

# Final Project Report

*Trevor Drees and Emily Howerton*

*December 15, 2019*

## Abstract

## Introduction and Data

As the world has become significantly more interconnected, the increased movement of people across the world has also facilitated an increase in the spread of species outside of their native ranges. When species that are introduced outside of their native range become invasive, they can have a detrimental impact on local biodiversity, quantity and quality of valuable ecosystem services, and other aspects of the local ecosystem. Because of this, the study of invasive species has been a large part of the ecology literature in recent decades, ranging from theoretical models of how they spread to investigating possible management practices that keep them at bay.

One particular invasive species that has greatly expanded its range due to human activity is *Carduus nutans*, also known as “musk thistle” or “nodding thistle”. This thistle is native to Europe and Central Asia, but has expanded its range into North America, Australia, and New Zealand, among other parts of the world (source). Within the U.S., this thistle has been reported in all U.S. states except for Alaska, Florida, Hawaii, Maine, and Vermont (source); the thistle may even be present in these states, but has not yet been reported. It is also been reported all Canadian provinces except Nunavut, Northwest Territories, and Yukon Territories (source).

*C. nutans* is considered to be a noxious weed in many U.S. states for several reasons. Because it can occur in very large numbers and grow to be quite large, this thistle may form dense and often impenetrable stands. The plant is also covered in numerous large spines, making it painful when touched as well as unpalatable to grazing animals. The adverse impacts of this weed on grazing can also lead to substantial economic losses.

Another reason why there is concern over *C. nutans* is because it has a high potential to spread locally when introduced, as it is wind dispersed and large pappi on the seeds allow them to be transported great distances. Models have been proposed to model such wind-driven seed dispersal (source), and such models have been applied to *C. nutans* (source), showing significant potential for long-range dispersal events. While there are abiotic and biotic factors that affect how far a seed like those in *C. nutans* can be dispersed, a noteworthy predictor of dispersal distance is seed terminal velocity. For seeds, a higher terminal velocity generally means a decreased dispersal distance; this is because a higher terminal velocity means the seed falls faster and thus spends less time in the air, which means less of an opportunity for wind to carry it further from its source.

However, it is not entirely clear what affects terminal velocity in *C. nutans* seeds, though the most obvious candidates would be physical properties of the seed such as shape and mass. In general, seeds with a larger area perpendicular to the direction of motion will have higher drag and a lower terminal velocity. Seeds with a higher mass will have a higher downward force (mg) from gravity and thus a higher air resistance force that must equal it to achieve terminal velocity, which leads to a higher terminal velocity since said resistance force is proportional to that velocity. However, the physical properties of the seed may be affected by the morphology and physiology of the parent plant; abiotic and biotic factors can affect the parent plant in such a way that may ultimately influence the terminal velocity of its seeds.

Given that there may be a link between abiotic influences on *C. nutans* and the terminal velocity of its seeds, we wish to investigate whether certain treatments applied to the plant before it flowers have any effect on seed dispersal capabilities. Any treatment effects that can reduce the dispersal capability of these thistles may then be used to inform management decisions. By using mowing treatment as well as a warming treatment

(and combinations of the two), we will examine the effects of said treatments on seed terminal velocity and thus on dispersal capability.

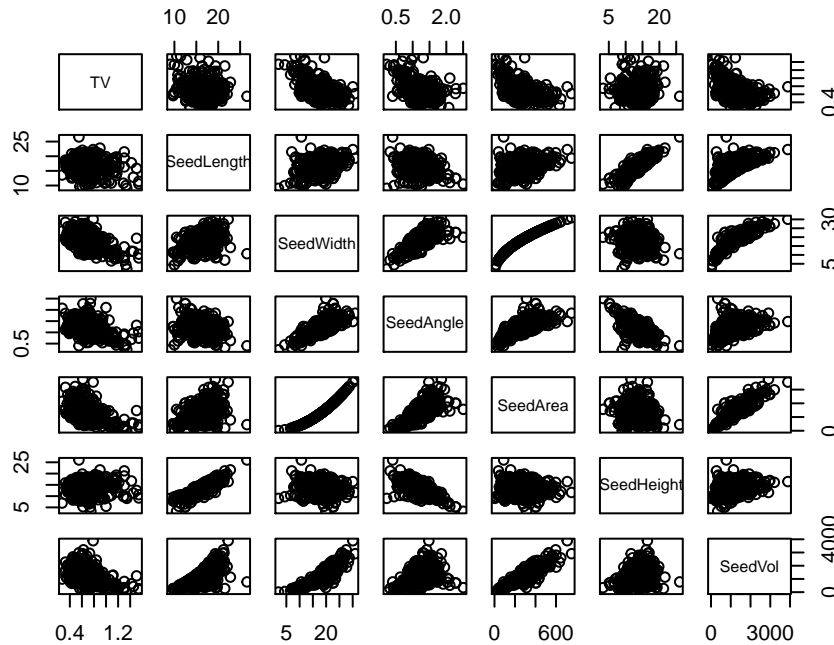
The data used to assess the effectiveness of the treatments were collected during a field experiment which crossed warming and mowing treatments. There were ten blocks, where each block subset into two plots - one warmed treatment and one ambient. Within a plot there are three positions, one for each mowing treatment. Ten seeds were planted at each position. One flower head was harvested from all individuals that survived to harvest date. Seeds were collected from individual flower heads and subsequently tested in a drop chamber. Seed drop tests were repeated until two drop times were recorded within 0.1 seconds. With these data, we want to assess the variation in drop time of seeds from plants under warming and/or mowing.

With this data, we will address three main research questions:

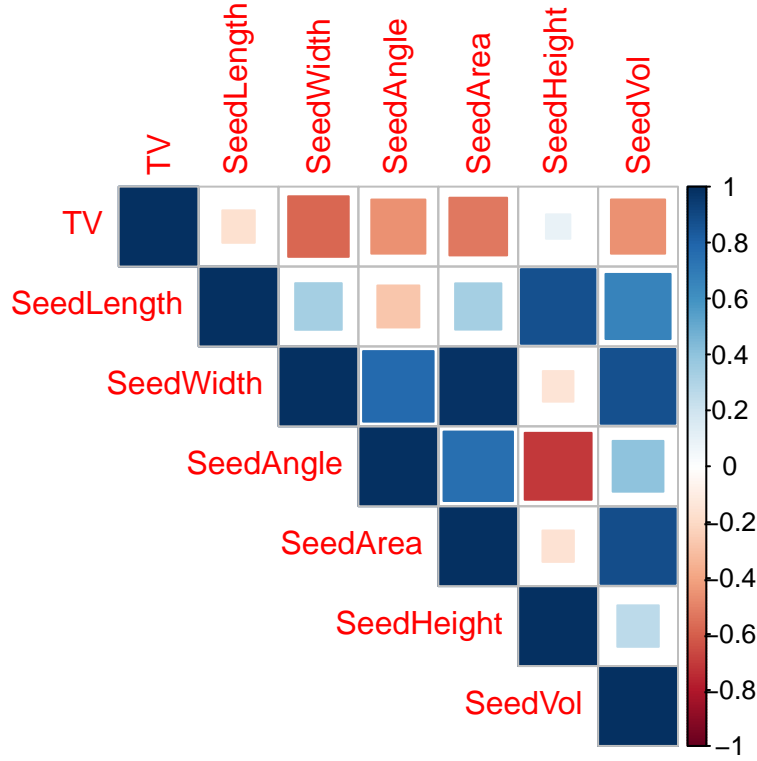
1. Is seed terminal velocity predicted by seed shape parameters?
2. To what extent do warming and mowing treatments change seed terminal velocity?
3. Can changes in terminal velocity by treatment be explained by changes in seed shape parameters?

## Research Question 1

Before any models were fit to the data, we first examined correlations between the various seed physical properties. The `pairs` plot for terminal velocity and the various physical properties is shown below:



Linear correlations between the variables can also be shown in a heatmap of the correlation matrix:



Between both of these plots, it is clear that multicollinearity will be an issue since several of the variables are highly correlated with each other (e.g. **SeedWidth** and **SeedAngle**, **SeedWidth** and **SeedArea**, etc.).

Our first model used the terminal velocity **TV** as a response and each of the physical properties as predictors. The resulting coefficients and their significance are listed below:

	Estimate	Std. Error	t value	Pr(> t )	VIF	Significance
(Intercept)	1.7939	0.2872	6.2465	0		***
SeedWidth	-0.0902	0.0294	-3.0629	0.0024	231.7271	**
SeedLength	0.2323	0.1394	1.6664	0.0965	1672.8695	.
SeedArea	-0.0012	0.0021	-0.5918	0.5544	918.6289	
SeedAngle	-0.2099	0.4397	-0.4774	0.6334	291.9196	
SeedHeight	-0.2336	0.1536	-1.5214	0.129	2709.2157	
SeedVol	4e-04	2e-04	2.2275	0.0265	124.7234	*

It is clear that this model is a poor fit for two main reasons. First, many of the terms are not significant to  $\alpha = 0.05$  and add no significant additional predictive power to the model. Second, the VIF for all of the coefficients is quite high, as was predicted.

Given that some of these variables must be removed to improve the fit of the model, we then performed variable selection on the full model to reduce clutter. The first method of selection was backwards selection from the full model; this involved removing subsequent variables based on their  $p$ -value; the variable with the highest  $p$ -value was removed, the resulting model examined, and the process repeated until all predictors were significant. The second method of selection involved backwards selection from the full model using AIC, where terms were removed until the AIC of the model was minimised. The third method of selection started from the full model and used **step** selection in both directions, while the fourth method started from the null model and added terms; again, in both cases, the selection continued until the AIC of the model was minimised.

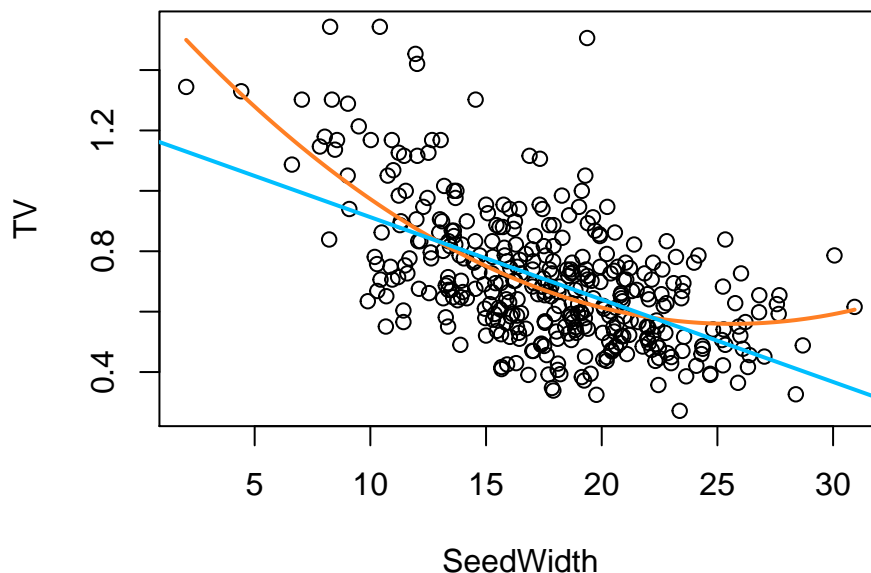
The results from the model selection can be seen below, along with the VIF for each term and AIC for each resulting model. All coefficients are significant to  $\alpha < 0.001$ .

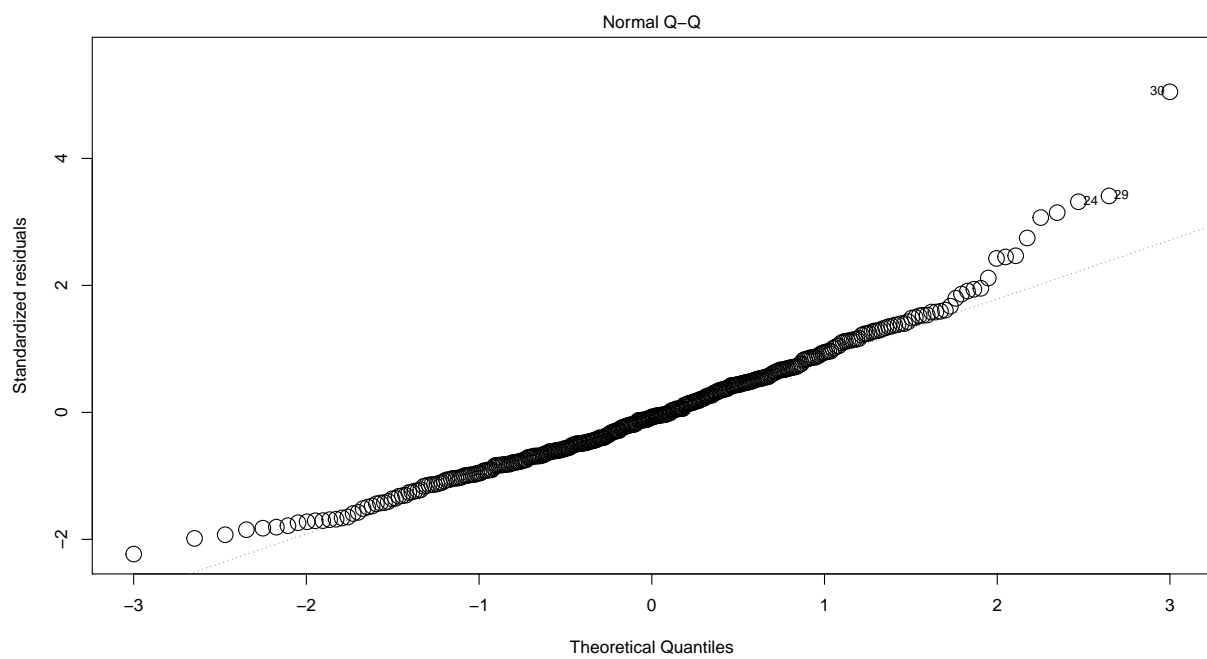
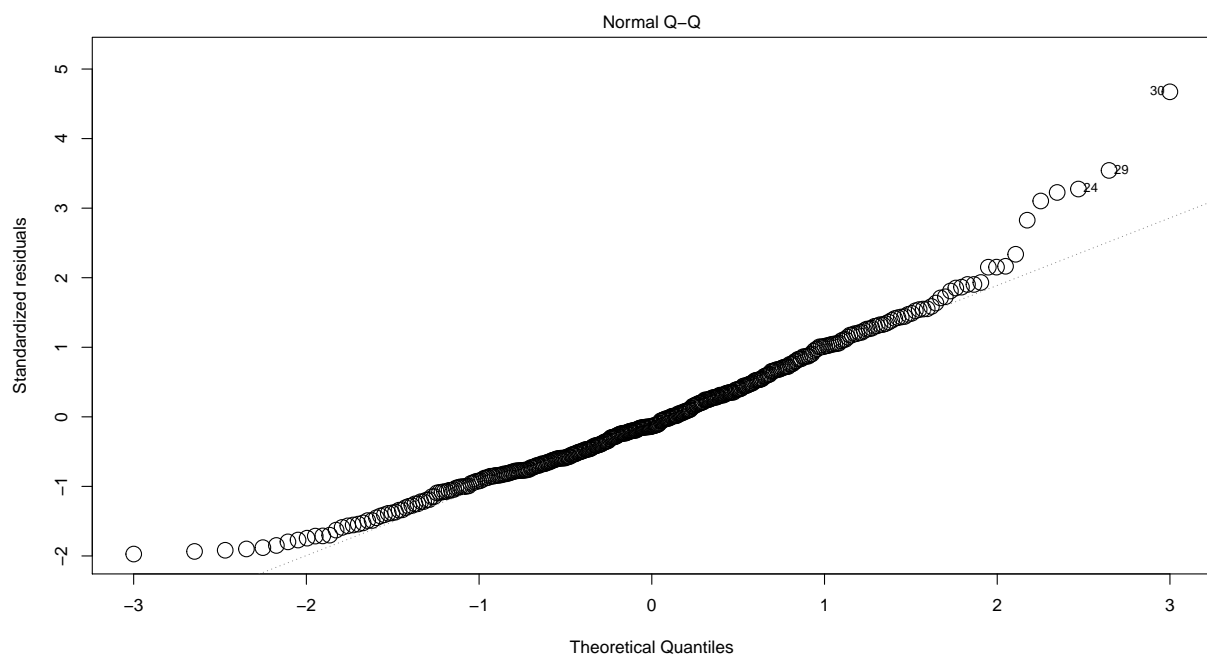
	Method 1	Method 2	Method 3	Method 4	Method 1,2,3 VIF	Method 4 VIF
(Intercept)	1.6698759	1.6698759	1.6698759	1.6694531		
SeedWidth	-0.1027270	-0.1027270	-0.1027270	-0.0865305	37.3815562	29.5080384
SeedLength	0.1536113	0.1536113	0.1536113		69.9871228	
SeedArea				-0.1470827		29.5080384
SeedAngle						
SeedHeight	-0.1470827	-0.1470827	-0.1470827		72.0248940	
SeedVol	0.0002749	0.0002749	0.0002749		14.8502466	
AIC	-1293.6000000	-1293.6000000	-1293.6000000	-1288.4000000	-1293.6000000	-1293.6000000
R <sup>2</sup>	0.3921194	0.3921194	0.3921194	0.3803212	0.3921194	0.3921194

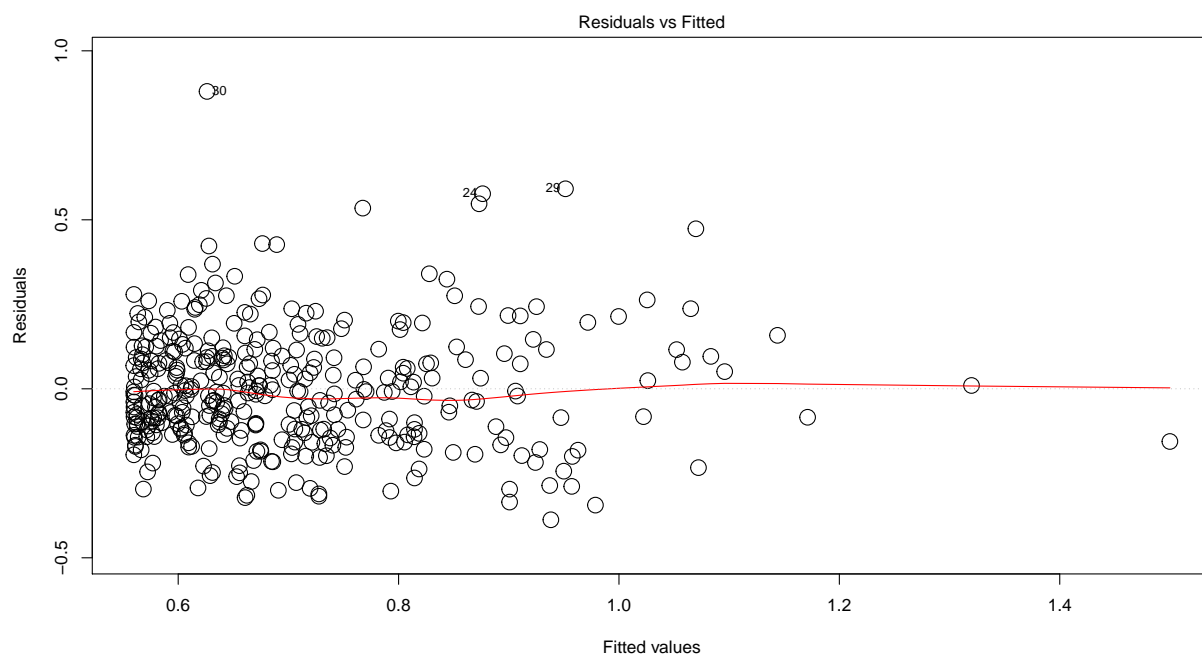
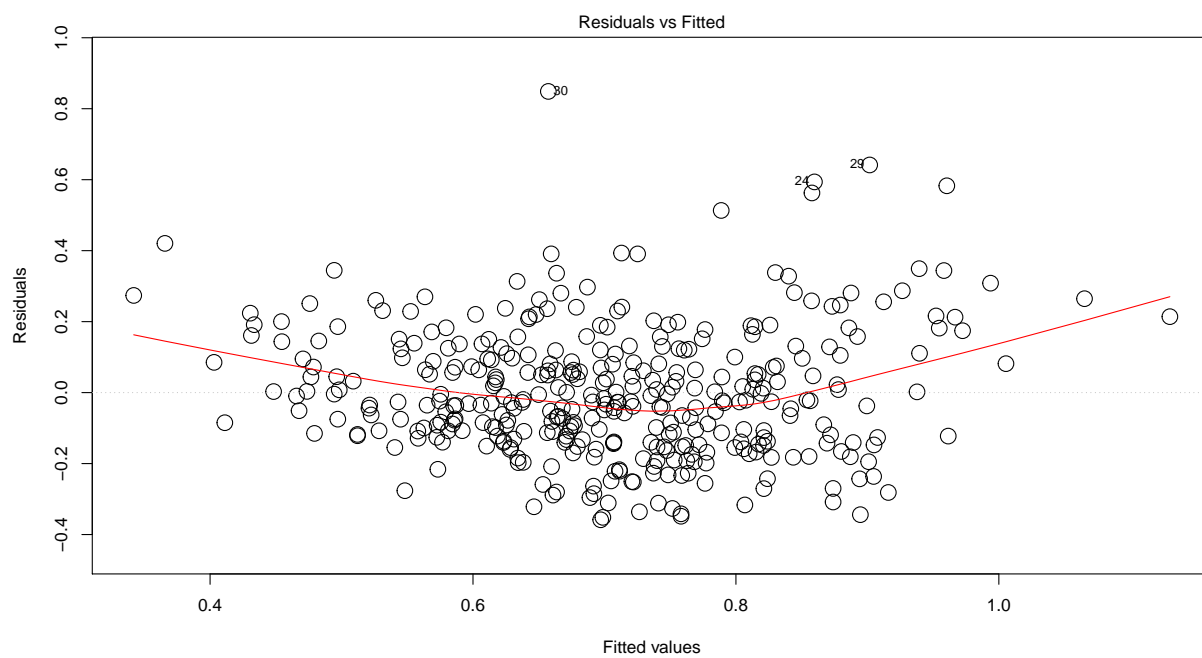
As we can see in the table, the first three selection methods produced the exact same result, while the fourth method produced a result with fewer terms. While the model derived from the first three methods has a higher R<sup>2</sup> and lower AIC than the model derived from the fourth method, it suffers greatly from high VIF. Because the VIF on the second model is much lower than that of the first and the difference in R<sup>2</sup> is very small, we will proceed with the second model, effectively trading a small increase in the amount of unexplained variance for a large decrease in multicollinearity.

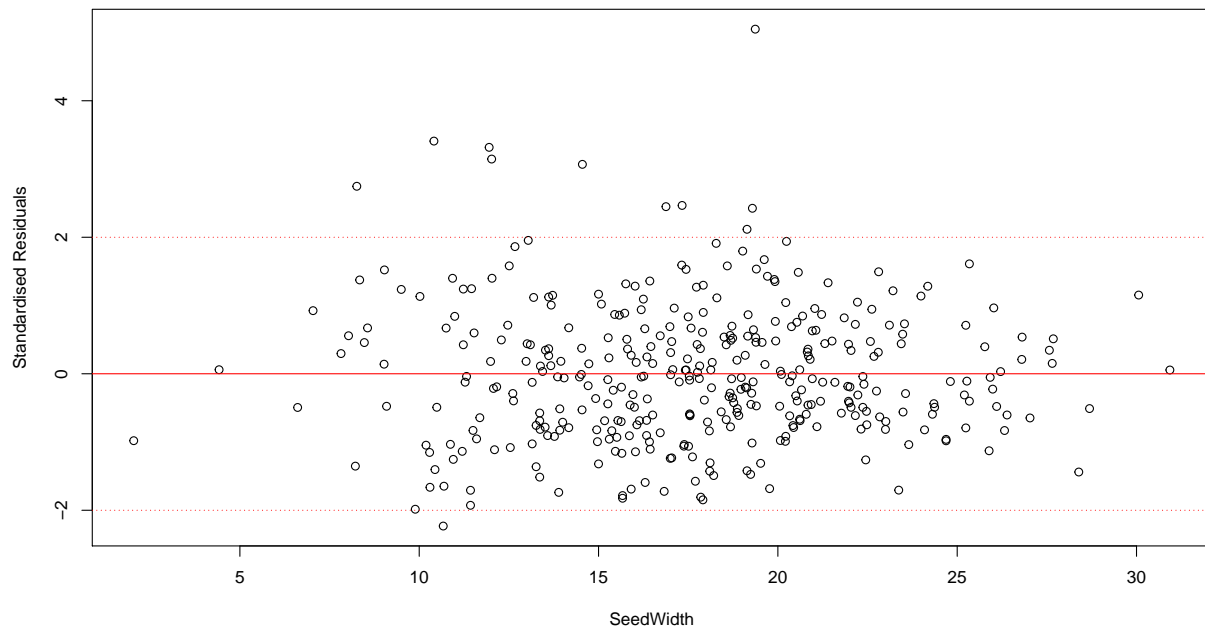
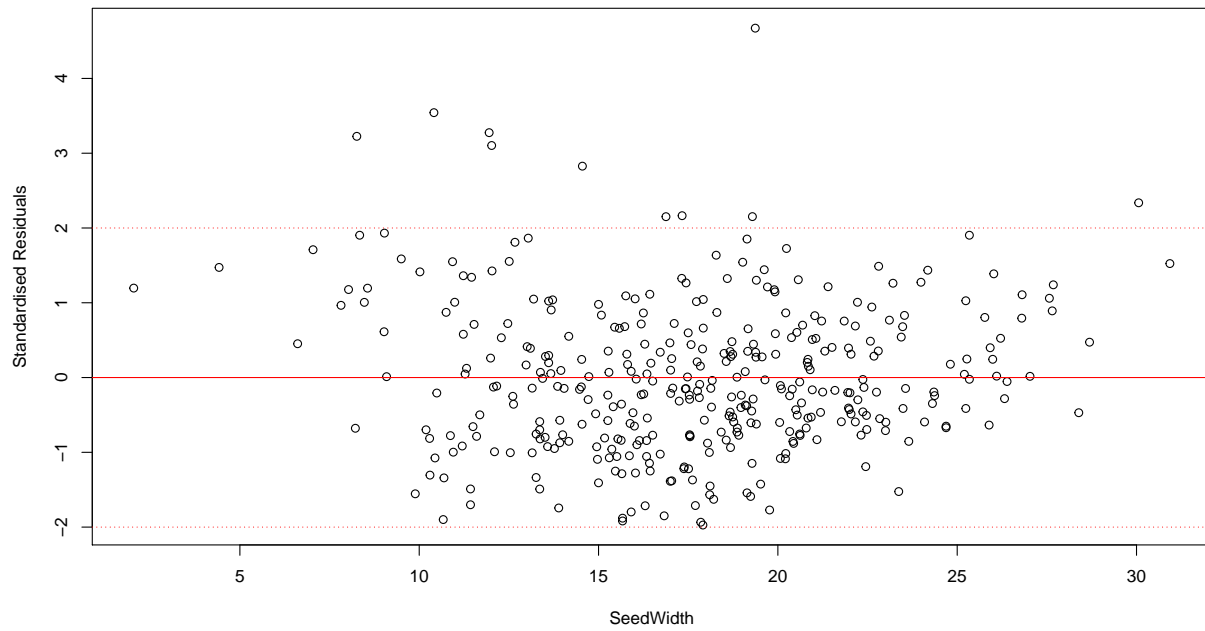
However, we still see a rather high VIF for the **SeedWidth** and **SeedArea** terms. While the **SeedArea** term may be inflating estimates when added to **SeedWidth**, a partial *F*-test confirms that its addition is statistically significant ( $F = 32.43$ ,  $p < 0.001$ ). Since **SeedArea** is proportional to the square of **SeedWidth**, we can simply express it as such in a polynomial regression instead of using a separate **SeedArea** variable for it.

Comparing this to a single-term model where VIF is not an issue, we see that a model with only **SeedWidth** is not as good of a fit, with an R<sup>2</sup> of 0.3274 compared to 0.3804 for the quadratic model. This is also clear when plotting the models against the data, as can be seen below:









## Research Question 2

Variable	AIC: Forward	AIC: Backward	AIC: Both	BIC: Forward	BIC: Backward	BIC: Both
Intercept	1	1	1	1	1	1
Mowing	0	1	1	0	1	1

Variable	AIC: Forward	AIC: Backward	AIC: Both	BIC: Forward	BIC: Backward	BIC: Both
Warming	0	1	1	0	1	1
Warming:Mowing	0	0	0	0	0	0

The following analysis will quantify the extent to which seed terminal velocity changed under warming and mowing treatments. Figure 1 shows the mean of each treatment. There are fewer points for late mowed plants, as many of them died due to the mowing treatment. To assess whether warming and mowing treatment significantly changed terminal velocity we used ANOVA models. We log-transformed terminal velocity to compensate for skewed right distribution of the terminal velocity data and significant deviations from error normality (Figure 2).

Using log-transformed terminal velocity, we use two methods to identify what is the most appropriate model to describe changes in terminal velocity based on the treatments.

## Research Question 3

## Discussion



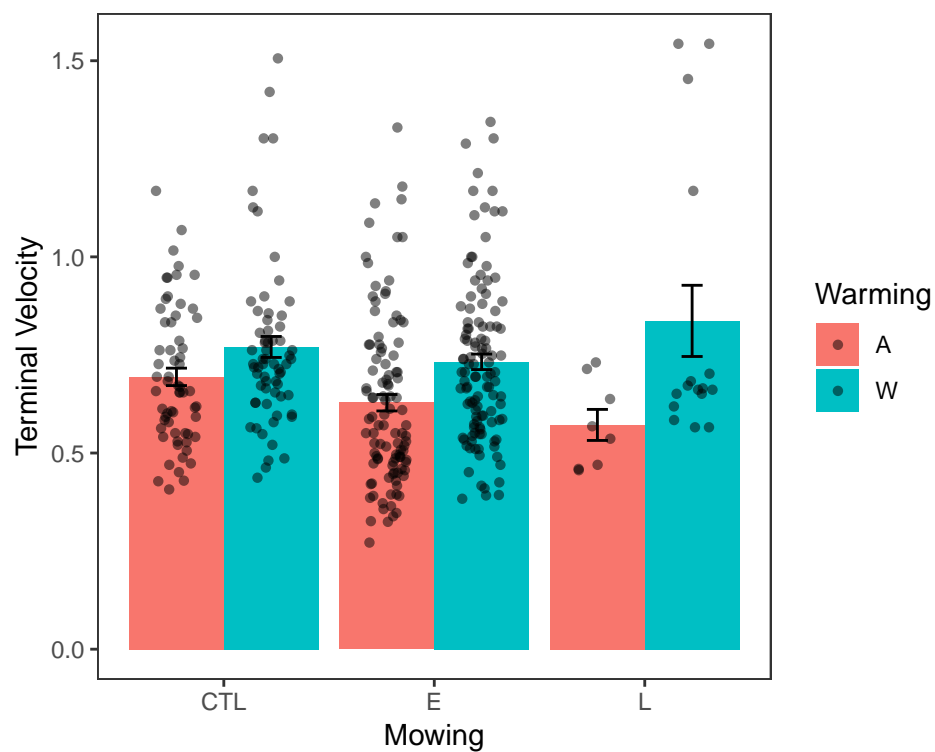


Figure 1: Terminal velocity for plants grown under warming and mowing treatments. Height of bar shows group mean, and error bars show one standard error. Actual data shown with black dots.

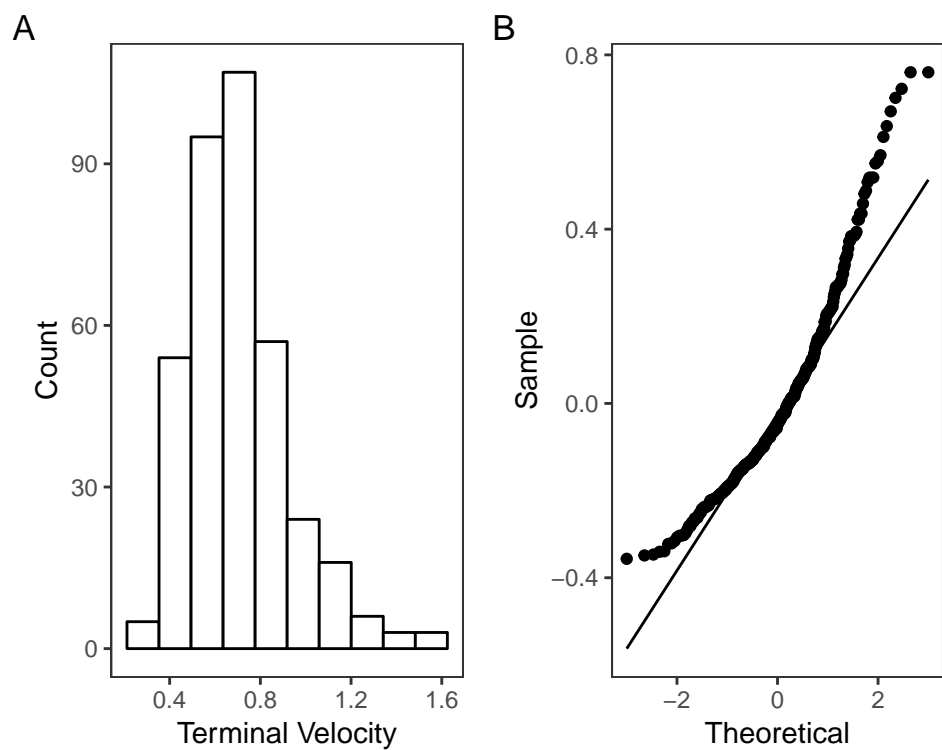


Figure 2: Justification for transformation of terminal velocity. Histogram of terminal velocity (A) and Normal-QQ plot of linear model, with both warming and mowing predictors.

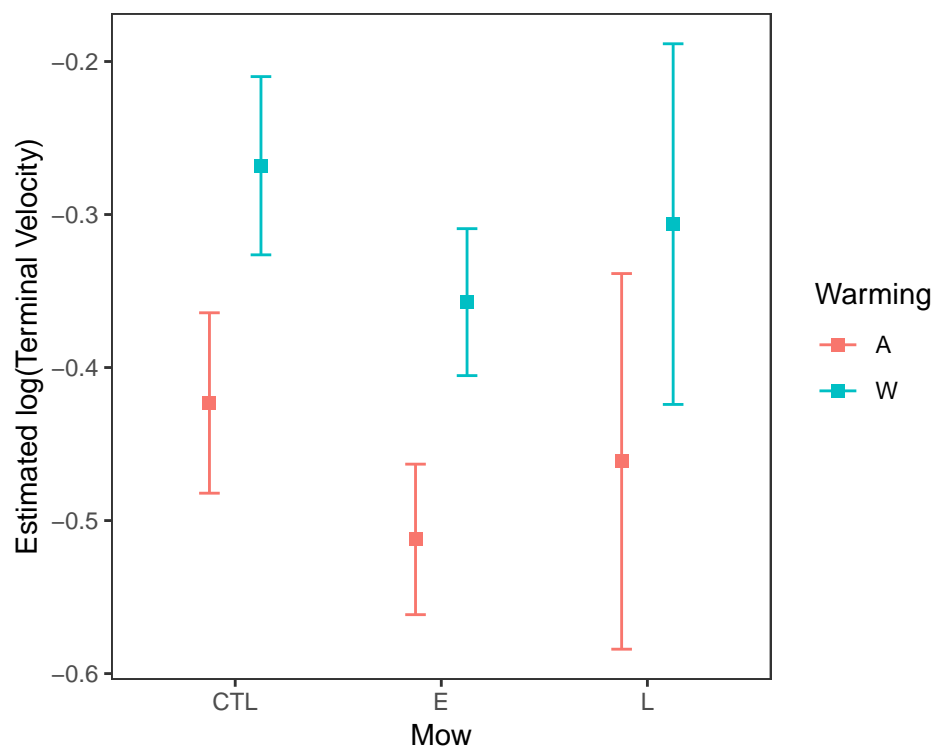


Figure 3: Results of best fitting ANOVA model, which included both warming and mowing as predictors. Square dots show mean predicted value and error bars show 95% confidence intervals.