ENVIRON710 Final Project Paper

Logan Dye, Jiahuan Li, Suri Sun

ABSTRACT

Urban forests provide various ecosystem services, and efficient integration of urban afforestation projects to establish healthy urban forests is critical for creating sustainable urban environments. Studies have shown that the success of afforestation projects and their soil characteristics can have a bidirectional effect on each other. However, it still remains a question as to how and what factors contribute to the correlation of these two factors. In this project, we examined data on soil characteristics collected as part of the New York MillionTree Initiative to identify 1) soil features that potentially contribute to the successful afforestation and 2) afforestation and soil physical characteristics that can potentially contribute to microbial biomass nitrogen (N) and carbon (C). These findings represent a preliminary exploration of the bidirectional effect between afforestation success and soil features and will have implications for promoting the most effective afforestation projects in urban areas.

INTRODUCTION

In urban areas, afforestation is an essential technique for mitigating the negative environmental impacts of urbanization. Afforestation is the process of establishing forests on land that previously did not have forests. By generating forests, it provides various ecosystem services, including air and water purification, climate regulation, and carbon sequestration (Dos-Santos et al., 2017 and Escobedo et al., 2019). Therefore, there is growing interest in exploring the influential factors and subsequent environmental effects of afforestation success as it is critical for creating a more sustainable urban environment.

The success of afforestation may be influenced by both manual factors, such as the diversity of planted tree species, and environmental factors, especially the soil conditions. In turn, afforestation also has the potential to provide multiple environmental benefits, such as soil improvement and carbon sequestration. Several studies have investigated the long-term bidirectional influence between afforestation success and soil characteristics. They showed that afforestation could improve soil characteristics such as microbial biomass C and N by providing additional organic matter and nutrients. On the other hand, the soil's quality and nutrient availability can also influence afforestation success. Organic carbon and nitrogen, for instance, can influence soil biogeochemical properties and functions by enhancing nutrient cycling and water retention, which ultimately supports plant growth (Ontl and Schulte, 2012). Therefore, the relationship between soil and afforestation success is bidirectional, with each influencing the other in a complex feedback loop. Understanding these interactions will be necessary to enhance our understanding of successful afforestation and help develop effective strategies to maximize the ecological benefits of urban forests.

To understand the mutual effects between soil characteristics and afforestation, we examined data on soil characteristics from a series of experimental restoration sites established between 2009 and 2011 as part of the MillionTreesNYC Initiative. This project aims to answer the following questions:

1. How does the probability of afforestation success (high or low) vary with different diversity treatments (high or low) and soil characteristics (quantified by organic and inorganic substances and microbial activities)?

2. How does soil biomass nitrogen (N) vary with afforestation characteristics (diversity and success), soil physical and biochemical characteristics (such as respiration and nitrification), and soil depth (0–10, 10–30, below 30 cm)?

3. How does soil biomass carbon (C) vary with afforestation characteristics, soil physical and biochemical characteristics, and soil profile depth?

Based on the findings of the existing literature, we hypothesize that soil characteristics and the diversity of tree species will significantly predict afforestation success. We also expect that soil biomass nitrogen and carbon will be positively correlated with afforestation success and proxies of soil quality.

METHODS

Data Collection

The dataset was part of the MillionTreesNYC Initiative effort of the New York City Department of Parks and Recreation (Mejia 2021). During 2009-2011, long-term research plots were set up at 10 afforestation sites of 7 municipal parks in New York (Figure 1). Each site was planted with either 2 or 6 species of trees as low and high diversity treatments (Downey et al. 2021). Soil core samples were collected from plots at each afforestation site between 2018-2019. The soil core samples were processed to measure physical features (bulk density, moisture content, rock mass, etc.) and biochemical contents (microbial biomass C and N, mineralization, respiration, etc.) (Downey et al. 2021). Afforestation success was also determined as low or high based on whether the plots had developed a closed canopy and leaf litter layer at the time when data were collected (Mejia 2021).

Figure 1. *Location of the 10 afforestation sites of 7 municipal parks in New York (Downey et al. 2021).*

Data preparation

We examined the dataset and edited it accordingly to ensure consistent namings and format. We deleted sampling-associated columns, including sample/soil core IDs, and removed unnecessary variables such as total core length, Munsell Color codes, and soil horizons. We removed incomplete variables that mainly consist of NAs and rows that contain at least one NA. We also omitted two extreme observations of Bulk Density and NH4 that fall beyond the reasonable range of the variables, which are highly likely due to recording errors. To best represent the effect of sampling dates, we separated years and months into two columns. We created a new nominal variable called "Season" (summer or fall) based on the specific sampling dates in relation to the summer solstice and autumn equinox. We created additional columns, used integers to denote character-based binary variables such as afforestation success and diversity (1=high, 0=low), and regrouped soil core sections into three categories (1=0-10cm, 2=10-30cm, 3=below 30cm). We made paired plots screened for strong correlations among independent variables and excluded TIN and mineralization from our analyses as having >0.9 correlation with other variables (Figure A in Appendix).

Data Analysis

The dependent variables of the first two linear mixed-effects models were microbial biomass nitrogen ("BiomassN") and microbial biomass carbon ("BiomassC"), with the remaining variables on soil characteristics as independent variables and site/plot as a combined random effect. These initial full models were then systematically reduced to obtain the minimum adequate models. Due to almost all variables being highly right-skewed based on the paired plots (Figure A in the Appendix), the variables used in the model were log-transformed for the linear mixed-effect model describing BiomassN using the logarithm of (original value + abs(minimum negative value) + 0.001). Following the reduction of the full model, the residual and Q-Q plots were examined to check for normality. Highly influential observations were identified using the outlier test and influence plot, and observations determined to be highly influential were removed from the dataset before refitting the model. The variance inflation factor (VIF) was calculated to check for leftover multicollinearity. We also checked the goodness of fit of the models.

The dependent variable for the third generalized linear mixed model was afforestation success ("success"), with site/plot as a combined random effect and the remaining ones (including biomass C and N) as independent variables. Because success is a binary variable, we used "binomial" as the family for this model. Initial nitrate ("NO2_NO3") was excluded from the analysis due to strong multicollinearity, and the model is stepwise reduced to obtain the minimum adequate model. Because the full model indicates huge variation in variable scales, independent variables were standardized by centering to z-scores. This, however, did not resolve the error that the model failed to converge. We also tried transforming the independent variables by taking the logarithm of (original value + abs(minimum negative value) + 0.001) to account for the issue of negative values, but the error still existed after transformation. Therefore we kept the variables untransformed and used the standardized ones in this model. We also calculated the variance inflation factor (VIF) to check for leftover multicollinearity.

RESULTS

The final cleaned dataset contains 326 observations on 21 variables in total. Of all, 8 are categorical variables, including park site ("Site"), study plot ("Plot"), afforestation success ("success," 1=high and 0=low), tree species diversity ("diversity," 1=6 species and 0=2 species), year (2018 or 2019), month (between 6-10), season (fall or summer), and core section (1=0-10cm, 2=10-30cm, 3=below 30cm). The rest are numeric variables, including rock mass ("RockMass_g," mean=10.24g), root mass ("RootMass_g," mean=0.34g), moisture ("Moisture_G," mean=0.12g), bulk density ("BulkDensity_g_cm3", mean=0.83g/cm3), biomass carbon ("BiomassC," mean=439.94ug/g dry soil), rate of respiration (“Respiration,” mean=9.9ugC/g soil/day), initial nitrate (“NO2_NO3”, mean=13.31ugN/g dry soil), initial ammonium (“NH4”, mean=0.976ug-N/g dry soil), biomass nitrogen (“BiomassN,” mean=34.75ug-N/g dry soil), potential net nitrification (“Nitrification,” mean=0.12ug-N/g dry soil/day), and denitrification enzyme assay (“DEA,” mean=136.62ug-N/g dry soil/hour).

Biomass N

For BiomassN, the null hypothesis is that no variables will significantly impact the amount of BiomassN present. The alternative hypothesis is that at least one variable will significantly impact the amount of BiomassN present. The minimum adequate model (lowest AIC) determined that the variables diversity, season, NO2_NO3, NH4, Nitrification, and BiomassC significantly impact the amount of BiomassN present when the plot and site are controlled as a random effect. The results linear mixed-effect model are as follows: The intercept (Diversity = Low, Season = Fall, 0 NO2_NO3, 0 NH4, 0 Nitrification, and 0 BiomassC) had a coefficient of -2.00 ($t = -8.571$, $df = 263.1$, $p < 0.001$). High species diversity had a coefficient of 0.19 ($t = 2.036$, $df = 33.4$, $p = 0.005$). NO2_NO3 had a coefficient of 0.39 ($t = -7.422$, $df = 155.5$, $p < 0.001$). The season summer had a coefficient of -0.22 ($t = -3.396$, $df = 303.6$, $p < 0.001$). NH4 had a coefficient of 0.21 ($t = 3.193$, $df = 308.7$, $p = 0.002$). Nitrification had a coefficient of 0.11 ($t = 3.661$, $df = 315.8$, $p < 0.001$). BiomassC had a coefficient of 0.69 ($t = 14.546$, $df = 290.5$, $p < 0.001$). The largest and most significant effect in relation to BiomassN was Biomass C, with a 0.69% expected increase in BiomassN for every 1% increase in BiomassC (Figure 2). Overall, 76% of the BiomassN variance was explained by the fixed effects in the model, and 78% of the BiomassN variance was explained by both the fixed and random effects in the model.

The assumption test results show a slight deviation from normality within the residuals (Figure B in Appendix). However, the residuals appear to be scattered randomly, as expected. No significantly influential outliers were detected, and the variance inflation factor (VIF) determined that no excess of multicollinearity was present within the model.

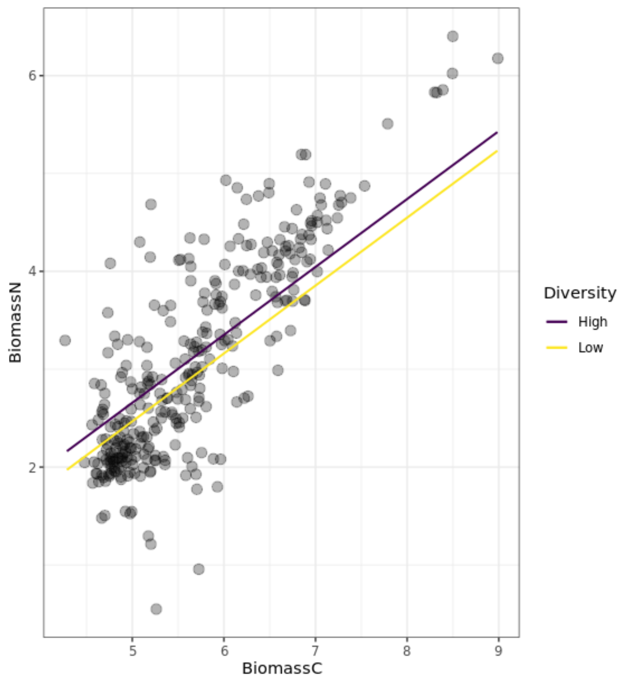

Figure 2. *The above figure represents the relationship between the most significant variable (BiomassC) and the dependent variable (BiomassN). The regression line was plotted twice. Once to represent high-diversity treatment sites and another to represent low-diversity treatment sites.*

Biomass C

Based on similar hypotheses to those for biomass nitrogen, the minimum adequate model for biomass carbon included the following main fixed effects in order of their effect size: respiration, moisture, core section depth, season, biomass N, and diversity. Additionally, the model accounted for site and plot differences by including them as a random effect. The model was fitted using the REML (restricted maximum likelihood) method, and Satterthwaite's method was used to estimate the degrees of freedom.

The intercept value is statistically significant ($t = 24.646$, $df = 238.6$, $p < 2.2e-16$), indicating that the model predicts a non-zero value for the biomass carbon when all other predictors are equal to zero. Specifically, the intercept value of 5.41 corresponds to a predicted mean value of biomass carbon of 224.5 g (exp^{5.41}) when all other numeric predictors are equal to zero, and the categorical predictors are high diversity, autumn, and core section depth in 0-10cm.

For the nominal explanatory variables, the low diversity group of 2 tree species had a significantly higher mean biomass carbon, with an increase of 12.5% ((exp(0.1179) - 1) * 100) compared to the high diversity group of 6 tree species ($t = 2.429$, $df = 39.4$, $p =0.020$). Moreover, the season variable significantly affects biomass C, with a decrease of 17.6% from autumn to summer ($t = -3.305$, $df = 305.6$, $p =0.001$). The core section depth predictor also had a significant effect on biomass C. Changing from the 0-10cm depth to 10-30cm led to a mean biomass carbon decrease of 26.8% ($t = -4.543$, $df = 297.4$, $p =8.07e-06$), while changing to above 30 cm caused a biomass decrease of 40.9% ($t = -5.758$, $df =281.5$, $p =2.22e-08$).

For the numerical explanatory variables, for every 1% increase in respiration, moisture, and biomass N, the mean biomass carbon will increase by 0.455% ($t = 8.085$, $df = 287.0$, $p =1.76e-14$), 0.383% ($t = 8.246$, $df = 315.4$, $p =4.46e-15$), and 0.149% ($t = 5.688$, $df = 314.7$, $p =2.94e-08$) accordingly.

The model's fit to the data is good based on both the marginal R-squared (R2m = 81.98%) and the conditional R-squared (R2c = 85.34%). The three largest effects are shown in the accompanying graphs. And all the fixed and random effects are illustrated in Appendix F. Additionally, the normality of the residuals and random effects appears to be satisfactory based on the performance plots (Appendix G). The residuals appear to be randomly scattered, and their mean value is close to zero. There are no significant outliers, with the maximum cook's distance being 1.43, as shown in the influence plot (Appendix H). Lastly, another performance plot indicates that the multicollinearity of the factors is low (Appendix I).

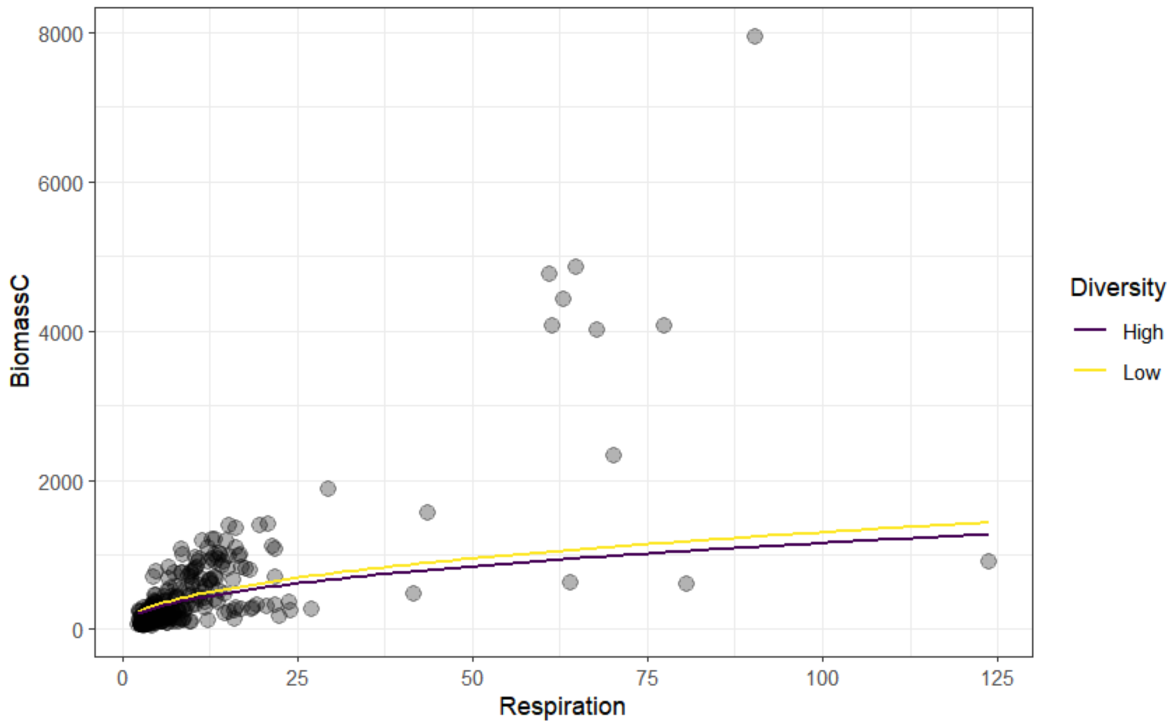

Figure 3. *Effects of respiration and diversity on biomass C.*

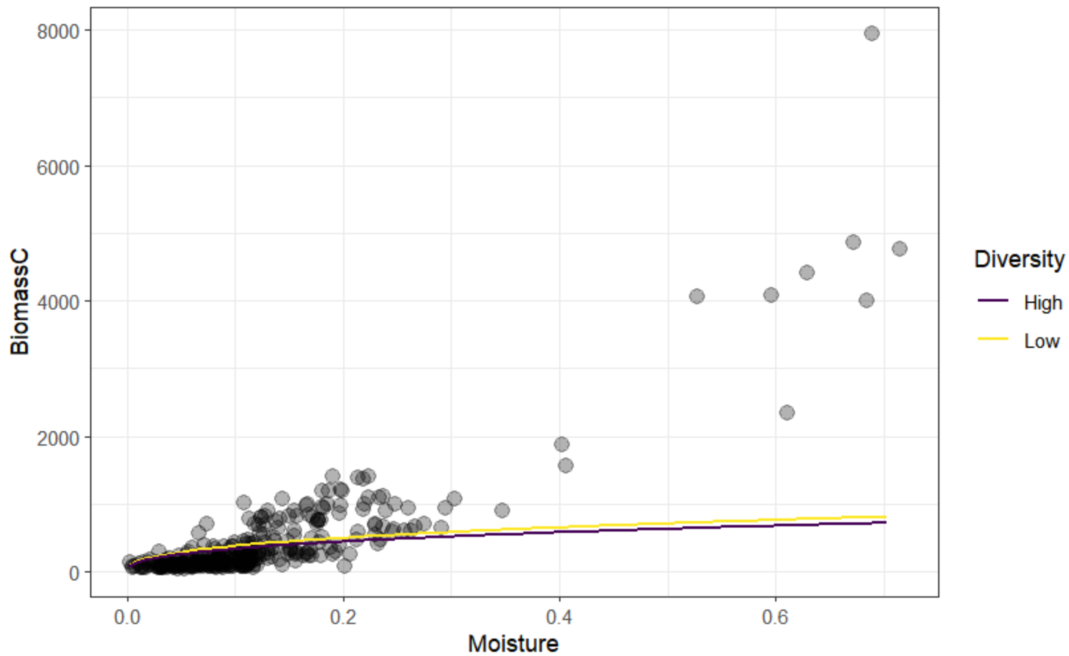

Figure 4. *Effects of moisture and diversity on biomass C.*

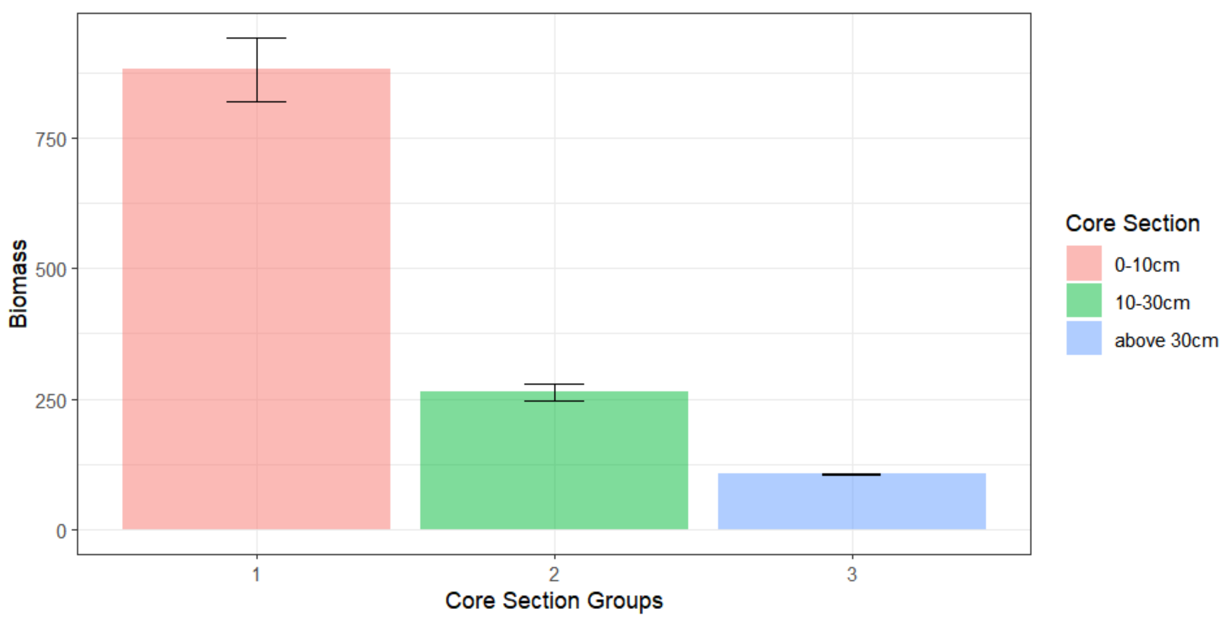

Figure 5. *Mean biomass carbon and differences among core section groups.*

Afforestation success

The null hypothesis of this model is that none of the independent variables significantly predict afforestation success, and the alternative hypothesis is that at least one of the independent variables significantly predicts success. The backward stepwise reduction did not yield a minimum adequate model for predicting afforestation success. We did not find significant effects of any predictor variables on the probability of success.

We also fitted the model that retained only the predictors of our interest, including diversity, biomass N, biomass C, root mass, bulk density, and core section. The VIF of all variables is below 2, which met the assumption of having no strong multicollinearity. The intercept represents the odds of success is 6.137904e+06 under high diversity treatment while biomass N, biomass C, root mass, and bulk density all equal 0 (estimate=16.46, z=3.32, p<0.001). None of the independent variables significantly affect predicting afforestation success after being centered by z-score. Based on the result we obtained, the odds of successful afforestation increase by approximately 13% when there is a low diversity of tree species (estimate=0.12, z=0.02, p=0.98), by over 100% with every additional ug of biomass N per gram of dry soil (estimate=0.91, z=0.106, p=0.92), and by 2.8% with every additional g/cm3 of bulk density (estimate=0.03, z=0.009, p=0.99). It decreases by approximately 13% with every additional ug of biomass C per gram of dry soil (estimate=-0.13, z=-0.014, p=0.99) and by approximately 20% with every additional gram of root mass (estimate=0.21, z=0.01, p=0.99). The poor fit of the model to the data also yielded an unexpected trend in the plot that deviates from our original expectation (Figure 6)

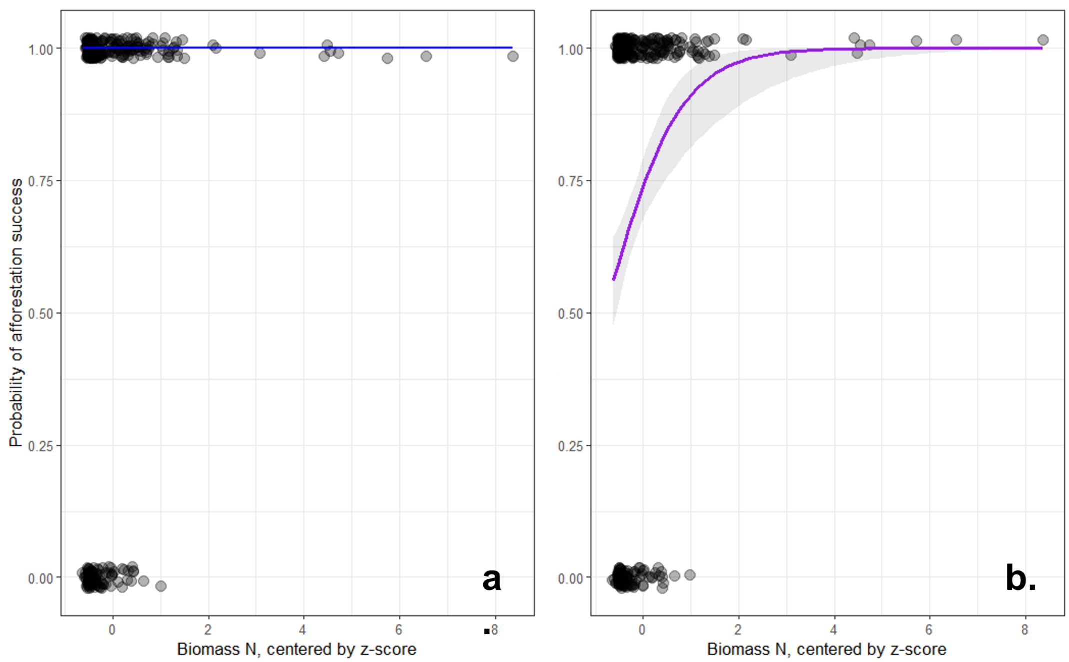

Figure 6. *Scatterplots of the probability of afforestation success against Biomass N, centered by z-scores, overlaid with the prediction curve generated based on coefficients from the glmer() model we fit (a) and the expected theoretical line (b).*

It is also noticeable that during the process of backward reduction, the model convergence issue disappeared and reappeared when certain variables were removed. We did not encounter the issue of model convergence in the model as presented above. Still, the summary results showed a warning on singularity when the bulk density ("BulkDensity_g_cm3") was excluded from the model.

DISCUSSION AND CONCLUSION

The first model concluded that six variables significantly impact the amount of BiomassN present in the soil. Diversity, NO2_NO3, NH4, Nitrification, and Biomass are all positively associated with BiomassN, while the season is negatively associated with BiomassN. These results make sense as higher values of NO2_NO3, NH4, Nitrification, and BiomassC are all associated with better, healthier soil. Since high levels of BiomassN are also associated with healthy soil, we can confirm that our independent variable, BiomassN, increases with other soil metric values. The most relevant finding of the BiomassN model is the impact diversity has on BiomassN. The model showed afforestation plots with high diversity (planted with six tree species rather than two) had higher BiomassN than afforestation plots with low diversity. This is interesting as it infers that having a higher species diversity in the canopy layer will lead to richer and healthier soils.

The findings from model 2 of biomass carbon provide important insights into the mutual effects between soil characteristics and afforestation. While some predictors, such as respiration, moisture, core section depth, biomass N, and season, have significant effects on biomass carbon, others mentioned in the hypotheses, such as the success of afforestation, are found to be insignificant. The reason why diversity is more significant than afforestation success could be due to several factors. One possible explanation is that different tree species have different functional traits, which can affect their ability on soil carbon sequestration (De Deyn et al., 2008). Moreover, the finding that lower diversity is associated with higher biomass C is not a universal pattern and may be context-dependent. For example, a study published in the journal Ecosystems found that increasing tree species diversity significantly increases soil carbon storage in a subtropical forest in China (Liu et al., 2018). The reasons for these discrepancies may include differences in environmental conditions, tree species composition, and management practices. Therefore, more research is needed to understand the complex relationship between tree diversity and soil carbon sequestration in different ecosystems. These findings highlight the importance of considering multiple factors when designing afforestation projects to maximize their carbon sequestration potential.

The generalized linear mixed model did not predict the success of canopy and leaf litter layer development at afforestation sites. This might be partially due to the fact that the soil characteristics were measured at the same time when afforestation success was determined. The probability of whether a site develops closed canopy cover and leaf litter layer may be more likely related to initial soil conditions when the trees were first planted rather than the most recent characteristics during sample collection. As a result, it is reasonable that the variables failed to predict success in the model we fit. This project can be hugely improved if we incorporate earlier data collected in 2009-2011. This may allow us to detect a much clearer trend in the strength of past soil conditions to predict future afforestation success.

It is also likely that a generalized linear mixed model might not be the best option to predict afforestation success from soil characteristic variables in this particular dataset. Previous studies used one-way ANOVA to compare soil characteristics between different diversity treatments and levels of success (Downey et al. 2021). It is possible that the setup of this study design and dataset may not best support fitting a binary logistic regression. Errors frequently occurred on model convergence and occasionally on singularity. Although we were not able to identify the potential factors that led to these issues, it is still likely that the distribution of the data might be one of the causes. It might be necessary to incorporate more sophisticated methods to transform or standardize the raw data to ensure that they meet the assumptions of the models that fit these variables the best.

Because the dataset we used in this project only examined selected afforestation sites in New York City, the scope of inference from this analysis will be restricted to the particular geographic location.

Additional studies will be needed to understand the connection between soils and afforestation progress in other places with different environmental conditions. Even more, data that cover a longer time frame will be preferred to further understand the effects of these afforestation projects in New York. Besides, it is also important to note that the number of levels of tree diversity is limited and therefore may not be sufficient to draw general conclusions on the effect of tree species diversity. Further studies with more tree diversity levels are needed to fully explore the relationship. Although some portions of our analyses yielded results different from our original hypothesis, our findings, in general, can provide the most preliminary insights for understanding the bidirectional effects between soil conditions and the effectiveness of afforestation.

Overall, the most important finding of our study is that species diversity was not significant to the success of the afforestation sites. This suggests that cities may be able to establish successful afforestation sites with a focused selection of a few tree species, which can potentially reduce the costs associated with planting and maintaining much more diverse species. If a city's goal is to establish a healthy canopy layer of trees, it is more important to focus on several species rather than diverting too many resources to planting and maintaining many different species. Despite the implications for resource allocation and feasibility, it is still crucial to pay attention to and further investigate the impact of tree diversity on other goals, such as habitat diversity and resiliency. Future studies should explore the influence of tree diversity on the long-term success and ecological sustainability of similar afforestation initiatives. Overall, our project will contribute to the growing body of literature on the complexities of urban afforestation ecosystems and inform decision-making in managing urban forests.

REFERENCES

De Deyn, G. B., Cornelissen, J. H., & Bardgett, R. D. (2008). Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecology letters*, *11*(5), 516-531.

Dos-Santos, A. R., de Oliveira, F. S., da Silva, A. G., Gleriani, J. M., Gonçalves, W., Moreira, G. L., ... & Mota, P. H. S. (2017). Spatial and temporal distribution of urban heat islands. *Science of the Total Environment*, 605, 946-956.

Downey, A. E., Groffman, P. M., Mejia, G. A., Cook, E. M., Sritrairat, S., Karty, R., Palmer, M. I., and McPhearson, T. (2021). Soil carbon sequestration in urban afforestation sites in New York City. *Urban Forestry & Urban Greening, 65*, 127342.

Escobedo, F. J., Giannico, V., Jim, C. Y., Sanesi, G., & Lafortezza, R. (2019). Urban forests, ecosystem services, green infrastructure and nature-based solutions: Nexus or evolving metaphors?. *Urban Forestry & Urban Greening*, *37*, 3-12.

Liu, X., Trogisch, S., He, J. S., Niklaus, P. A., Bruelheide, H., Tang, Z., ... & Ma, K. (2018). Tree species richness increases ecosystem carbon storage in subtropical forests. *Proceedings of the Royal Society B*, *285*(1885), 20181240.

Ontl, T. A., & Schulte, L. A. (2012). Soil carbon storage. *Nature Education Knowledge*, *3*(10).

APPENDIX 1 Plots

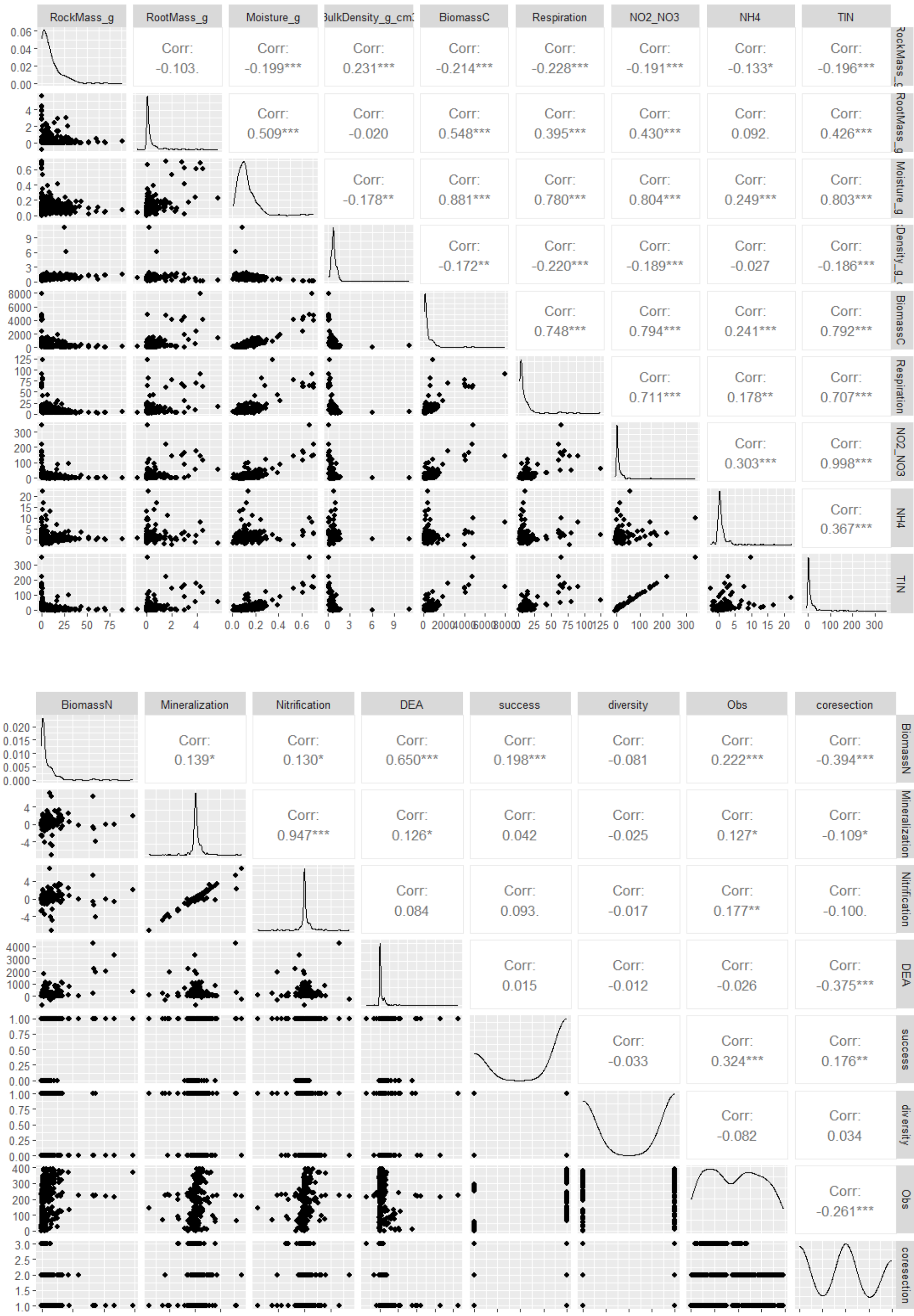

Figure A. *Exploratory plots of all variables generated using the function Ggpairs. Almost all of the variables are strongly right skewed.*

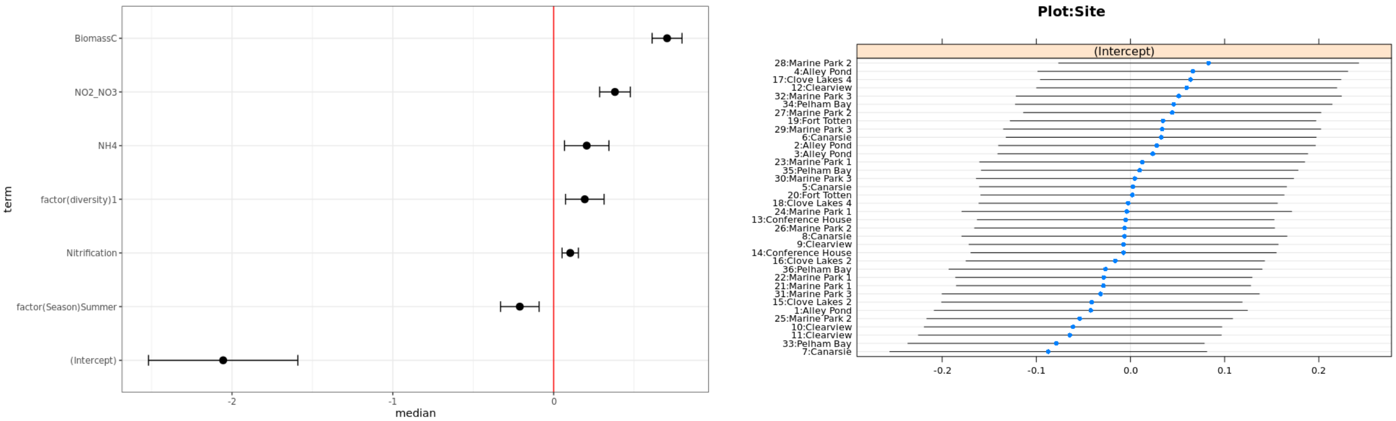

Figure B. *Fixed and random effects for models on Biomass N.*

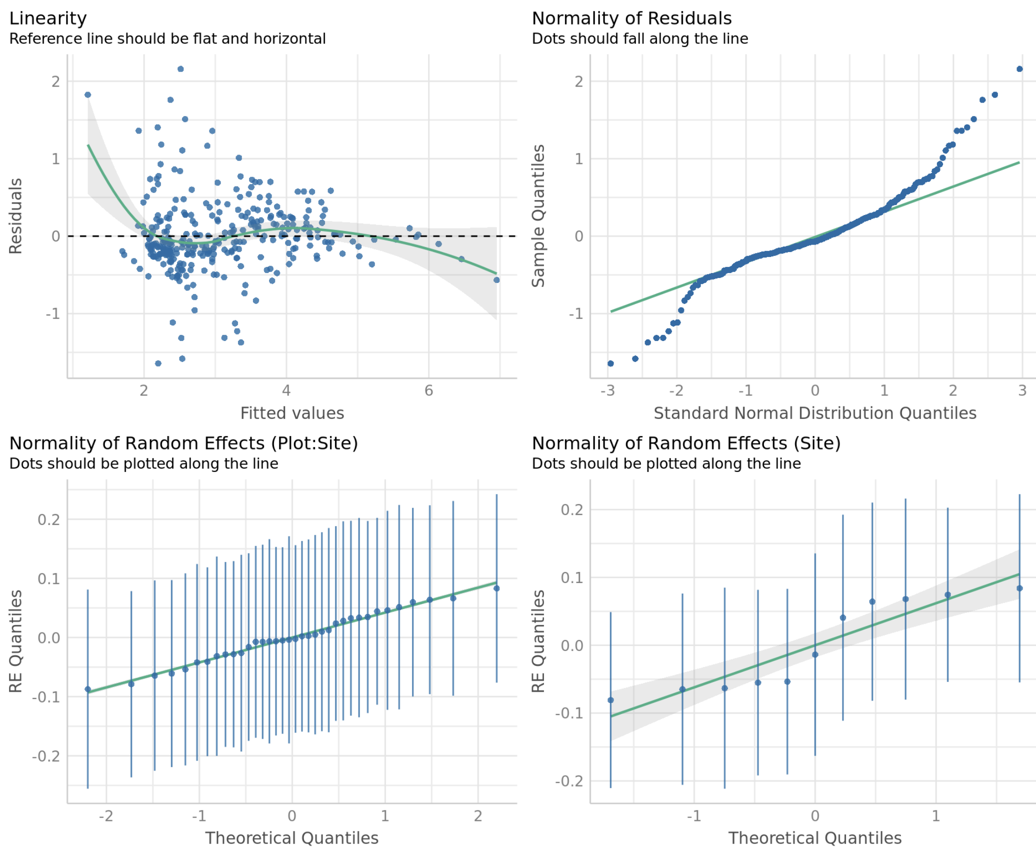

Figure C. *Residual plot, QQ plot, and normality plots of the random effect for model on Biomass N.*

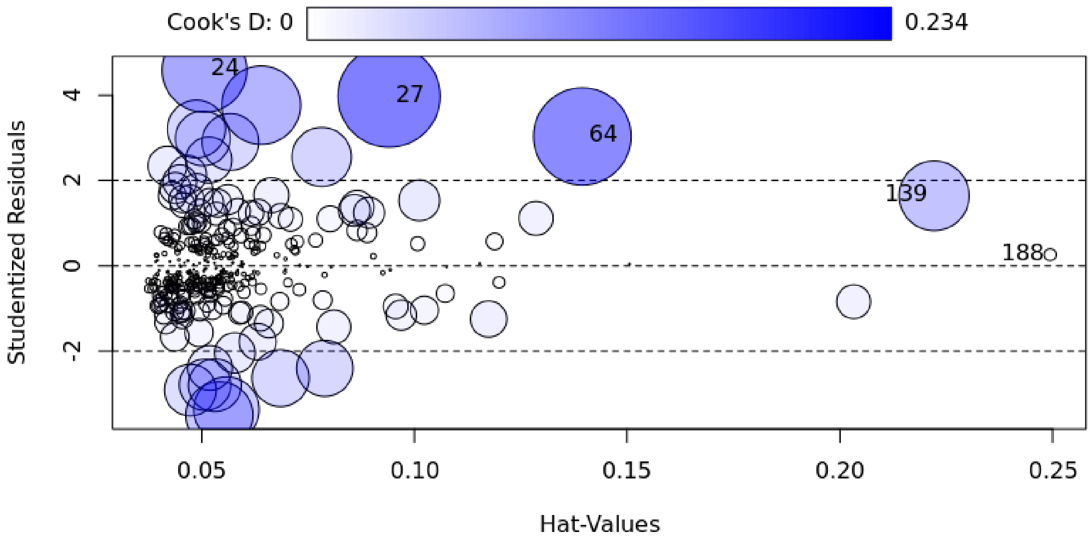

Figure D. *Outlier plot of the model on Biomass N*

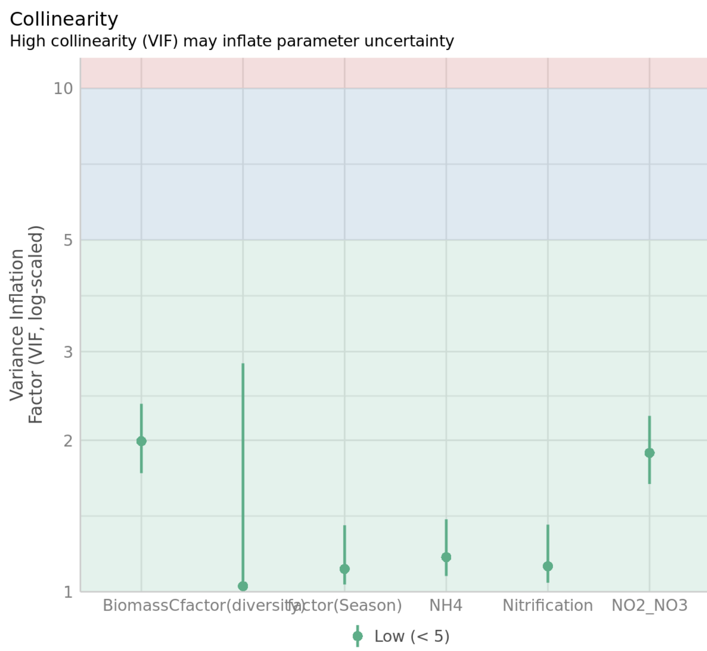

Figure E. *Collinearity plot of the model on Biomass N*

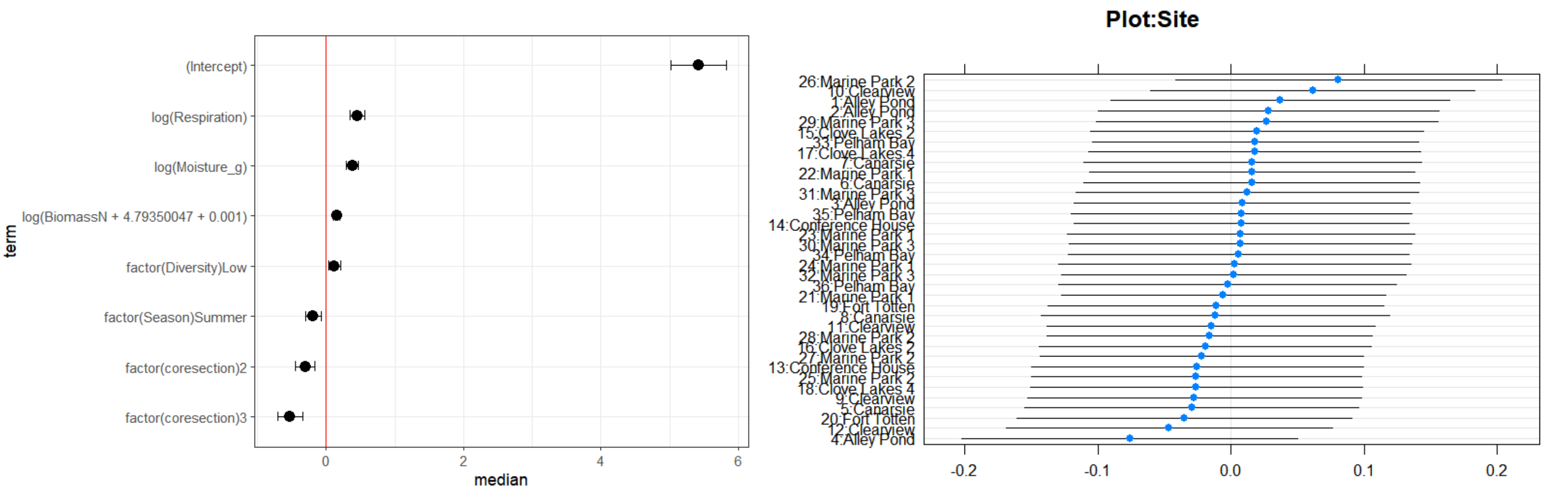

Figure F. *Fixed and random effects for models on Biomass C.*

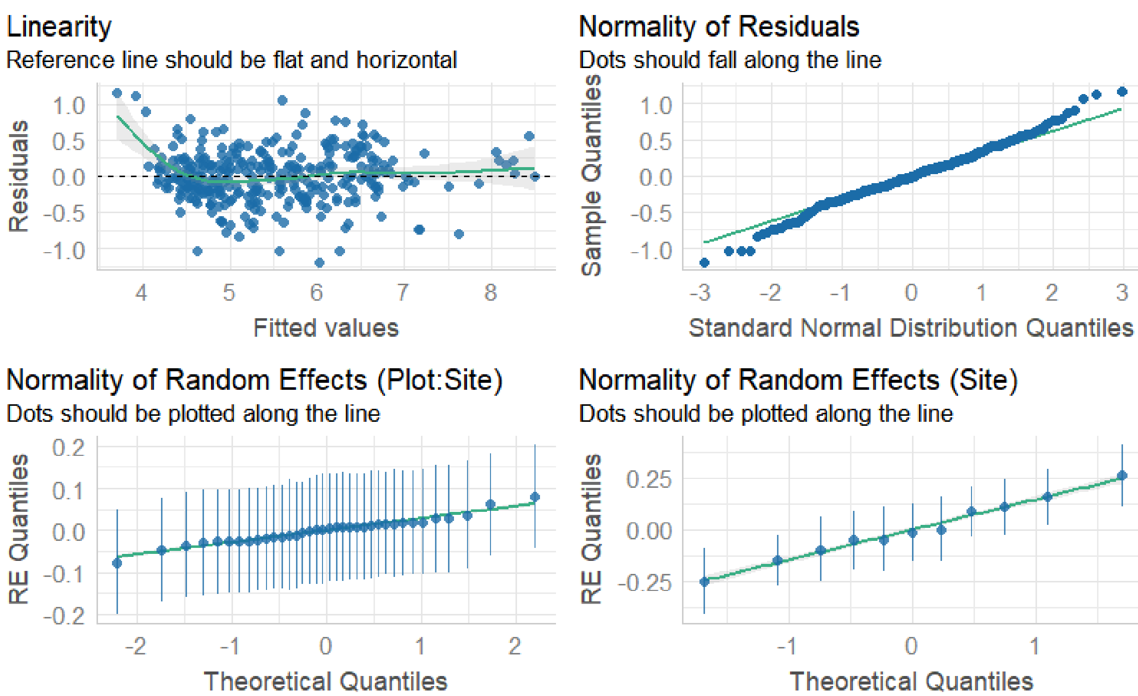

Figure G. *Residual plot, QQ plot, and normality plots of the random effect for model on Biomass C.*

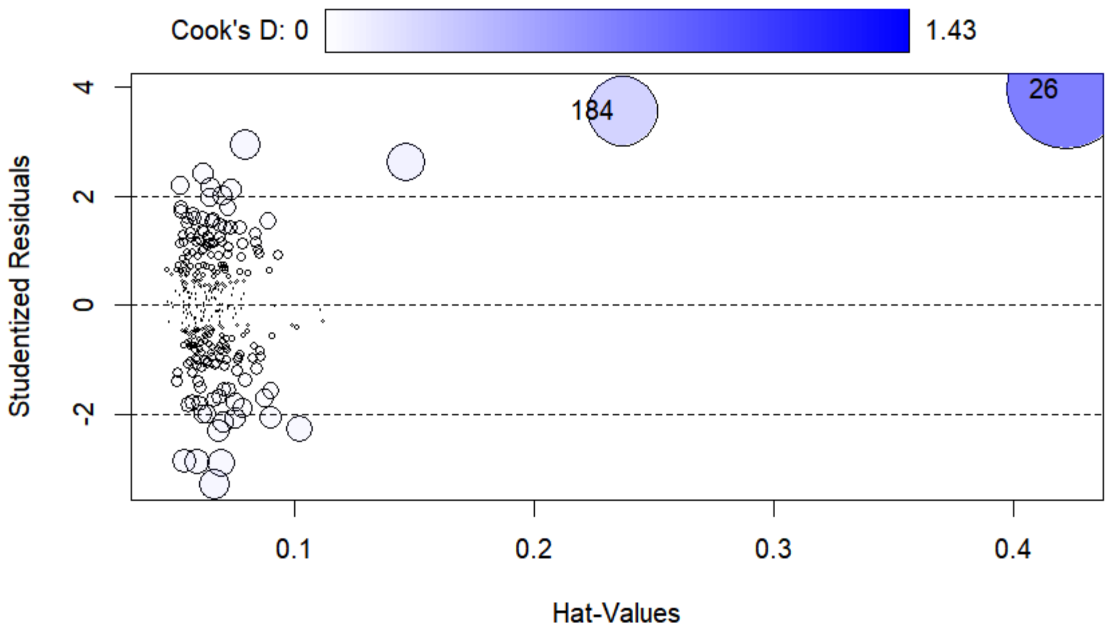

Figure H. *Outlier plot of the model on Biomass C*

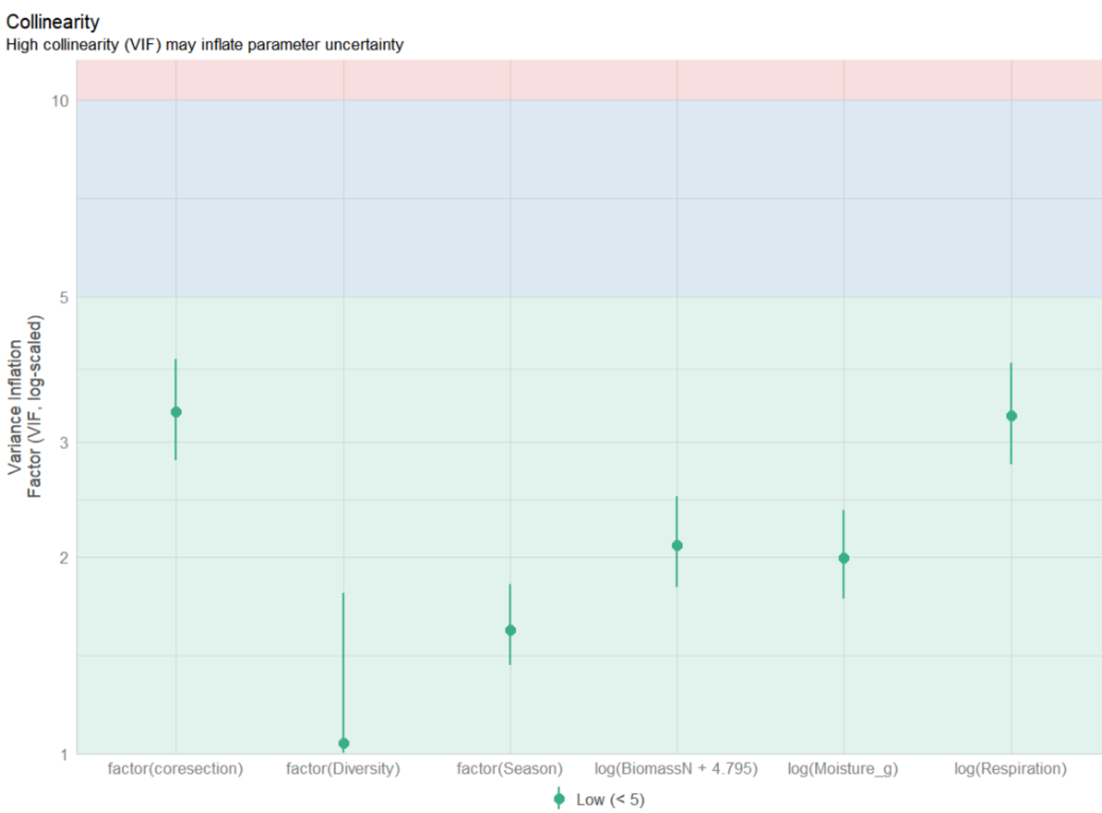

Figure I. *Collinearity plot of the model on Biomass C*