the CEH Woodland Survey database (Kirby et al., 1971) and justify the selection of these predictor variables by citing support in the published literature. This is important due to the model selection and averaging procedure used as part of this analysis which requires justification for the inclusion of variables from prior knowledge.

Studies of beech and hornbeam woodlands (Kooijman and Cammeraat, 2010) report lower species richness at lower soil pH and when soil organic matter content is higher. This was true for both taxa,

suggesting that the effect was not the result of the physical barrier created by slowly degrading beech litter.

Cornwell & Grubb (2003) showed that the greatest richness in woodlands occurred on nutrient rich soils, reporting a unimodal response with a peak around Ellenberg N of 7 for shade loving species. Since most macronutrients are available around pH5/6 this implies that a unimodal richness response in woodlands should be expected.

Ohlemuuler & Wilson, (2000) showed a decrease in richness with latitude in New Zealand for woody species, but this variation was not seen in herbaceous species. In studies on forests stretching from the equator to 60° latitude, Gillmann (2014) found that net primary production increases toward the equator and concludes that, since vascular plant richness is correlated with net primary production, richness will also increase. However, this effect may not be evident given the modest change in latitude in our UK data.

The 2010 Lawton report (Lawton et al, 2010) summarised the need for more connectivity within our landscape. Many authors, in many different settings, have shown that fragmentation reduces biodiversity. Brudvig et al (2009) showed that plants which depend on animals for dispersal increased when corridors connected experimental patches of regenerating pine forest. Thiele et al., (2018) studied plant species richness in agricultural, fragmented landscapes and concluded that connectivity increased the richness of grassland and wetland plants. Hannay et al. (1997) used the amount of forested area around a woodland to quantify isolation, and isolation and showed that richness decreased as the amount of surrounding woodland decreased. Petit et al., (2004) showed that fragmentation constrains the dispersal of ancient woodland species in British lowland woodlands. Dzwonko & Loster (1988) showed that woods which that had been isolated for longer in the agricultural landscapes in the Carpathian foothills had lower species richness.

Habitat heterogeneity is known to be important for biodiversity (Gardener, 2010). It has been shown to increase richness when it takes the form of woodland management (Boch et al, 2013, Schmidt, 2005), presence of less shaded features such as gaps, rides and paths, number of soil types, the number of different habitats the woodland contains (Honnay et al, 1999) and windthrow (Smart et al, 2014).

SPECIES AREA CURVES AND DRIVERS AND z VALUES

Species area curves describe the change in the number of species accumulated over an area. They can be used to predict numbers of species over areas larger than those at which sampling is possible or to see if sampling is exhaustive. These curves are often described by the power model,

$$S = cA^z$$

Where S is the number of species in area A, c is the number of species in a unit area and the exponent z describes how the slope of the curve changes with area. There have been many debates about the biological meaning of the z value. Z values of around 0.25 were first proposed by Preston (1960) as being the result of sampling from log-normal species distributions. This value would be expected to occur when sampling from similar habitats, for example from nested quadrats. Other authors have suggested that the z value has no direct biological meaning (Connor et al, 1983, Tjorve & Tjorve, 2017) but is a result of species distribution and aggregation. Using neutral models, Cencini et al (2012) predicted that population size affects the shape of the curve while Sugihara (1979) concludes that the value of 0.25 reflects community structure.

In this study, species were accumulated in different ways depending on the scale. At the smaller scale, species area curves were generated for species across nested areas. These are referred to as Type I curves (Scheiner, 2003). At this scale a z value of around 0.25 is expected. At larger scales we also accumulated species across all the non-contiguous quadrats in the site. These curves are referred to as species accumulation curves and at this scale habitat heterogeneity is likely to exert a greater influence on the z value (Storch et al, 2014). In similar work, Shen et al (2009) considered the effect of tree density, soil nutrients, pH, slope, elevation and aspect and heterogeneity and concluded that dispersal and habitat heterogeneity jointly are required to explain the shape of species area curves. Baldi & Sadler (2017) concluded that habitat heterogeneity, as quantified by the number of habitat types as shown through CORINE land cover maps, was more important than area for expressing species richness.

In this work, the combined influence of local abiotic factors and woodland heterogeneity on the exponent of curves constructed over these two different scales, was examined.

HABITAT HETEROGENEITY – THE USE OF ZETA DIVERSITY

Habitat heterogeneity is known to have a major influence on biodiversity (Gardener, 2010), but it is not trivial to quantify and has been expressed in many ways. Honnay et al., (1999) used length of rides, number of soil types and a ratio of perimeter and area. Schmidt (2005) used the presence of logging

trails, and Bàldi (2008) used the number of CORINE land cover codes. Here we quantified heterogeneity by combining several variables (Table 2) into a Heterogeneity Index (HI). The index has some shortcomings since it does not enumerate the numbers of each feature that occur in the woodland but merely records a feature as present irrespective of quantity or extent. This will reduce the sensitivity of the index. In addition, the number of elements identified and will depend on how exhaustively the entire site was surveyed.

A new metric was also generated based on zeta diversity (Hui & McGeoch, 2014) to provide a simple, objective measure of heterogeneity in species composition between plots within each woodland.

ZETA DIVERSITY

Zeta diversity is the average number of species shared between multiple plots, (Hui & McGeoch 2014) and can be used to show compositional change in species assemblages, (McGeogh et al., 2017) and model drivers of species turnover, (Latombe et al., 2017).

As an example, consider four plots with sets of species S_1, S_2, S_3, S_4. Let S_i represent the set and number of species in plot 1. Four orders of zeta can be calculated in the following way.

$$\zeta_1 = \frac{S_1 + S_2 + S_3 + S_4}{\binom{4}{1}} \qquad 1$$

$$\zeta_2 = \frac{S_1 \cap S_2 + S_1 \cap S_3 + S_1 \cap S_4 + S_2 \cap S_3 + S_2 \cap S_4 + S_3 \cap S_4}{\binom{4}{2}} \qquad 2$$

$$\zeta_3 = \frac{S_1 \cap S_2 \cap S_3 + S_1 \cap S_2 \cap S_4 + S_1 \cap S_3 \cap S_4 + S_2 \cap S_3 \cap S_4}{\binom{4}{3}} \qquad 3$$

$$\zeta_4 = \frac{S_1 \cap S_2 \cap S_3 \cap S_4}{\binom{4}{4}} \qquad 4$$

$S_i \cap S_j$ is the number of species shared between plots i and j. For a woodland with 16 plots, zeta can be calculated to ζ_{16} as shown in figure 1.

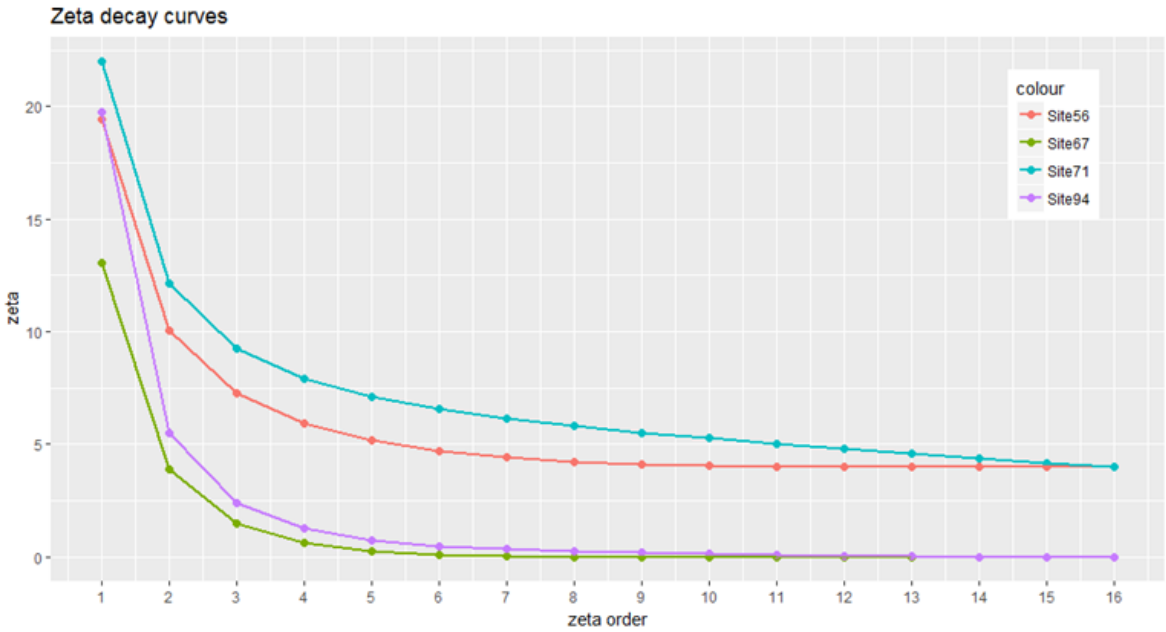

Figure 1. Examples of zeta decay curves calculated for 4 sites in the Bunce survey. The gradient of the initial part of the curve was used to create a metric and used as an explanatory variable to represent species heterogeneity in the woodland.

ζ_1 is the average number of species in each plot and is therefore equivalent to mean α diversity. ζ_2, is the average number of species shared between every combination of two plots.

The slope of the zeta decay curve varies with the probability of finding new species in new plots (Hui & McGeogh, 2014) and the number of common and rare species (Latombe et al., 2017). If plots contained the same species, then $\zeta_1 - \zeta_2 = 0$ and the initial gradient of the curve = 0. If all plots contained different species then $\zeta_2 = 0$, and the initial gradient of the curve would equal ζ_1.

A homogeneous woodland is more likely to contain species that occur in many plots, whereas a heterogeneous woodland is more likely to have plots containing different species. This suggests a method of comparing heterogeneity between woodlands using the ratio,

$$\zeta_r = \frac{\zeta_1 - \zeta_2}{\zeta_1},$$

ζ_r is then the initial gradient of the zeta decay curve normalised by ζ_1 to give a value between 0 and 1.

The more species-compositionally homogeneous the woodland, the closer the value of ζ_r to zero.

The zeta values for each woodland were calculated using the zetadiv function in R, (Latombe et al., 2018). The ζ_r coefficient was calculated and added to the set of predictor variables.

In summary, we sought to answer three questions:

1) Which variables had most influence on species richness of the entire woodland?

2) Which variables had most influence on the exponent of the species area curve and did this change with scale?

3) Can the zeta diversity metric ζr be used to represent compositional heterogeneity?

Materials and methods

DRIVERS OF VASCULAR PLANT SPECIES RICHNESS ACROSS ENTIRE WOODLANDS AND AT TWO DIFFERENT SCALES.

The drivers of vascular plant species richness were studied in two ways. Firstly, by examining the effect of a range of variables on the species richness of a fixed area within 103 different woodlands and, secondly, by looking at the scale dependence of species richness as expressed in the canonical species area curve (SAC), $S = cA^z$, and the effect of habitat heterogeneity and environmental conditions on the exponent, *z* of these curves. The same modelling was applied to both tasks, that is, species richness and the z values at the two different scales were used as response variables across the same set of explanatory variables.

DATA

We used a dataset extracted from the UK Centre for Ecology and Hydrology's woodland survey database (Wood et al 2017). One hundred and three woodlands across Great Britain were surveyed between 2000 and 2003. Sixteen square plots were randomly located in each woodland, each having a total area of 200m^2. Further, each plot was itself nested with sub-areas of 4m^2, 25m^2, 50m^2, 100m^2 and 200m^2.

All vascular plants and a restricted list of bryophytes were recorded within each plot by searching each concentric nest starting at the central nest and then working outwards only recording extra species recorded in each larger nest. Additional plot data were collected, including diameter at breast height (DBH) of trees, live basal area (LBA) of trees and shrubs, soil pH, soil organic matter (SOM) content, National Vegetation Classification (NVC) code, and major soil group (MSG). Additional information was also recorded where pre-specified lists of woodland attributes and features occurred in each plot. These included signs of woodland management, the presence of riparian zones

and open areas such as glades or rides. Table 1 summarises the information taken from the dataset and details how it was transformed to create predictor variables.

A sample-based estimate of the λ diversity of each woodland was calculated to be the unique number of species in the total 16 plots surveyed. The variables within the database were at different scales, plot level and woodland level. Plot level variables, such as pH, DBH or SOM, were scaled appropriately to the richness by taking the mean per woodland. Site level variables were buffer, number of NVC codes, number of major soil groups, area ratio and HI

Table 1 Summary of predictors created from the variables in the dataset. Some plot scale variables, such as pH, were averaged across the woodland to give a site level variable.

Variable	Description
Heterogeneity Index (HI)	The presence of features that diversify the woodland were recorded where present. These included ponds, ditches, buildings, rock exposures and dead wood. The count off recorded features was used as an explanatory variable - HI. The greater the HI, the more elements within the woodland that could increase richness by increasing heterogeneity of conditions.
Northing	Latitude
Buffer	The proportion of land around the wood (to 3500m from the edge) that is not arable, improved grassland or urban. The larger the buffer, the more the surrounding area can be considered as a benign mosaic of habitats assumed to buffer the woodland against loss of typical woodland species and incursion of nutrient-demanding, ruderal species. This variable therefore equates to the inverse of isolation and measures the favourability of the matrix within which each woodland was embedded.
Number of major soil groups (MSG)	The number of major soil group codes in each woodland. This is an alternative indicator of woodland heterogeneity.
Number of National Vegetation Classification (NVC) codes	The number of different NVC codes in each woodland. The NVC code represents the assignment of the plant species assemblage to a phyto-sociological classification of British plant communities. The number of different NVC codes therefore describes the number of different plant assemblages present. This is equivalent to the number of CORINE land cover codes used by Baldi (2008). Although the NVC code might correlate with richness, the number of codes will not, since it is possible that each code could represent a low richness assemblage. However, since NVC codes represent a different species assemblage with different abiotic factors, if there are more codes, it is more likely that the woodland has greater heterogeneity.
pH	The pH of a 15 x 5cm soil sample from the centre of each plot
Soil Organic Matter (SOM) content	The SOM content
Live Basal Area (LBA)	The LBA of all saplings, trees and shrubs recorded in a plot. This variable is a measure of shading.
Diameter at breast height (DBH)	The DBH of all trees recorded in each plot. As with LBA, this variable is a proxy for shading.
Tree Density (TD)	Count of trees per plot. Again, tree density is an alternative indicator for shading.
Area ratio	The ratio of area to perimeter of the site. This is a measure of the number of habitats the woodland is exposed to (Honnay et al, 1999).
ζ_r	The ratio of ζ_1-ζ_2 to ζ_1, quantifying species turnover between plots.

STATISTICAL MODELLING

Two statistical modelling approaches were taken to address the first question. First, model selection using Akaike Information Criteria followed by model averaging (Burnham and Anderson, 2004) and second, a gradient boosting machine (GBM) (Natekin and Knoll, 2013). Both of these approaches were used to examine which of the predictor variables had the greatest effect on the species richness of the same area, 3200m^2 i.e. across the 16 x 200m^2 plots in each of the 103 woodlands.

The model selection procedure only was then repeated using the exponent, z, of the species area curves as the response with the same set of predictor variables over two scales; 4m^2-200m^2 (within-plot level) and 200 – 3200m^2 (between-plot level). Finally, we evaluated the importance of ζ_r as a heterogeneity measure by observing the behaviour of this predictor in the above models.

MODEL SELECTION AND AVERAGING

Model selection followed by model averaging is an ideal method for isolating variables which are most important for predicting a response when the relationships are linear, (Kimberley et al, 2014; Symonds & Moussalli, 2010). The process is frequently used in ecology and has been described in detail elsewhere (Burnham and Anderson, 2004; Johnson and Omland, 2004; Wagenmakers and Farrell, 2004; Zuur, Ieno and Elphick, 2010; Symonds and Moussalli, 2011). Therefore, a brief synopsis only of the steps involved will be given here, and is summarised in figure 2.

Firstly, predictor variables were chosen which are known to have biological significance, (see table 1, steps 1-4, 7-9,12 in figure 2). This is because model selection is a relative measure. Including predictors with low predictive power will result in a poor set of models. This method will always produce a “best” model set across which predicators are averaged, but that model set is only “best” relative to all the other models generated. Thus, a poor set of predictors will be unable to generate a set of models with good predictive power, (Symonds and Moussalli, 2011). This means that standard data exploration protocols for linear models was followed, such as examining collinearity between predictors (step 19, figure 2) and considering R^2 of the initial linear model (step 21 figure 2), (Zuur, Ieno and Elphick, 2010; Freckleton, 2011)). In this case we found that all variables had Spearman and Pearson correlation coefficients below 0.54. The variance inflation factors of the models were also examined and were all below 2. None of the models showed any pattern in the residuals. For the species richness models R^2 was 0.47. For nest z, 0.33 and for plot z, 0.47.

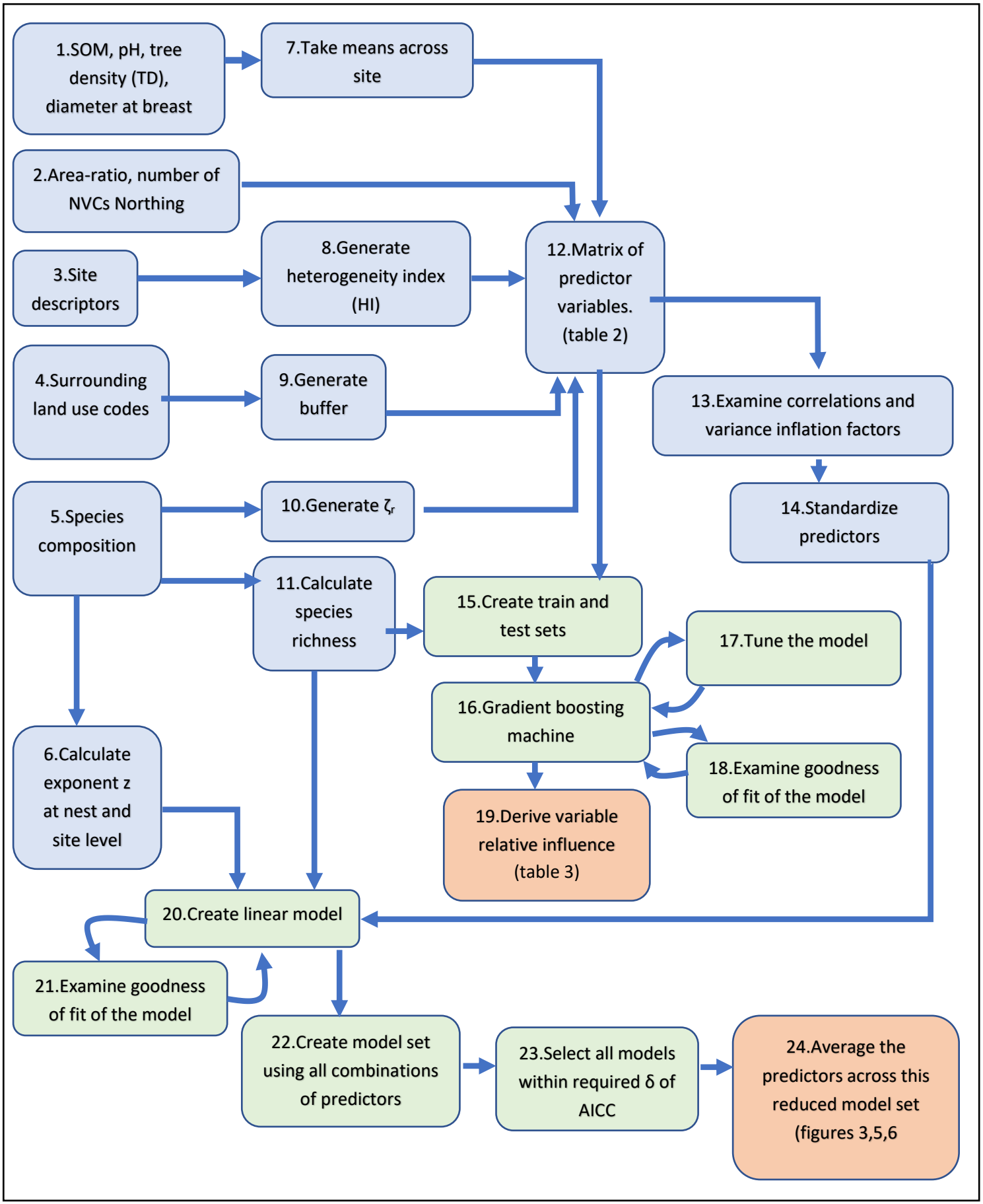

Figure 2. Steps involved in modelling species richness and z slopes using model averaging and GBM. Blue = model inputs, Green = quantitative modelling steps, Orange = model outputs. This diagram summarises the standard processes followed when carrying out statistical modelling using model selection and model averaging or decision trees.

Once the predictor variables were selected the dataset was standardized (step 14 figure 2) as suggested by Gelman, (2008). The input predictors are standardized such that two standard deviations of change in the predictor variable will result in the effect size change in the response. A linear model was created using all the variables (step 20 figure 2). The dredge function in the MuMin package in R (Barton, 2018) was used to create a set of models containing all permutations of the variable set (step 22 figure 2). These were ranked according to the AIC (step 23 figure 2). The model with the lowest AIC is considered the model that best represents the data.

All models within an AIC of between two to seven can be considered as equally good (Burnham et al., 2010) and used to create a model set. The parameter effect sizes are then averaged across this model set and a 95% confidence interval obtained (step 24 figure 2). If this confidence interval does not cross zero, these predictors are considered significant, (Kimberley et al., 2014). In this work, models within AIC of 2 ($\delta < 2$) of the highest ranked model were initially selected. The effect of increasing δ was then examined. If more variables were introduced which had confidence intervals that crossed zero, the lower value was used to avoid including redundant models (Grueber et al., 2011). The relative importance of each effect size was also obtained, (Burnham, 2015). Relative importance is a measure of the probability that a variable is present in the model set, so the greater the importance of a parameter, the more likely it is to be found in a model. A relative importance of 1 means that a variable appears in every model in the model set.

GRADIENT BOOSTING MACHINES

In addition to model selection and model averaging, a gradient boosting machine was used. The benefit of using these two modelling methods is their ability to detect different responses. Model-averaging detects linear relationships. Decision tree methods such as GBMs may not be as successful at detecting linear relationships, but they can detect more complex patterns in the data, (James et al, 2013).

Again, the strengths of this method and the procedure for its application have been examined and described by other authors (Ath, 2007; Elith, Leathwick and Hastie, 2008; Natekin and Knoll, 2013)) and a brief summary will therefore only be given here.

Gradient Boosting Machines (GBMs) are ensemble decision tree algorithms where recursive binary splits divide the data into regions, such that the root mean square error (rmse) of the response variable is minimized, (James et al., 2013). In addition, GBMs weight the observations such that those which were difficult to classify after a split are given increased weight. The next tree is then generated using these weighted observations. GBMs are suited to analysing complex ecological datasets because they can fit non-linear relationships and are not sensitive to scale, (De'Ath & Fabricius, 2000; Elith et al.,

2008). They have been shown to be successful, for example, in finding environmental variables which explain species richness of vertebrate groups (Mouchet et al., 2015). In this analysis the gbm package in R was used, (Ridgeway, 2017).

The data was split into train and test sets using a 75/25 ratio. The GBM also uses a random subset of training data and so automatically provides a validation set. The values that are not selected can be used to give an "out of bag" (OOB) prediction error for the model.

Hyper parameters define the structure of the GBM and must be chosen to optimise the model performance. This was achieved through a hyper parameter tuning grid; a matrix containing a set of permutations of hyper parameters. The model was run across the hyper parameter tuning grid and the hyper parameters which minimized the OOB error were used in the final model.

Over fitting was addressed using the ntrees function of the gbm library. This function looks at the prediction error on the OOB set, and the number of trees is set as the value when this error stops decreasing and starts to increase. The rmse error of the train was half that of the test set, suggesting some overfitting. However, the test set rmse was comparable to that obtained using model-averaging.

To extract the variables which most influence species richness the relative influence (RI) of each variable was used, (Natekin & Knoll, 2013). The percentage decrease in rmse is averaged for each variable over each time a variable was chosen to make a binary split in the data. The variable which has greatest RI is best at splitting the data and predicting the response.

Decision tree methods such as GBMs sometimes do not give stable results for RI, (Nicodemus et al., 2007). Therefore, each model was repeated 100 times and the variables ranked by the average values of RI. The value of RI is relative and therefore subjective. In this work variables which had RI above 10% were selected as important for species richness.

SCALE DEPENDENCE OF SPECIES RICHNESS – SPECIES AREA CURVES

We calculated species area curves across the nests from $4m^2$ to $200m^2$ and across sites, from $200m^2$ to $3200m^2$. The former curves correspond to type I nested curves as described by Scheiner (2003). In order to derive z at this smaller scale the species were accumulated across the five nests in each woodland with plot as a random effect.

At the larger scale, species were accumulated across the plots from $200m^2$ to $3200m^2$, as shown in figure 3. At this scale, species can be accumulated in 16! alternative ways i.e. every sequential arrangement of each of the 16 plots. The average of a subset of these accumulation curves is often used to calculate the slope of the species area curve. However, there are potential problems with this

method when the area over which the accumulation occurs is heterogeneous, and spatial autocorrelation should be taken into account, (Bacaro et al, 2012; Bevilacqua et al, 2018, Chiarucci et al, 2009). This averaging method is effectively selecting all species for accumulation as if they have equal probabilities of occurring in any plot in the woodland, i.e. the woodland is homogeneous. This is not the case for this data, the species are grouped into assemblages. Therefore, we took a different approach to maintain this spatial information.

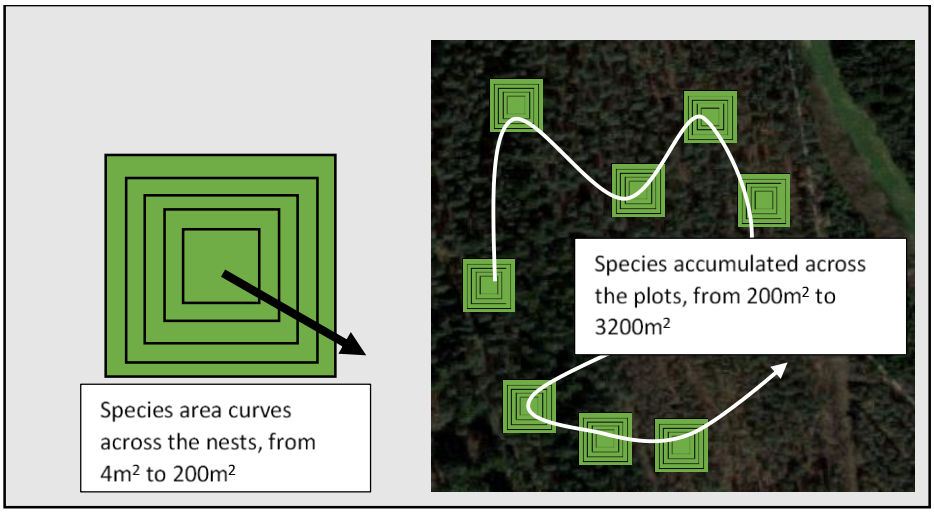

Figure 3. Two scales of species area dependence. At the within-plot scale species area curves were calculated for areas from 4m^{2}to 200m^2. At the between-plot scale species were accumulated across the 16 plots such that area increased from 200m^2 to 3200m$^{2.}$

To overcome this in a simple to execute algorithm, two curves were created. First, species were accumulated in such a way as to give a minimum gradient to the species accumulation curve. To do this, the richest plot is selected as a starting plot. The plot which shares the most species with the starting plot is accumulated next, and so on. This is equivalent to grouping all similar species assemblages together. Secondly, species were accumulated to give a maximum gradient. For this curve, the most species poor plot is selected first. The second plot to be accumulated is the one that has the least number of shared species with the starting plot. This is equivalent to arranging the plots in the most heterogeneous manner. Since these two cases represent two extremes in the woodland (the most homogenous and the most heterogeneous) an average was taken of the two resulting curves. Whilst this average does have the effect of homogenising the woodlands to some extent, we expect that this is not as extreme as averaging over all possible accumulation curves and will preserve some spatial heterogeneity.

This spatially explicit method is similar to a type III curve, as described by Scheiner (2003), where mean diversities of adjacent plots are calculated, but here adjacency is in terms of species assemblage's

similarity or dissimilarity. Using this method to fit log(S) to log(A) we found that 94% of the variance in log(S) was explained for three quarters of the woodlands.

At both scales the slope parameters of the log/log transform of the curves were used as the response variable z.

These linearized slope coefficients from the species area curves were then subject to explanatory modelling with the same variables used to model species richness, as listed in table 1.

Results

SPECIES RICHNESS OF FIXED AREA OF 3200M^2

Figure 4 shows the model-averaged effect sizes and 95% confidence intervals for the total richness (λ diversity) accumulated across all the 16 200m^2 plots in each woodland. Area of semi-natural habitat around each site (Buffer), number of NVC codes (no_NVC), Heterogenity Index (HI) and ζ_r all had a significant positive effect on species richness with similar effect sizes. Mean diameter at breast height (meandbh) and mean SOM had a negative effect, again, with similar effect size. All significant variables were included in every model in the model set except ζ_r which shows an 80% probability of inclusion. The effect sizes of all significant variables were similar, with no one parameter clearly having a greater influence on richness than any other.

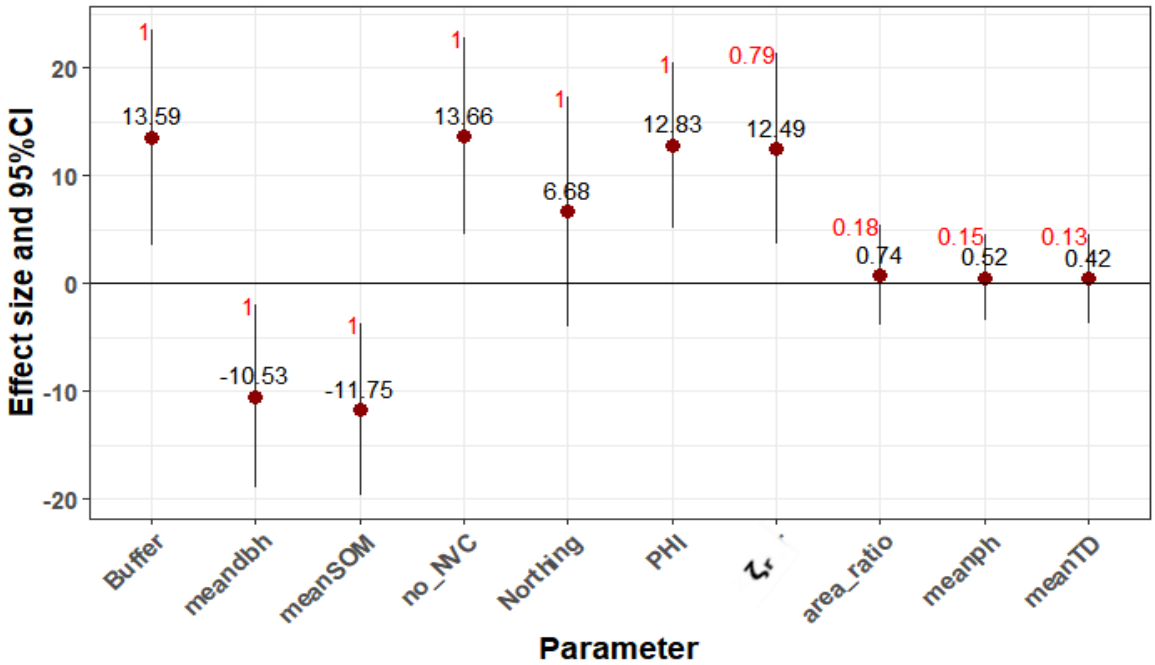

Figure 4. Model averaged effect sizes for woodland richness modelled with all variables using δ < 2 for the model set. Variable importance is in red. The points show the average effect size and the bars show the 95% confidence intervals.

Table 2 shows the results from the modelling using the GBM to find the variables which have the most influence on species richness. Parameters with relative influence of 10 or above have been selected as being important for predicting richness. ζ_r and Heterogeneity Index (HI) are again shown as having an impact on species richness. The GBM also selected mean LBA, mean pH and Northing.

Together, model-averaging and the GBM suggest that nine variables affect species richness of the woodlands, were woodland is defined as the 3200m^s covered by the area of the 16 plots. Three of these represent heterogeneity: ζr, number of NVC codes and HI. Of these, number of NVC codes and HI are present in every model in the model set and ζr is present in 80% of the models. The GBM selects ζ_r as the most influential variable.

Table 2. Most influential variables from the gradient boosted machine

Parameter	Relative influence
ζ	16.37
Northing	13.25
meanLBA	11.42
PHI	10.88
mean pH	9.92

SCALE DEPENDENCE OF SPECIES RICHNESS

Figure 5 below shows the results of calculations to compute the exponent, z, at the two scales of this work.

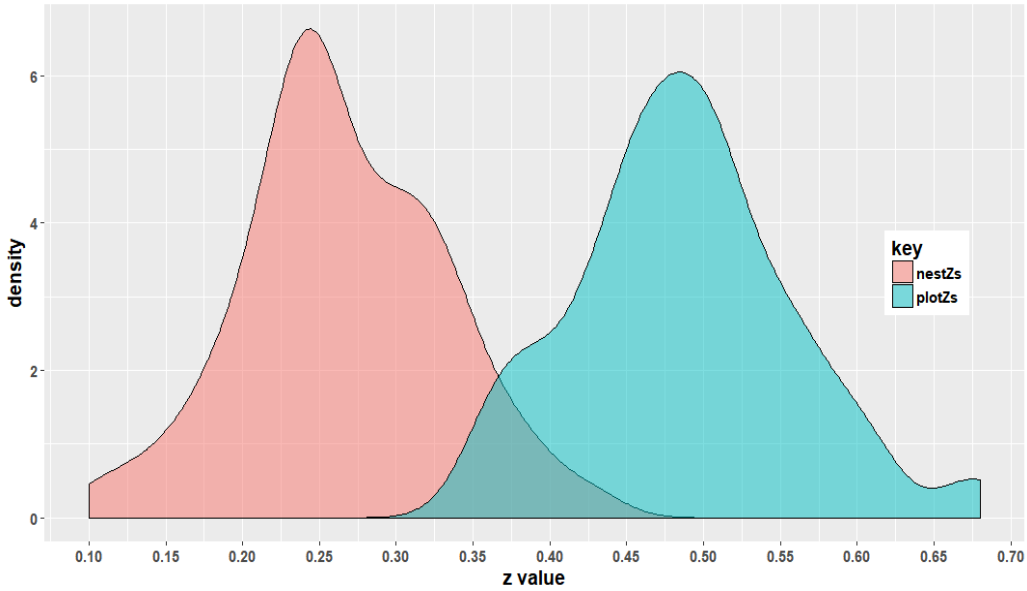

Figure 5. Distribution of z values at two scales. The pink distribution shows within-plot scale – 4m^2 to 200m ^{2}. The blue distribution shows between-plot scale – 200m^2 to 3200m^2

At the smaller within-plot scale the exponents are distributed around 0.25, whilst at the larger, between-plot scale they are distributed around 0.5.

Figure 6 shows the model-average effect sizes for (within-plot) nest z. The site level heterogeneity parameters HI and ζ_r have a positive effect on the value of z while mean SOM has a negative effect. The mean SOM is contained in every model in the model set, but the probability of inclusion of HI is 82% and for ζ_r only 57%, this implies that ζ_r is not an important parameter at nest scale. HI is the most significant parameter, the effect size double that of mean SOM.

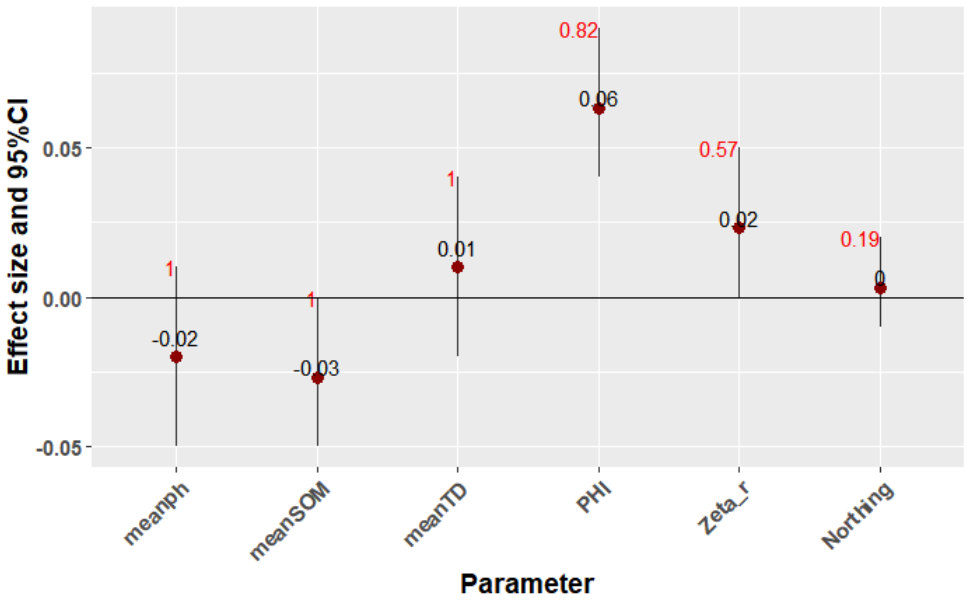

Figure 6. Model averaged effect sizes for nest z modelled with all variables using δ < 1.5 for the model set.

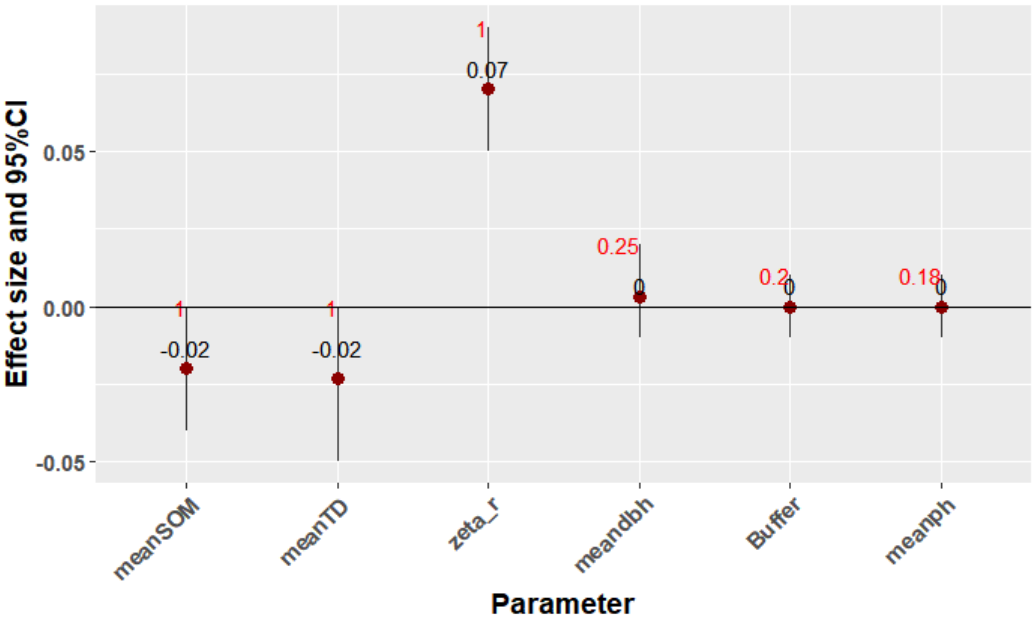

Figure 7. Model averaged effect sizes for plot z modelled with all variables using δ < 1.5 for the model set.

Figure 7 shows the model-average effect sizes for the slope of the between-plot specie area curve z. ζ_r is highly significant and has a positive effect on plot z. At this larger scale the effect of ζ_r, which quantifies species turnover across the woodland, has much greater significance than at the smaller nest scale. Mean tree density and mean SOM are also significant, although less so than ζ_r and their effect sizes are less than a third of ζ_r. Both mean SOM and mean tree density have a negative effect on plot z.

Discussion

VARIABLES AFFECTING SPECIES RICHNESS

We would expect that a suite of factors significant influence α, β and λ diversity in each woodland site. Our results indicate that the strongest effect is attributable to the heterogeneity of abiotic conditions at plot and site scale. We discuss the importance of heterogeneity and other influences below.

HETEROGENEITY

The greater the HI, the more this indicates evidence of woodland management, open areas and riparian zones. This implies that all these features are associated with elevated plant species richness. The number of NVC codes and ζr quantify the number of different species assemblages. Different species assemblages are more likely to occur if environmental conditions differ across the woodland – for example, different levels of soil moisture or different light levels, again stressing the importance of heterogeneity within the woodlands. Moreover, it appears that increased site heterogeneity not only leads to increased gamma diversity, as seen by increased site species richness, but also increased beta diversity, as seen by the dependence of nest z on HI.

MEAN-DBH AND MEAN-LBA

Model-averaging shows mean DBH to have a negative effect on richness and the GBM selected mean LBA as influential. This is to be expected, as an increase in either variable implies succession in the woodland (Packham et al., 2001; Smart et al., 2014) which leads to increased homogeneity because of a strengthening of the shading filter on the understorey species pool (Keith et al., 2009), again indicating the importance of woodland management schemes as well as sufficient area for natural disturbance dynamics to vary the successional status of the woodland.

AREA OF SEMI-NATURAL HABITATS AROUND THE SITE (BUFFER)

The amount of semi-natural habitats around the woodlands (buffer) had a significant positive effect on richness and occurred in every model in the model set. The size of this buffer relates to the amount of the woodland that is exposed to natural habitats. The greater that value, the more protected the woodland is from agrochemicals and airborne environmental pollutants; particularly around the edges (Draaijers, et al., 1988), which are linked to reduced species richness (Bobbink et al., 1998; Pallet et al., 2016, Simkin et al., 2016; Stevens et al., 2016). Previous research has suggested that buffer regions should be included around woodlands for various reasons, including protection from anthropogenic influences (McWilliam et al., 2010) and vehicle pollution (Bignal et al., 2008). A larger the buffer may also imply a more connected the woodland; habitat fragmentation is known to have a negative impact on biodiversity, (Quine, 2011, Watts et al 2008). It is not possible within this dataset to ascertain the mechanism for the increase in richness with buffer, which might not be metapopulation buffering and reduction of extinction probability but could be because a high amount of seminatural habitat is a surrogate for the woodland being on a substrate and in a climatic space that favours a high gamma diversity.

NORTHING

Northing is selected as an influential variable by the GBM. This would appear to be in contradiction to previous authors, (Gilman et al., 2014; Ohlemuller & Wilson, 2000, Gaston, K.J, 2000) who suggest that species richness would be expected decrease with latitude. However, the large-scale effects which are behind the generally accepted increase in species richness toward the equator are unlikely to be seen across the small change in latitude of our study, and also are likely to be dominated by other factors. For example, a bivariate plot of Northing with mean TD (figure 8) shows a negative correlation, suggesting that the more northerly woods are less dense and hence the increased richness could be due to the addition of more light-demanding species.

The Scottish woodlands might also consist of plant communities which tend to be richer. For example, if Scottish woodlands were predominantly W6 eutrophic Alder woodlands, they would be richer than W14 beech woods which are common in the South. However, figure 9 shows that this is not the case. The distribution of NVC communities is similar in the North and South. What is noticeable, however, is that for the same community, richness can be greater in Scottish woodlands, (figure 10). This suggests that other variables, not availabe in this dataset, may be responsible for the increased richness in Scottish woodlands, or that a combination of positive predictors is more likely to co-occur in the more Northern woodlands.

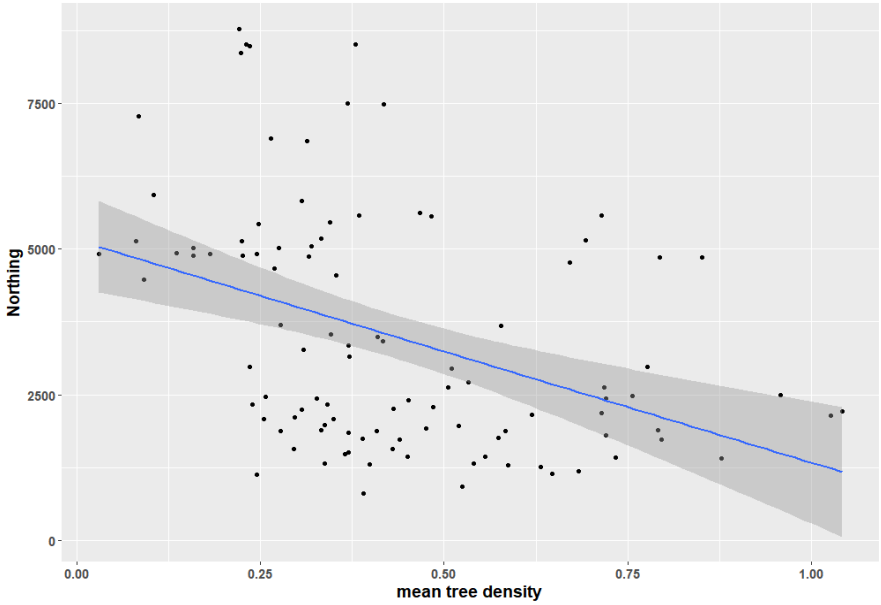

Figure 8. Scatter plot of northing with mean tree density suggests that the northern woodlands are more open. This could allow the inclusion of more light-loving species assemblages and hence add to species richness.

Distribution of NVC codes in English and Welsh woodlands

Distribution of NVC codes in Scottish woodlands

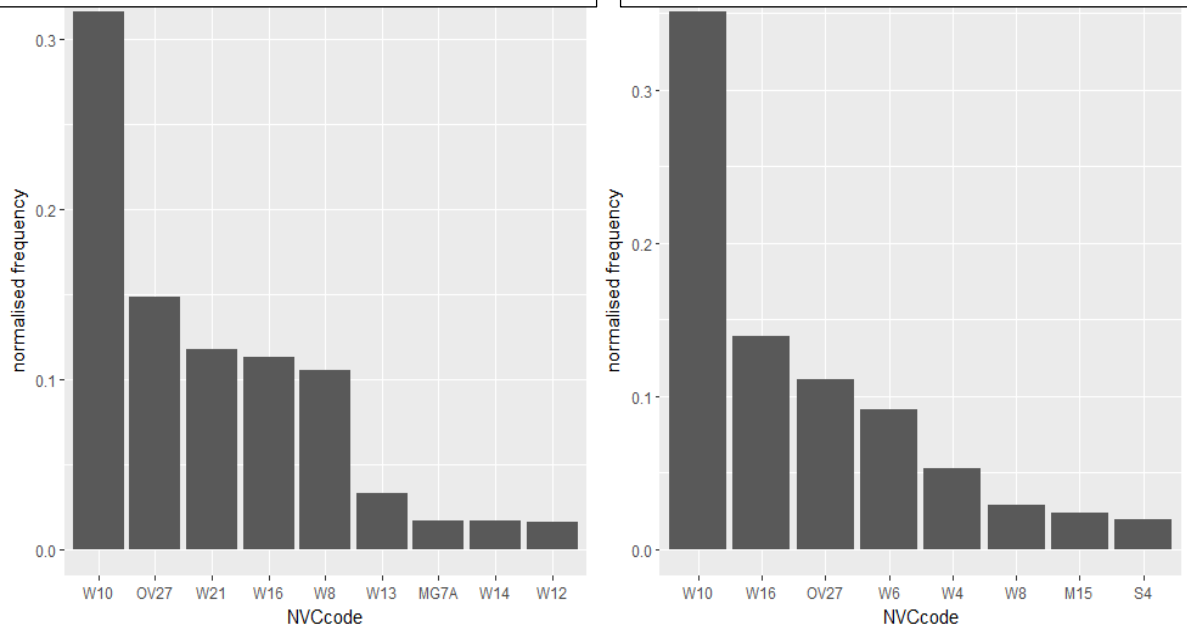

Figure 9. Distribution of NVC codes in Scotland and England and Wales. In both areas 30% of communities are W10, Quercus robur woodlands. These communities have one of the widest range of species richnesses of all NVC communities in this dataset.

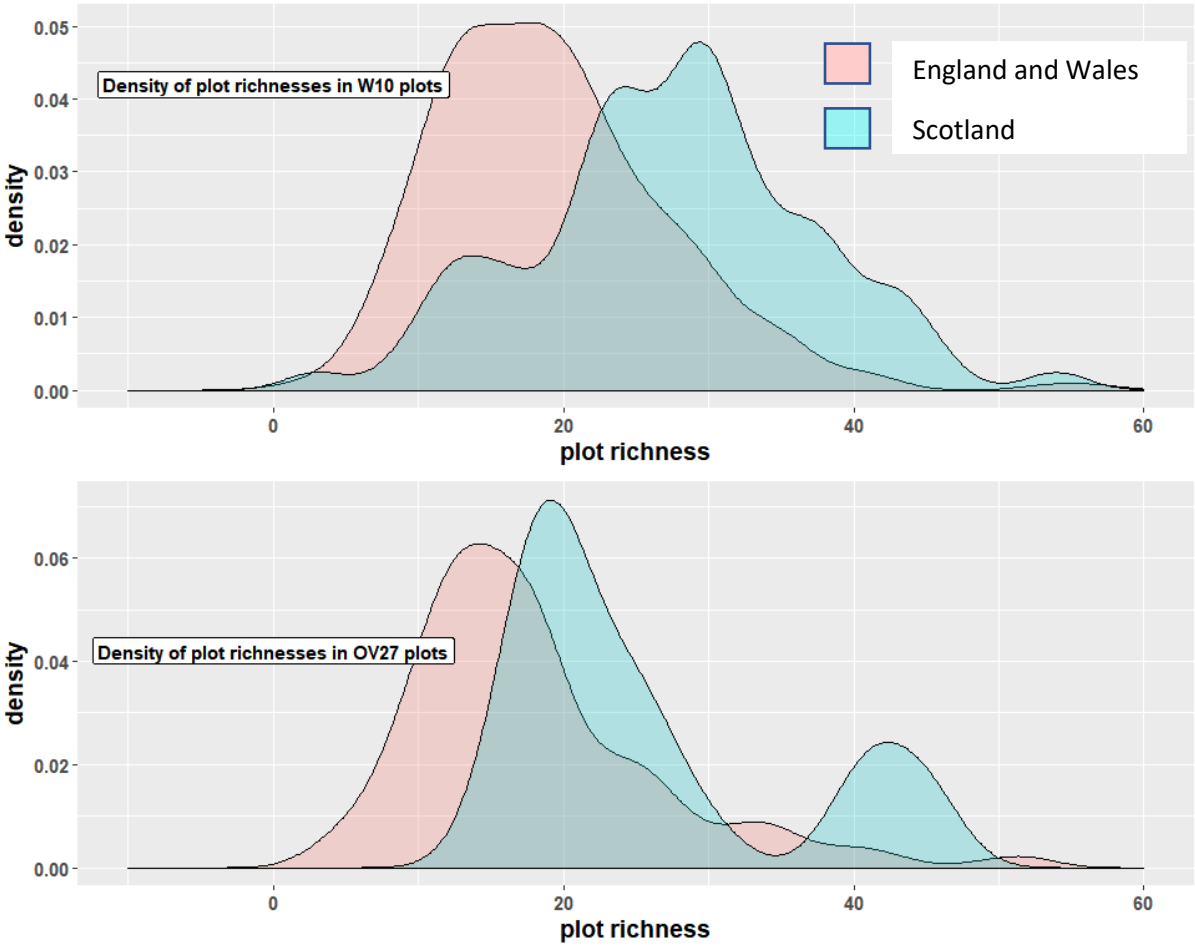

Figure 10. Density of plot richness for the most common NVC codes in the dataset in Scotland and England and Wales. Given the same NVC community, richness tends to be greater in Scotland for these common communities. Although OV27 is not a woodland community, it was was common in this dataset, demonstrating open areas within the wood.

SOIL ORGANIC MATTER

SOM content was shown to have a significant and important negative effect on species richness in agreement with Koojiman & Cammeraat (2010). Organic soils have low pH and low pH soils in woodlands results in low species richness, (Cornwell & Grubb, 2003). However, in this data, very few of the plots were located on organic soils. The six NVC codes shown in table 4 represent 82% of all plots surveyed and give an indication of the majority of soil types.

Table 4. Soil types for the most common NVC codes in this data

NVC code	% of plots	Soil type
W6 *Alnus glutinosa-Urtica diocica*	5%	Eutrophic moist soil alluvial or on flood fen peat
W8 *Fraxinus excelsior – Acer campestre – Mercurialis perennis*	9.5%	Dry calcareous soils
W21 *Crataegus monogyna – Hedera helix* scrub	10%	Various, base rich
W16 *Quercus* spp. – *Betula* spp. – *Deschampsia flexuosa*	11.5%	Very acidic, oligotrophic, pH<4, sandy podzolic
OV27 *Epilobium angustifolium* Open habitat	14%	Various, fertile, damp
W10 *Quercus robur – Pteridium aquilinum – Rubus fruticosus*	32%	Neutral brown earths

Of the most common NVC types in this dataset, only W6, 5% of the total, are potentially on peat, but W6 is a high nutrient, high richness community. It is therefore unlikely that the reduced richness with SOM is caused by reduced richness expected on organic soils.

The reduced richness could be due to low pH on other soil types since increased SOM has been shown to decrease pH (Russell, 1960; Williams & McDonald, 1960). On the other hand, adding SOM can also increase pH, depending on the initial acidity of the soil and the nature of the litter (Ritchie & Dolling, 1985). In this analysis, increased SOM appears to correlate with either decreased or no change in pH (figure 11).

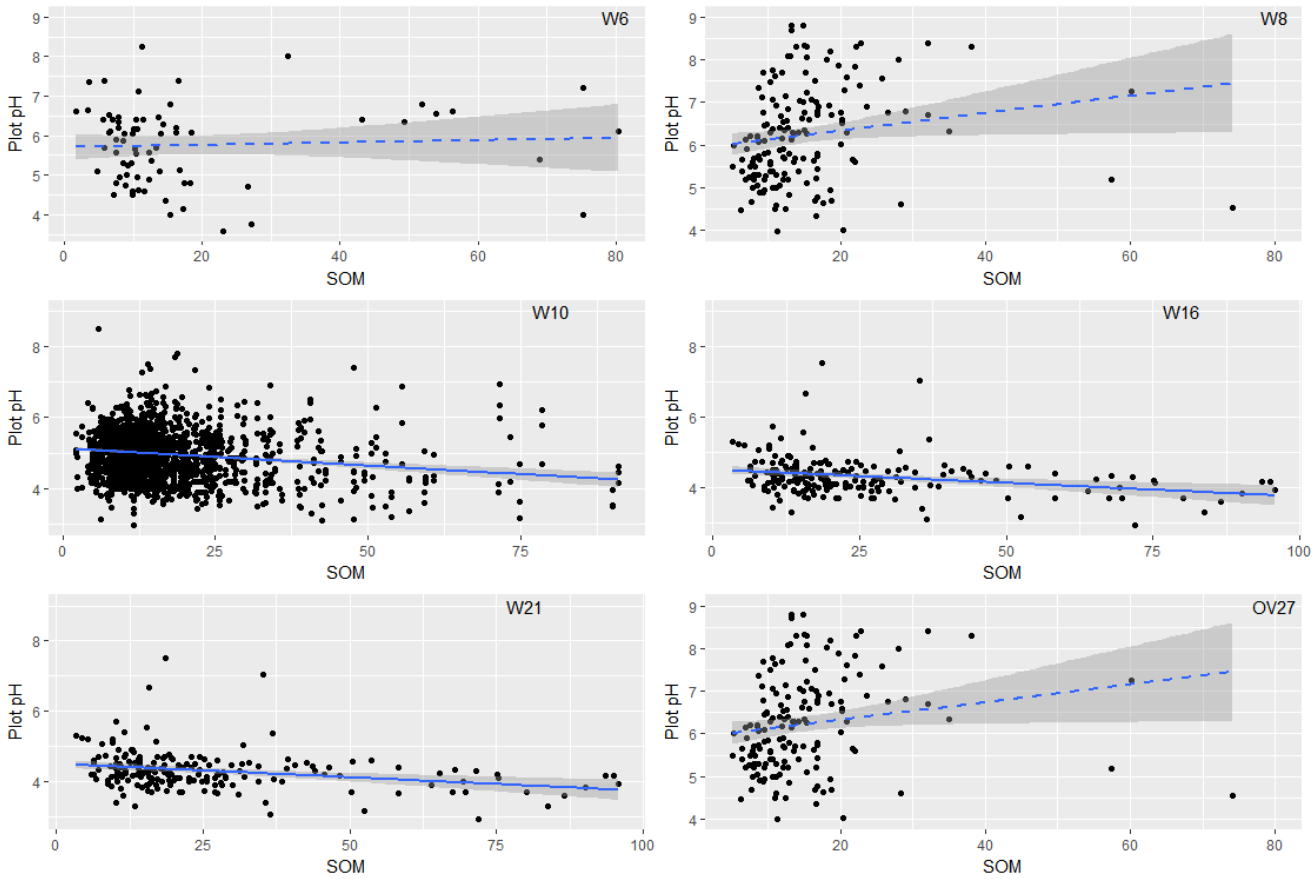

Figure 11. Change in plot pH with SOM content. For W10, W16 and W21 plots the pH reduces with increased SOM but this effect is not clear for other plot types

Figure 12 shows that all the common NVC communities tend to show reduced richness as SOM increases, regardless of soil type. Given also that the NVC code represents a community with similar physical conditions, such as nutrient levels, soil pH and light levels, and while the response to richness with increased SOM is seen across all communities, this finding implies that the variation in richness with SOM is unlikely to be due solely to one of those physical conditions, but perhaps to another factor not measured in this dataset.

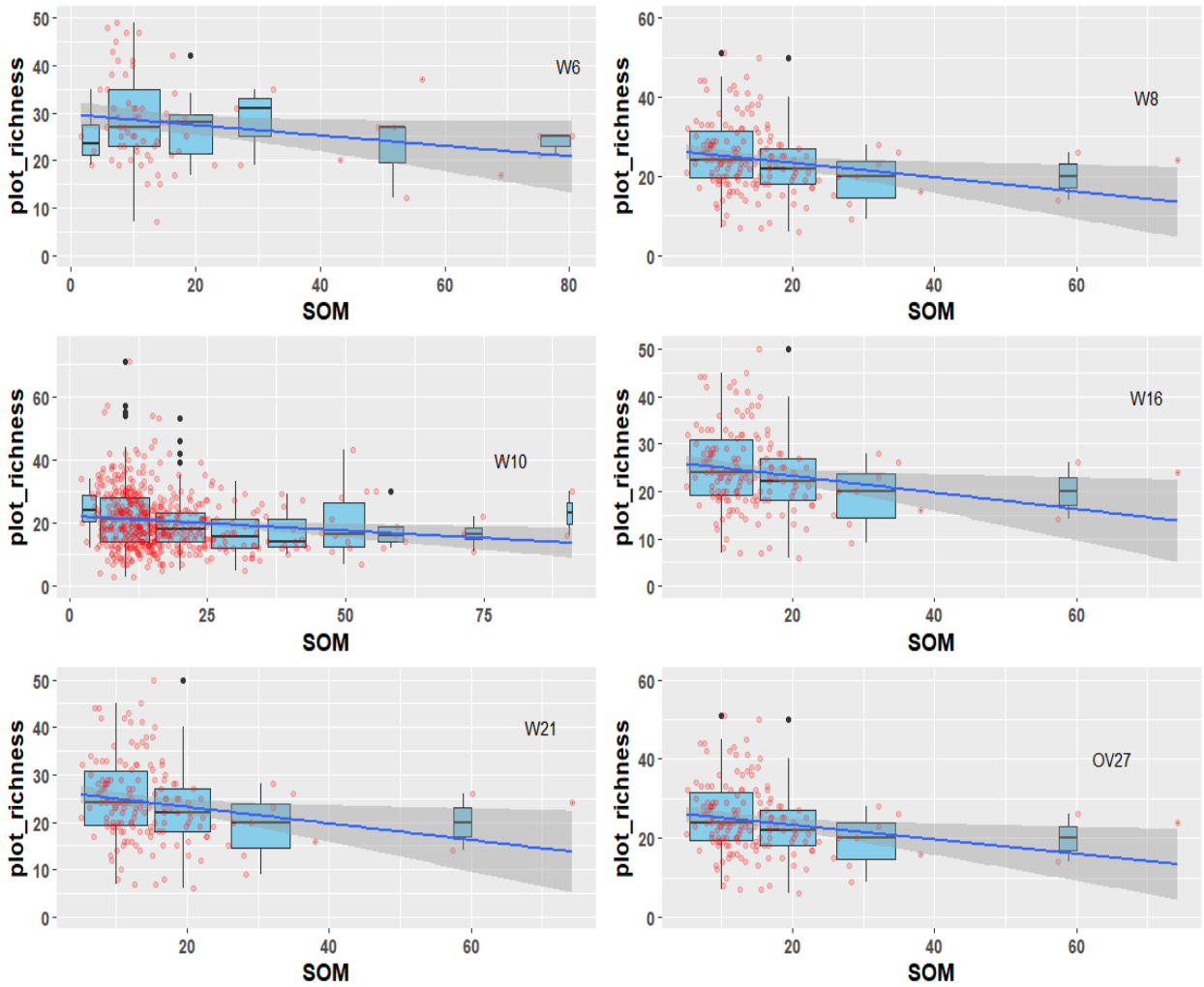

Figure 12. Plot richness with SOM for the six most common NVC communities in this data. The blue boxplots show the range in richness when grouped, the red dots are the raw ungrouped data. The variation in richness is high when SOM is low and vice versa.

Increased nitrogen deposition can both increase SOM (Fang et al., 2014; Zak et al., 2016) and reduce richness (Bobbink et al., 1998; Pitcairn et al., 2002; Simkin et al., 2016; Stevens et al., 2016). A smaller buffer could allow increased nitrogen deposition, (Draaijers et al., 1988; Kennedy & Pitman, 2004), which might be reflected in the Ellenberg N values (Maskell et al., 2010). Therefore, if woodlands with smaller buffers showed both increased Ellenberg N values and increased SOM, nitrogen deposition might be the mechanism behind the change in SOM and the reduced richness. In support of this figure 13 shows that buffer is negatively correlated with Ellenberg N values, but, in contradiction, it is positively correlated with SOM content. The latter association may be confused by the location of plots within the woodland.

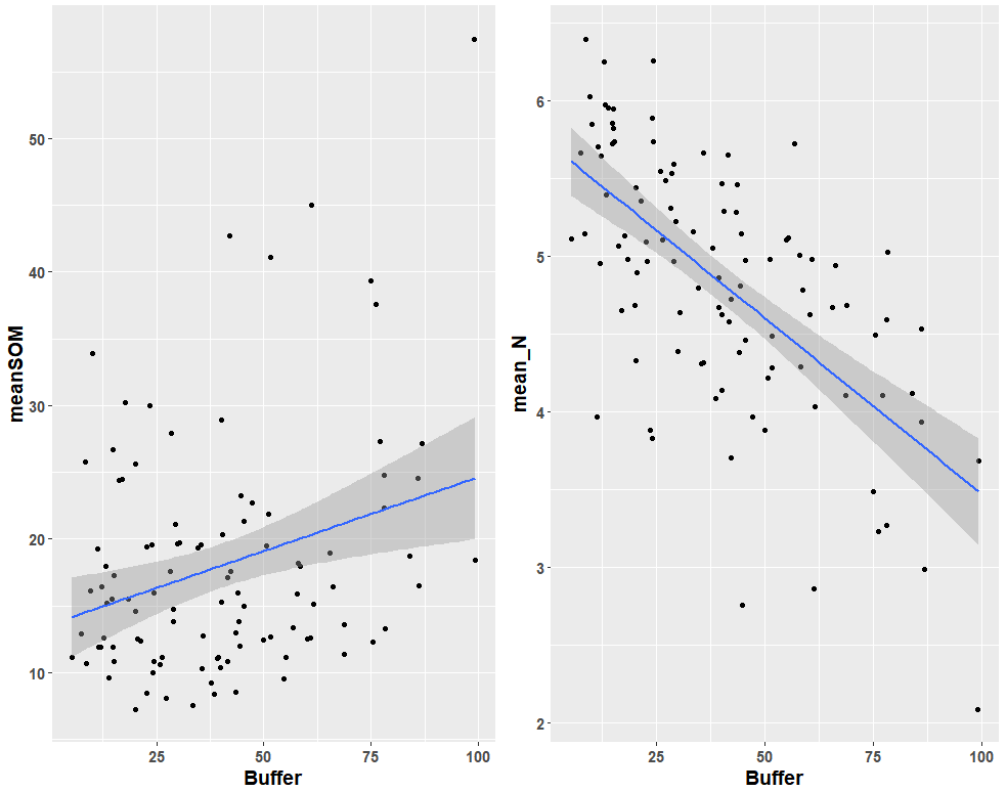

Figure 13. Change in mean SOM and Ellenberg N values with Buffer size. If nitrogen deposition was leading to increased SOM, then woodlands with smaller buffers would have greater SOM and larger Ellenberg N. The graph on the right shows that smaller buffers correlates with higher Ellenberg N, but not to greater mean SOM.

In summary, whilst SOM is negatively correlated with species richness, the causal mechanism is not evident. However, other authors, using modelled estimates of nitrogen deposition rather than Ellenberg N values with this dataset have shown positive correlation between SOM and modelled nitrogen deposition, (Kirby, 2005). It is also interesting to note that nitrogen deposition values were on average lower in Scotland (RoTaP, 2012), which may be another factor contributing to the increased richness of the more Northern woodlands.

SOIL pH

The GBM selected mean pH as important for predicting species richness. Figure 14 demonstrates that in this data plot richness does have a unimodal response to plot pH, in agreement with Cornwell & Grubb, (2003)

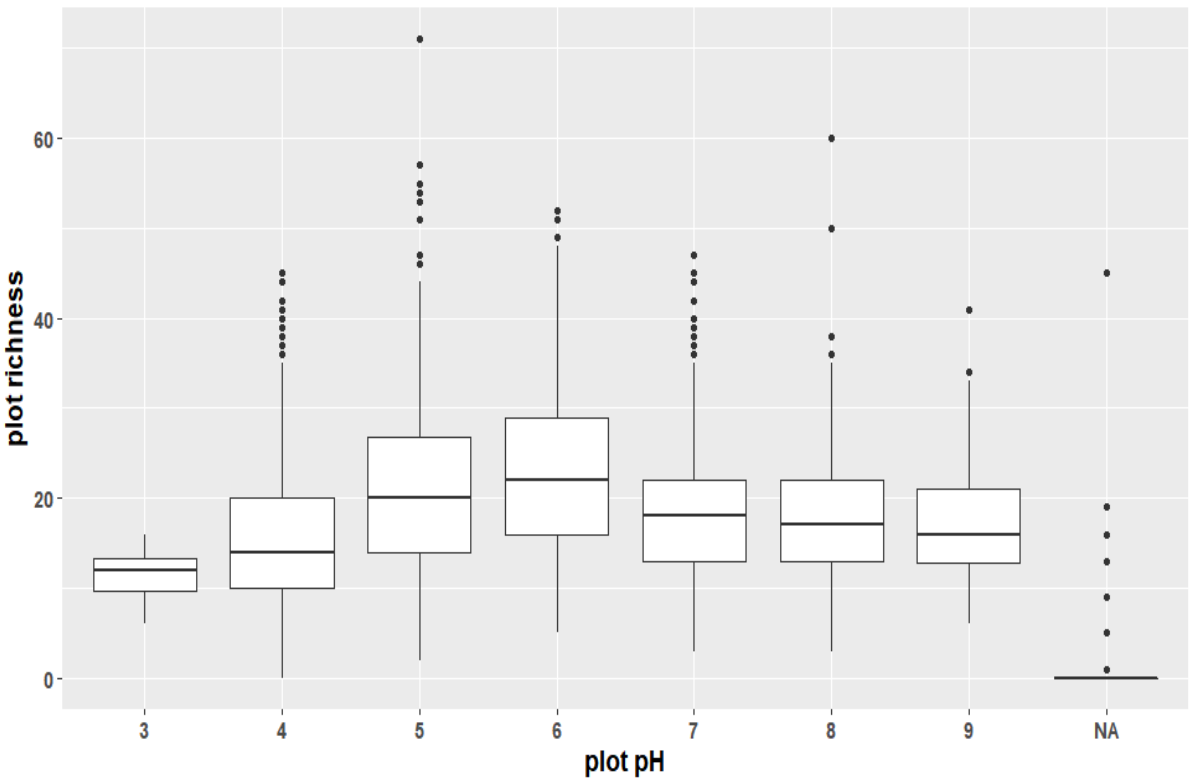

Figure 14. Plot richness response to pH can be seen to be unimodal with a peak around pH6.

EFFECT OF DRIVERS OF SPECIES RICHNESS ON GRADIENT OF SPECIES AREA CURVE AT TWO SCALES.

The fitted nest *z* values were distributed around 0.25. This agrees with the value of z predicted by other authors (Preston, 1962, Connor et al., 1983; Sugihara, 1976). At the larger scale, plot z was found to be distributed around 0.5. These two values correspond to those found in a similar study at two different scales by Crawley & Harral (2001) for vascular plants in the county of Berkshire, South East England. These authors comment that the smaller z value is expected at square meter scales, whereas the larger z value reflects large changes in species richness as new habitats are added when species are accumulated at larger scales.

At smaller scales, HI has a significant positive effect on nest z while SOM has a negative effect. This model suggests that biological variables appear to be influencing the slope of the species area curves, in contradiction to the remarks of other authors (Connor et al., 1983; Sugihara, 1976; Tjorve & Tjorve, 2017), but this is not necessarily the case. Both Sugihara (1979) and Tjørve & Tjørve (2017) point to the role of community structure and aggregation in shaping the species area curve. HI, by introducing elements into the woodland that increase plant species richness, is likely, on average, to be increasing the richness of each plot. This will imply a more diverse community over which the curve is

constructed. Similarly, SOM can have a homogenizing effect on the plots due to loss of species richness, either directly (Kooijmann & Cammeraat, 2010) or because increased SOM indicates increased nitrogen deposition and eutrophication that favours the abundance of competitive species potentially suppressing species diversity in the understorey, (Maskell et al, 2010, Pallet et al 2016). Both HI and SOM can be seen therefore, as influencing community structure either positively or negatively and so effecting nest z. It should also be noted that model-averaged R^2 for nest z was only 0.31, and that of the full model 0.33 so neither of the variables explained substantial variance in nest z.

At larger scales ζ_r has a significant positive effect on plot z, and mean SOM and mean TD have a significant negative effect. ζ_r is significant at the larger scale because it represents species turnover across the woodlands. The greater the species turnover, the greater the expected slope. Similarly, increasing mean TD and mean SOM are likely to have a homogenising effect on the woodland and therefore, to reduce plot z. At the larger scale, it is species turnover and heterogeneity that has been shown to influence the exponent of the species accumulation curves rather than biological factors. At this scale the model was more successful at explaining the variance in plot z, having R^2 of 0.44 for the model-averaged model and 0.47 for the full model.

VARIABLES NOT INCLUDED IN THIS ANALYSIS

Statistical modelling using model averaging explained 47% of the variance in plant richness. Since we chose only those variables that we both knew were likely to affect richness and that were available within the dataset, several important potential variables were omitted which may improve the explanatory power of the model. Woodland age and past land use have been shown to affect species composition (Hermy, 1994) and richness (Dzwonko & Loster, 1989; Peterken & Game 1984). Ancient woodlands are more likely to contain long-lived, poorly dispersing perennials whose migration to new woodland sites is limited (Kimberley, Blackburn & Whyatt, 2015). This constrains the potential richness of new woodlands, particularly if they are isolated. Ancient cultivation has also been shown to induce a species compositional gradient and thereby increase richness (Dupouey et al., 2002).

Plant available phosphorus has been shown to reduce species richness of native plants in alder forests (Hrivnak et al., 2015). Phosphorus eutrophication favours competitive species particularly at higher light levels and this can result in the exclusion of other species (Keersmaeker et al., 2004).

The amount of light reaching the woodland floor influences richness, with greater richness at higher light levels, (Kenedy & Pitman, 2004). However, eutrophication together with high light levels encourages shade-intolerant, competitive species which tend to reduce richness, (Cornwell & Grubb,

2003, Keersmaeker et al., 2004). Conversely, low pH in conjunction with low light levels can result in very little ground cover, (Cornwell & Grubb, 2003)

USE OF ζ_r AS HETEROGENEITY METRIC

Results presented here suggest that ζr is a useful way to quantify species turnover. This variable was selected as the most influential by the GBM and was significant when used in model-averaging. The fact that ζ_r was a significant parameter when modelling the slope of species accumulation curves across the wood, but not when modelling the slope of species area curves across plots, demonstrates that it quantifies species turnover. This metric could be easily generated from species lists as an indicator to woodland managers of the number of different habitats within the woodlands.

SYNTHESIS

Habitat heterogeneity was found to be important for species richness in UK broadleaved woodlands. Heterogeneity in these models included coppicing, riparian zones and open areas, suggesting that all these factors need to be considered if woodland plant diversity is to be increased. Isolation was negatively correlated with species richness and therefore any possibility to increase the size and connectivity of new woodlands should be a priority. The northern woodlands were found to be richer than those in the south, and more importantly, in some cases the same plant communities were richer in Scotland than the south. The reason for this was unclear and further work could be done, for example, examining interactions or nitrogen deposition data.

The parameter effect sizes for all the variables were similar, so that no one factor can be said to be most important for richness. This implies that in order to increase woodland richness all must be addressed simultaneously, and that one single factor will neither explain the species richness or lack of diversity in our woodlands.

References

Amar, Arjun & M Hewson, C & M Thewlis, R & Smith, Ken & Fuller, Robert & Lindsell, Jeremy & Conway, Greg & Butler, Simon & A Macdonald, M. (2006). What's Happening to Our Woodland Birds?

Bacaro, G. *et al.* (2012) 'The spatial domain matters: Spatially constrained species rarefaction in a Free and Open Source environment', *Ecological Complexity*. doi: 10.1016/j.ecocom.2012.05.007.

Báldi, A. (2008) Habitat heterogeneity overrides the species–area relationship. *Journal of Biogeography*, 35 (4), 675–681.

Barton, K., (2018). MuMIn: Multi-Model Inference. R package version 1.42.1. https://CRAN.R-project.org/package=MuMIn

Baselga, A., Jimenez-Valverde, A., Niccolini, G. (2007) A multiple-site similarity measure independent of richness. *Biology Letters*, 3 (6), 642–645.

Baselga, A. (2013) Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. *Ecography*, 36 (2), 124–128.

Baude, MJ *et al* (2016) Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature* **530**, 85-88.

Bevilacqua, S. *et al.* (2018) 'An approach based on the total-species accumulation curve and higher taxon richness to estimate realistic upper limits in regional species richness', *Ecology and Evolution*, 8(1), pp. 405–415. doi: 10.1002/ece3.3570.

Bignal, K.L., Ashmore, M.R., Headley, A.D. (2008) Effects of air pollution from road transport on growth and physiology of six transplanted bryophyte species. *Environmental Pollution*, 156 (2), 332–340.

Bobbink, R., Hornung, M. & Roelofs, J.G.M. (1998) The effects of air-borne nitrogen pollutants on species diversity in natural and semi-natural European vegetation. *Journal of Ecology*, 86 (5), 717–738.

Boch, S., Prati, D., Müller, J., Socher, S., et al. (2013) High plant species richness indicates management-related disturbances rather than the conservation status of forests. *Basic and Applied Ecology*, 14 (6), 496–505.

Brainard, J., Bateman, I. J. and Lovett, A. A. (2009) 'The social value of carbon sequestered in Great Britain's woodlands', Ecological Economics. Elsevier B.V., 68(4), pp. 1257–1267. doi: 10.1016/j.ecolecon.2008.08.021.

Brudvig, L.A., Damschen, E.I., Tewksbury, J.J., Haddad, N.M., et al. (2009) Landscape connectivity promotes plant biodiversity spillover into non-target habitats. *Proceedings of the National Academy of Sciences,* 106 (23), 9328–9332.

Burnham, K.P., Anderson, D.R., Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer; 2010

Burnham, K.P., Anderson, D.R., Huyvaert, K.P. (2010) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65 (1), 23–35.

Burnham, K, P., (2015) Multimodel inference: understanding AIC relative variable importance values. Available from: https://sites.warnercnr.colostate.edu/anderson/wp-content/uploads/sites/26/2016/11/AICRelativeVariableImportanceWeights-Burnham.pdf. (Accessed 10th August 2018)

Butchart, S. H. M. *et al.* (2010) 'Global biodiversity: Indicators of recent declines', *Science*, 328(5982), pp. 1164–1168. doi: 10.1126/science.1187512.

Cencini M, Pigolotti S, Muñoz MA. What ecological factors shape species-area curves in neutral models? *PLoS One*. 2012;7(6):e38232. doi:10.1371/journal.pone.0038232

Chiarucci, A., Bacaro, G., Rocchini, D. & Fattorini, L. (2008) Discovering and rediscovering the sample-based rarefaction formula in the ecological literature. *Community Ecology*, 9 (1), 121–123.

Chiarucci, A., Bacaro, G., Rocchini, D., Ricotta, C., Palmer, M.W., Scheiner, S.M., (2009), Spatially constrained rarefaction: incorporating the autocorrelated structure of biological communities into sample-based rarefaction, *Community Ecology*, 10(2), pp209 – 214

Connor, E. F., McCoy, E. D. and Cosby, B. J. (1983) 'Model discrimination and expected slope values in species-area studies.', *American Naturalist*, 122(6), pp. 789–796. doi: 10.1086/284171.

Corney, P. M. *et al.* (2004) 'The effect of landscape-scale environmental drivers on the vegetation composition of British woodlands', *Biological Conservation*, 120(4), pp. 491–505. doi: 10.1016/j.biocon.2004.03.022.

Cornwell, W.K., Grubb, P.J. (2003) Regional and local patterns in plant species richness with respect to resource availability. *Oikos*, 100 (3), 417–428.

Crawley, M.J. (2001) Scale Dependence in Plant Biodiversity. *Science*, 291 (5505), 864–868.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., et al. (2007) Random Forests For Classification In Ecology. *Ecology*, 88 (11), 2783–2792.

De'ath, G., Fabricius, K.E. (2000) Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. *Ecology*, 81 (11), 3178.

Defra, (2018), The 25 Year Environment Plan, Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf, (Accessed 8th August 2018)

Draaijers, G., Ivens, W., Bleuten, W. (1988) Atmospheric deposition in forest edges measured by monitoring canopy throughfall. *Water, Air, and Soil Pollution*, 42 (1-2).

Dupouey, J.L., Dambrine, E., Laffite, J.D., Moares, C. (2002) Irreversible Impact of past Land Use on Forest Soils and Biodiversity. *Ecology*, 83 (11), 2978.

Dzwonko, Z. & Loster, S. (1989) Distribution of Vascular Plant Species in Small Woodlands on the Western Carpathian Foothills. *Oikos*, 56 (1), 77.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, 77 (4), 802–813.

Fang, H.J., Cheng, S.L., Yu, G.R., Yang, X.M., et al. (2014) Nitrogen deposition impacts on the amount and stability of soil organic matter in an alpine meadow ecosystem depend on the form and rate of applied nitrogen. *European Journal of Soil Science*, 65 (4), 510–519.

Freckleton, R.P. (2010) Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioural Ecology and Sociobiology*, 65 (1), 91–101

Friedman, J.H., Meulman, J.J., (2003) Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 22 (9), 1365–1381.

Fuller, R.J., Warren, M.S., (1993), Coppiced woodlands: their management for wildlife, Available from: http://jncc.defra.gov.uk/pdf/pubs93_Coppicedwoodlands.pdf (Accessed 8th August 2018)

Gardener, T. (2010), Monitoring forest biodiversity, Earthscan Publications,UK

Gaston, K.J., Spicer, J.I. (2004) Biodiversity: an introduction. Blackwell.

Gelman, A. (2008) Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine,* 27 (15), 2865–2873.

Gillman, L.N., Wright, S.D., Cusens, J., Mcbride, P.D., et al. (2014) Latitude, productivity and species richness. *Global Ecology and Biogeography,* 24 (1), 107–117.

Grueber, C.E., Nakagawa, S., Laws, R.J. & Jamieson, I.G. (2011) Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology*, 24 (4), 699–711.

Hayhow DB, Burns F, Eaton MA, Al Fulaij N, August TA, Babey L, Bacon L, Bingham C, Boswell J, Boughey KL, Brereton T, Brookman E, Brooks DR, Bullock DJ, Burke O, Collis M, Corbet L, Cornish N, De Massimi S, Densham J, Dunn E, Elliott S, Gent T, Godber J, Hamilton S, Havery S, Hawkins S, Henney J, Holmes K, Hutchinson N, Isaac NJB, Johns D, Macadam CR, Mathews F, Nicolet P, Noble DG, Outhwaite CL, Powney GD, Richardson P, Roy DB, Sims D, Smart S, Stevenson K, Stroud RA, Walker KJ, Webb JR, Webb TJ, Wynde R and Gregory RD (2016) State of Nature 2016. The State of Nature partnership.

Hermy, M. (1994) Effects of former land use on plant species diversity and pattern in European deciduous woodlands. *Biodiversity, Temperate Ecosystems, and Global Change*, pp123–144.

Honnay, O., Hermy, M., Coppin, P. (1999) Effects of area, age and diversity of forest patches in Belgium on plant species richness, and implications for conservation and reforestation*. Biological Conservation,* 87 (1), pp73–84.

Hrivnák, R., Slezák, M., Jarčuška, B., Jarolímek, I., et al. (2015) Native and Alien Plant Species Richness Response to Soil Nitrogen and Phosphorus in Temperate Floodplain and Swamp Forests. *Forests*. 6 pp 3501-3513

Hubbell, S.P. (1997) A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs*, 16 (5).

Hui, C. & Mcgeoch, M.A. (2014) Zeta Diversity as a Concept and Metric That Unifies Incidence-Based Biodiversity Patterns. *The American Naturalist,* 184 (5), pp684–694.

James, G., Witten, D., Hastie, T., Tibshirani, R., An Introduction to Statistical Learning: with Applications in R. New York, Springer Verlag; 2013.

Johnson, J. B. and Omland, K. S. (2004) 'Model selection in ecology and evolution', *Trends in Ecology and Evolution*, 19(2), pp. 101–108. doi: 10.1016/j.tree.2003.10.013.

Keersmaeker, L.D., Martens, L., Verheyen, K., Hermy, M., et al. (2004) Impact of soil fertility and insolation on diversity of herbaceous woodland species colonizing afforestations in Muizen forest (Belgium). *Forest Ecology and Management*, 188 (1-3), pp291–304.

Keith, S.A., Newton, A.C., Morecroft, M.D., Bealey, C.E., et al. (2009) Taxonomic homogenization of woodland plant communities over 70 years. *Proceedings of the Royal Society B: Biological Sciences*, 276 (1672), pp3539–3544.

Kennedy, F. &Pitman, R. (2004) Factors affecting the nitrogen status of soils and ground flora in Beech woodlands. *Forest Ecology and Management,* 198 (1-3), pp 1–14.

Kirby, K. J. *et al.* (2005) 'Long term ecological change in British woodland (1971–2001). A re-survey and analysis of change based on 103 sites in the Nature Conservancy "Bunce 1971" woodland survey.', *English Nature Research Reports*, 653(May 2014), p. 139.

Kirby, K.J.; Smart, S.M.; Black, H.I.J.; Bunce, R.G.H.; Corney, P.M.; Smithers, R.J.; Shaw, M.W. (2013). Woodlands survey flora data 1971-2001. NERC Environmental Information Data Centre. https://doi.org/10.5285/2d023ce9-6dbe-4b4f-a0cd-34768e1455ae

Kimberley, A., Blackburn, G.A., Whyatt, J.D., Kirby, K., et al. (2013) Identifying the trait syndromes of conservation indicator species: how distinct are British ancient woodland indicator plants from other woodland species? *Applied Vegetation Science*, 16 (4), pp667–675

Kimberley, A., Blackburn, G.A., Whyatt, J.D., Smart, S.M. (2014) Traits of plant communities in fragmented forests: the relative influence of habitat spatial configuration and local abiotic conditions. *Journal of Ecology*, 102 (3), pp 632–640.

Kimberley, A., Blackburn, G.A., Whyatt, J.D., Smart, S.M. (2015) How well is current plant trait composition predicted by modern and historical forest spatial configuration? *Ecography*,) 39 (1), pp67–76.

Kooijman, A.M. & Cammeraat, E. (2010) Biological control of beech and hornbeam affects species richness via changes in the organic layer, pH and soil moisture characteristics. *Functional Ecology*, 24 (2), pp469–477.

Latombe, G., Hui, C., Mcgeoch, M.A. (2017) Multi-site generalised dissimilarity modelling: using zeta diversity to differentiate drivers of turnover in rare and widespread species. *Methods in Ecology and Evolution*, 8 (4), pp431–442.

Latombe, G., McGeoch, M.A., Nipperess, D.A., Hui, C., (2018). zetadiv: Functions to Compute Compositional Turnover Using Zeta Diversity. R package version 1.1.1. https://CRAN.R-project.org/package=zetadiv

Lawton, J.H., Brotherton, P.N.M., Brown, V.K., Elphick, C., Fitter, A.H., Forshaw, J., Haddow, R.W., Hilborne, S., Leafe, R.N., Mace, G.M., Southgate, M.P., Sutherland, W.J., Tew, T.E., Varley, J., &

Wynne, G.R. (2010) Making Space for Nature: a review of England's wildlife sites and ecological network. Report to Defra.

Liaw, A., Wiener, M., (2002). Classification and Regression by randomForest. R News 2(3),18--22.

Marrs, R. H. *et al.* (2013) 'Native dominants in British woodland – a potential cause of reduced species-richness?', *New Journal of Botany*, 3(3), pp. 156–168. doi: 10.1179/2042349713y.0000000028.Maskell, L.C., Smart, S.M., Bullock, J.M., Thompson, K., Stevens, C.J., (2010) Nitrogen deposition causes widespread loss of species richness in British habitats, *Global Change in Biology*, 16, pp 671-679

McCulloch, J. S. G. and Robinson, M. (1993) 'History of forest hydrology', Journal of Hydrology, 150(2–4), pp. 189–216. doi: 10.1016/0022-1694(93)90111-L.

McGeoch, M.A .,Latombe, G.,Andrew, N.R., Nakagawa,S.,Nipperess, D.A., Roige,M., Marzinelli, E.M., Campbell, H.A., Verges,A Thomas, T.,Steinberg,P. D.,Selwood, K. E.,Hui,C.,(2017), The application of zeta diversity as a continuous measure of compositional change in ecology. Biorxiv, Available from www.biorxiv.org/content/early/2017/11/09/216580.full.pdf+html Accessed (12th August 2018)

Mcwilliam, W., Eagles, P., Seasons, M. & Brown, R. (2010) The housing-forest interface: Testing structural approaches for protecting suburban natural systems following development. *Urban Forestry & Urban Greening,* 9 (2), pp 149–159.

MetOffice, (2018), Weather and climate data, Available from: https://www.metoffice.gov.uk/climate/uk/regional-climates/ws#rainfall Accessed (12th August 2018)

Mouchet, M., Levers, C., Zupan, L., Kuemmerle, T., et al. (2015) Testing the Effectiveness of Environmental Variables to Explain European Terrestrial Vertebrate Species Richness across Biogeographical Scales. *Plos One,* 10 (7).

Natekin, A. Knoll, A. (2013) Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*. (Online) 7.

Natural England and RSPB (2014) Climate change adaptation manual, Available from: http://publications.naturalengland.org.uk/publication/5629923804839936, Accessed (12th August 2018)

Nicodemus, K.K., Wang, W., Shugart, Y. (2007) Stability of variable importance scores and rankings using statistical learning tools on single-nucleotide polymorphisms and risk factors involved in gene × gene and gene × environment interactions. *BMC Proceedings*, 1 (Suppl 1).

Ohlemuller, R., Wilson, J. (2000) Vascular plant species richness along latitudinal and altitudinal gradients: a contribution from New Zealand temperate rainforests. *Ecology Letters*, 3 (4), pp 262–266.

O'Brien, E., (1998), Water energy dynamics, climate, and prediction of woody plant species richness: an interim model, *Journal of Biogeography*, 25, pp 379-398

Oksanen, J., F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, R, P., O'Hara, O, B., Simpson, L, G., Solymos, P., Stevens, H, H., Szoecs, E., Wagner, H. (2018). vegan: Community Ecology Package. R package version 2.5-1. https://CRAN.R-project.org/package=vegan

Ostle, N. J. *et al.* (2009) 'UK land use and soil carbon sequestration', *Land Use Policy*, 26(SUPPL. 1), pp. 274–283. doi: 10.1016/j.landusepol.2009.08.006.

Packham, J.R., Harding, D.J.L., Hilton, G.M., Stuttard, R.A., (2001), Functional ecology of woodlands and forests, The Netherlands, Kluwer.

Paillet, Y., Bergès, L., Hjältén, J., Ódor, P., Avon,C., Bernhardt-Romermann, M., Bijlsma, R., Bruyn, L., Fuhr, M., Grandin, U., Kanka, R., Lundin, L., Luque, S., Magura, T., Matesanz, S., Meszaros, I., Sebastia, M., Schmidt, W., Standovar, T., Tothmerez B., Uotila, A., Valladares, F., Vellak,K., Virtanen, R., (2010) Biodiversity Differences between Managed and Unmanaged Forests: Meta-Analysis of Species Richness in Europe. *Conservation Biology*, 24 (1), pp101–112.

Pallett, D., Pescott, Schäfer, S. (2016) Changes in plant species richness and productivity in response to decreased nitrogen inputs in grassland in southern England. *Ecological Indicators,* pp 6873–81.

Palpurina, S., Wagner, V., von Wehrden, H., Hájek, M., Horsák, M., Brinkert, A., Hölzel, N., Wesche, K., Kamp, J. , Hájková, P., Danihelka, J., Lustyk, P., Merunková, K., Preislerová, Z., Kočí, M., Kubešová, S., Cherosov, M., Ermakov, N., German, D., Gogoleva, P., Lashchinsky, N., Martynenko, V., Chytrý, M. and Grytnes, J. (2017), The relationship between plant species richness and soil pH vanishes with increasing aridity across Eurasian dry grasslands. *Global Ecol and Biogeography*, 26: pp 425-434.

Petit, S., Griffiths, L., Smart, S.S., Smith, G.M., et al. (2004) Effects of area and isolation of woodland patches on herbaceous plant species richness across Great Britain. *Landscape Ecology*, 19 (5), pp463–472.

Peterken, G.F. & Game, M. (1984) Historical Factors Affecting the Number and Distribution of Vascular Plant Species in the Woodlands of Central Lincolnshire. *The Journal of Ecology*,72 (1), 155.

Pitcairn, C., Skiba, U., Sutton, M., Fowler, D., et al. (2002) Defining the spatial impacts of poultry farm ammonia emissions on species composition of adjacent woodland ground flora using Ellenberg

Nitrogen Index, nitrous oxide and nitric oxide emissions and foliar nitrogen as marker variables. *Environmental Pollution,* 119 (1), pp 9–21.

Preston, F.W. (1962) The Canonical Distribution of Commonness and Rarity: Part I. *Ecology*, 43 (2), 185.

Quine, C. P., C. Cahalan, A. Hester, J. Humphrey, K. Kirby, A. Moffat, and G. Valatin. (2011). Ch 8 Woodlands. In: R. Watson and S. Albon, editors. UK National Ecosystem Assessment. UNEP & Defra, pp 242-293

Richerson, P.J. & Lum, K.-L. (1980) Patterns of Plant Species Diversity in California: Relation to Weather and Topography. *The American Naturalist*, 116 (4), pp 504–536.

Ridgeway, G with contributions from others (2017). gbm: Generalized Boosted Regression Models. R package version 2.1.3. https://CRAN.R-project.org/package=gbm.

Ritchie, G. &; Dolling, P. (1985) The role of organic matter in soil acidification. *Australian Journal of Soil Research,*23 (4), 569.

RoTAP (2012) *Review of Transboundary Air Pollution: Acidification, Acidification, eutrophication, ground level ozone and heavy metals in the UK.*, *Centre for Ecology & Hydrology on behalf of Defra and the Devolved Administrations*.

Russell, J. (1960) Soil fertility changes in the long-term experimental plots at Kybybolite, South Australia. II. Changes in phosphorus. *Australian Journal of Agricultural Research,* 11 (6), 926.

Scheiner, S.M., (2003), Six types of species area curves, *Global Ecology and Biogeography*, 12, pp 441 - 447.

Schmidt, W., (2005), Herb layer species as indicators of biodiversity of managed and unmanaged beech forests, *Snow and landscape research.* 79 (1,2).

Shen, G., Yu, M., Hu, X.-S., Mi, X., et al. (2009) Species–area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology,* 90 (11), pp 3033–3041.

Simkin, S.M., Allen, E.B., Bowman, W.D., Clark, C.M., et al. (2016) Conditional vulnerability of plant diversity to atmospheric nitrogen deposition across the United States. *Proceedings of the National Academy of Sciences,* 113 (15), pp 4086–4091.

Slee, B., Kyle, C., Biodiversity and Woodland Ecosystems, Available from: https://www.hutton.ac.uk/sites/default/files/files/Biodiversity%20&%20Woodland%20Ecosystems(1).pdf., (Accessed 8th August 2018)

Smart, S.M., Ashmore, M.R., Hornung, M., Scott, W.A., et al. (2004) Detecting the Signal of Atmospheric N Deposition in Recent National-Scale Vegetation Change Across UK. *Water, Air, & Soil Pollution,* 4 (6), pp 269–278.

Smart, S.M., Kirby, K., Cornway, P., Wood, C.M., (2013), Woodlands survey of Great Britain 1971 - 2001, Available from: http://nora.nerc.ac.uk/id/eprint/504075/1/N504075CR.pdf, (Accessed 10th August 2018)

Smart, S.M., Ellison, A.M., Bunce, R.G.H., Marrs, R.H., et al. (2014) Quantifying the impact of an extreme climate event on species diversity in fragmented temperate forests: the effect of the October 1987 storm on British broadleaved woodlands. *Journal of Ecology*, 102 (5), pp 1273–1287.

Stevens, C.J., Payne, R.J., Kimberley, A. Smart, S.M., (2016) How will the semi-natural vegetation of the UK have changed by 2030 given likely changes in nitrogen deposition? *Environmental Pollution,* 208 Part B pp 879–889.

Storch, D., Kiel, P, Kuni, W.E., (2014), Scaling communities and biodiversity, available from: http://www.cts.cuni.cz/~storch/publications/Storch_et_al_2014_Scaling_communities_biodiversity.pdf, (Accessed 10th August 2018)

Stroh, P.A., Leach, S.J., August, T.A., Walker, K.J., Pearman, D.A., Rumsey, F.J., Harrower, C.A., Fay, M.F., Martin, J.P., Pankhurst, T., Preston, C.D. & Taylor, I. (2014). A Vascular Plant Red List for England. Botanical Society of Britain and Ireland, Bristol.

Symonds, M.R.E., Moussalli, A. (2010) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*. (Online) 65 (1), pp 13–21.

Sugihara, G., (1979), Minimal community structure: an explanation of species abundance patterns, *The American Naturalist*, 116(6) pp 770-787.

Tittensor, D., Walpole, M., Hill, S., Boyce, D., Britten, L., Burgess, G., Butchart, H. M., Leadley, W., Regan, P., Alkemade, E., Baumung, R., Bellard, R., Bouwman, C., Bowles-Newark, A., Chenery, N.M., Christensen, W., Cooper, V., Crowther, H.R., Ye, A., (2014). A mid-term analysis of progress toward international biodiversity targets. *Science*. 346.

Thiele, J., Kellner, S., Buchholz, S. & Schirmel, J. (2018) Connectivity or area: what drives plant species richness in habitat corridors? *Landscape Ecology*, 33 (2), 173–181.

Tjørve, E., Tjørve, K.M. (2017) Species-Area Relationship. *eLS*, 1–9.

Wagenmakers, E. J. and Farrell, S. (2004) 'AIC model selection using Akaike weights', *Psychonomic Bulletin and Review*, 11(1), pp. 192–196. doi: 10.3758/BF03206482.

Watts, K., Quine, C.P., Eycott, A.E., Moseley, D., Humphrey, J.W., Ray, D., (2008) Conserving forest biodiversity: recent approaches in UK forest planning and management in: Patterns and Processes in Forest Landscapes, Springer Verlag.

Williams, C. & Donald, C. (1957) Changes in organic matter and pH in a podzolic soil as influenced by subterranean clover and superphosphate. *Australian Journal of Agricultural Research*, 8 (2), 179.

Zak, D.R., Freedman, Z.B., Upchurch, R.A., Steffens, M., et al. (2016) Anthropogenic N deposition increases soil organic matter accumulation without altering its biochemical composition. *Global Change Biology*, 23 (2), pp 933–944.

Zuur, F.A., Ieno, E.N., Elphick, C.S., (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution,* 1, pp 3 -14