2006

Modeling Inter-plant Interactions

Jessica Larson

University of Central Florida

Part of the Mathematics Commons

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

https://stars.library.ucf.edu/etd/1033
MODELING INTER-PLANT INTERACTIONS

by

JESSICA R. LARSON
B.S. University of Central Florida, 2004

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Mathematics
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Fall Term
2006
ABSTRACT

The purpose of this paper is to examine the interactions between two plant species endemic to Florida and develop a model for the growth of one of the plant species. An equation for the growth of *Hypericum cumulicola* is developed through analyzing how the distance to and the height of the nearest *Ceratiola ericoides* (Florida rosemary) affects the growth of *Hypericum cumulicola*. The hypericums were separated into five separate regions according to the distance to the nearest rosemary plant. The parameters for a basic growth equation were obtained in each of the five regions and compared to each other along with the average deviations in each of the five regions. Analysis of the five separate regions aided in the creation of different growth equations that each encompassed all of the regions together. Four different growth equations are developed and then compared and analyzed for their accuracy.
# TABLE OF CONTENTS

ABSTRACT................................................................................................................................... iii  
TABLE OF CONTENTS............................................................................................................... iv  
LIST OF FIGURES ........................................................................................................................ v  
LIST OF TABLES ......................................................................................................................... vi  
CHAPTER ONE: INTRODUCTION............................................................................................. 1  
CHAPTER TWO: LITERATURE REVIEW ................................................................................ 3  
CHAPTER THREE: METHODOLOGY ....................................................................................... 5  
CHAPTER FOUR: DATA ANALYSIS......................................................................................... 7  
CHAPTER FIVE: SUMMARY AND DISCUSSION ..................................................................... 28  
APPENDIX A: ESTIMATING ERRORS IN LINEAR FITS..................................................... 31  
APPENDIX B: MODEL ERROR ............................................................................................... 36  
LIST OF REFERENCES ............................................................................................................ 38
LIST OF FIGURES

Figure 1: Scrub habitat gap. ............................................................................................................. 2

Figure 2: Final height of the hypericums in region RM5 versus their initial heights after one year. .................................................................................................................................................. 7

Figure 3: The model given by equation (1) superimposed on the data from region RM5. ........... 10

Figure 4: The model given by equation (2) superimposed on the data from region RM5. .......... 10

Figure 5: The average value of $\alpha$ and its expected variation in each region versus the average distance for each region. ........................................................................................................ 12

Figure 6: The average value of $\beta$ and its expected variation in each category versus the average distance of each region........................................................................................................... 13

Figure 7: Three models of $\beta$ versus average distance: the fitted smooth function, the stepwise function, and the linear piecewise-continuous function......................................................... 15

Figure 8: Equation (8) Final Height versus Initial Height............................................................. 18

Figure 9: Equation (7) Final height versus Initial height............................................................. 19

Figure 10: Equation (6) Final Height versus Initial Height........................................................... 19

Figure 11: Equation (4) Final Height versus Initial Height............................................................ 20

Figure 12: Equation (6) $dH_f$ versus distance to the nearest rosemary......................................... 21

Figure 13: Equation (4) $dH_f$ versus distance to the nearest rosemary......................................... 22

Figure 14: Equation (6) $dH_f$ versus rosemary height............................................................... 25

Figure 15: Equation (4) $dH_f$ versus rosemary height............................................................... 26
# LIST OF TABLES

Table 1: Distance regions........................................................................................................................................ 6

Table 2: Initial model comparison....................................................................................................................... 9

Table 3: $a$, $b$, their expected variations ($\delta a$ and $\delta b$), the square of the average deviation ($\varepsilon^2$),
   and the average distance between the hypericum and its closest rosemary for the 5 distance
   regions................................................................................................................................................................... 12

Table 4: Average deviations and expected variations for the four models considered. .................. 18

Table 5: The slopes of the deviations versus the distance to the nearest rosemary and their
   expected variations for Equations (6) and (4) in each of the 5 regions....................................................... 23

Table 6: The slopes of the deviations versus the height of the nearest rosemary and their
   expected variations for Equations (6) and (4) in each of the 5 regions....................................................... 27
CHAPTER ONE: INTRODUCTION

The purpose of this paper is to mathematically analyze the spatial relationship between two Florida scrub plants: scrub hypericum (*Hypericum cumulicola*) and Florida rosemary (*Ceratiola ericoides*). Particularly, we analyze the relationships between the heights of the individual plants, the distance between a scrub hypericum and its closest rosemary neighbor, and the growth of the hypericum.

*H. cumulicola* is a small herbaceous species whose maximum height is approximately 70 cm while Florida rosemary is a relatively large shrub with a maximum diameter of about 3 m. Thus, we will study relationships between a large plant and a smaller plant. We collected field data and determined how the distance between the rosemary and hypericum and rosemary size affect the growth of the hypericum.

Scrub habitat is an ecosystem in which fire is a regular occurrence and a major influence on the vegetation (Menges and Kohfeldt 1995). An important characteristic of Florida scrub is that there are definite areas of low-density foliage, called gaps, dispersed throughout an overall area of high-density foliage (Quintana-Ascencio *et al.* 1998).
The gaps typically range in size from 2m to 14m in length and width. Gaps decrease in size as time-since-fire increases because, in the absence of fire, plants are growing back. *H. cumulicola* almost only occurs within these open areas, gaps, while rosemary is the dominant species in the matrix vegetation and also occurs both within the gaps (Quintana-Ascencio *et al.* 1998). Populations of *H. cumulicola* within a gap tend to decline with time-since-fire (Menges and Kohfeldt 1995). As rosemary and other plants recover with time-since-fire, the gaps become overgrown thereby reducing the open spaces required for the survival of *H. cumulicola* (Quintana-Ascencio *et al.* 1998). This study of the growth of *H. cumulicola* is important because it provides us with information on how plant communities organize in dynamic systems. Here, rosemary is a dominant shrub species within the scrub habitat and *H. cumulicola* is almost entirely restricted to inter-shrub gaps (Quintana-Ascencio *et al.* 2004).
CHAPTER TWO: LITERATURE REVIEW

The effects of disturbance on population dynamics and habitat are important when considering proper conservation and management practices (Quintana-Ascencio and Morales-Hernández 1997). Fire is a natural disturbance in the rosemary phase of sand pine scrub with this type of community burning approximately every 20-80 years (Quintana-Ascencio and Morales-Hernandez 1997). When a community is burned by fire, Florida rosemary individuals are killed along with most of the lichen groundcover leaving open sandy patches that serve as micro-habitats for many endemic herb species like *H. cumulicola* (Quintana-Ascencio and Morales-Hernandez 1997). As scrub habitat recovers with time-since-fire, rosemary shrubs and lichens grow back recovering their populations and decreasing the size and frequency of open sandy patches. This leads to a decline in herbaceous species (Menges and Kohfeldt 1995). Survival and growth of many scrub endemics such as hypericum is higher in more recently burned patches (Quintana-Ascencio and Morales-Hernandez 1997). In this case, the relationship between rosemary and *H. cumulicola* can be used to determine optimum time intervals for prescribed controlled burns.

A former spatially explicit simulation considered a variety of characteristics regarding both *H. cumulicola* and rosemary (Quintana-Ascencio, unpublished). Data included in this model consisted of survival, growth, recruitment, plant size, conspecific density and distance to the nearest rosemary for the hypericum and rosemary growth and size. Empirical functions for survival, growth, fecundity, and distribution of plants were developed using the software package, SPSS, and then used in Matlab to create a simulation of the spatial patterns of hypericum and rosemary within a gap. Here we go in a different direction. We constructed a
mathematical model of the growth of hypericum and seek to determine how its growth (height) is affected by the distance to the nearest rosemary and the height of that rosemary.

Ecology is the study of the relations and interactions between organisms and their environment. In attempting to model biological populations, a commonly used method is one where the models are based on individuals because “individuals differ from each other, even within the same species and age, so each interacts with its environment in unique ways” (Grimm, 3). “[All] that an individual does – grow, develop, acquire resources, reproduce, interact – depends on its internal and external environments” (Grimm, 3), so it is important to note that in modeling this particular biological phenomenon, a certain degree of seeming randomness can always be expected to be present. The seeming randomness exhibits itself through individual variability, called demographic stochasticity and variation in environmental factors, called environmental stochasticity (Haccou, 5).

In addition to individual variations, the environment also impinges on the individuals. Thus in examining the growth of hypericum, there will always be a certain spread in the data because it is “impossible to record all the conditions that have repercussions on the life career of an individual” (Haccou et al., 4). In other words, two plants that have the same initial height may not have the same final height due to the combination of the stochastic nature of the individuals and the stochastic environmental conditions in which they exist. Some examples of these factors that can influence individuals differently include genetic differences between plants, nutrient availability to plants, and herbivore activity on the plants. These are just a few examples of an extensive list of influencing features in the natural world.
CHAPTER THREE: METHODOLOGY

Throughout this analysis of the growth of hypericum, we used field data collected between 1994 and 1995 at Archbold Biological Station on 1057 individuals (Quintana-Ascencio, et al., 2004). The collected data included the height of the hypericum at the beginning of the year (this will be referred to as the initial height), the final height after one year, survival after one year, the distance to the nearest rosemary and its height, and the site location. Other information about the hypericums, including survival, number of reproductive structures, and site location, was collected as well, but the additional information was not used in this study. Out of the 1057 plants, only plants which survived the entire year (and thus had a measured final height) were used. This left a total of 652 hypericums.

The remaining hypericums were then split into five regions according to their distance to the nearest rosemary. These regions were hypericums where the distance from the hypericum to the nearest rosemary fell between 0-50cm, 50-100cm, 100-200cm, 200-300cm and over 300cm. The five different regions were labeled as in Table 1. We assumed for this study that the hypericums in regions RM5 were sufficiently far away so as to be not influenced by any rosemary plants.

The 652 hypericums were the only collected data analyzed during the study. In the following chapter, the approach in treating this data in order to model the final height of a hypericum after one year, given its initial height and the distance to the closest rosemary and the height of that rosemary is described.
Table 1: Distance regions.

<table>
<thead>
<tr>
<th>Distance Regions</th>
<th>Distance (cm)</th>
<th>Number of Plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM1</td>
<td>1-50</td>
<td>55</td>
</tr>
<tr>
<td>RM2</td>
<td>51-100</td>
<td>84</td>
</tr>
<tr>
<td>RM3</td>
<td>101-200</td>
<td>147</td>
</tr>
<tr>
<td>RM4</td>
<td>201-300</td>
<td>73</td>
</tr>
<tr>
<td>RM5</td>
<td>301 and above</td>
<td>293</td>
</tr>
</tbody>
</table>
CHAPTER FOUR: DATA ANALYSIS

In analyzing the relationship between the growth of hypericum and the distance of the hypericum to its closest rosemary neighbor, we first examined the hypericums that were in the RM5 region (greater than 300cm from the closest rosemary). Analysis of this region was assumed to show the growth trends in hypericum when it is not influenced by rosemary. By plotting the final height of the hypericum versus the initial height, we could then see the general trend of the relationship between the initial height and the growth rate of the hypericum.

Figure 2 shows the relationship between the final height of the hypericum versus its initial height for the RM5 region plants.

![Figure 2: Final height of the hypericums in region RM5 versus their initial heights after one year.](image)

The observed scatter in the data points is assumed to be due to individual variations and/or environmental effects, as mentioned earlier. In spite of this, one can still see that the growth
trends in hypericum were such that smaller plants have a higher relative growth rate than larger plants. Simply from the general slope seen in the data, one notes that plants with a smaller initial height experience a larger percentage of growth than plants with higher initial heights. One notes that the general trend is to level off at some maximum height. Figure 2 shows that this maximum height for hypericum is about 70cm and the plants tend to level out or saturate around this peak height. This result is not surprising. Many living organisms start off small and grow to a maximum height at which they level off, due to trade-offs between growth/survival and reproduction (Silvertown and Charlesworth, 272). For example, reproducing at a smaller size (younger age) may slow growth and increase the risk of death while not producing a high number of reproductive structures. Reproduction has a high cost because it requires energy which may otherwise be used by a plant for growth and survival. For different species, there are different optimum ages and sizes at which a plant stops putting most of its energy into growth and survival and starts concentrating its efforts on reproduction (Silvertown and Charlesworth, 273). For hypericum, the data indicates that this height is around 40cm-70cm.

From Figure 2, looking at the overall shape, one sees that one could model this trend in the data by a function like a square root (or any power less than unity) or a logarithm. So, we begin by determining which such function would come closest to fitting the observed data. Letting $H_f$ be the final height of the hypericum (cm) and $H_o$ the initial height of the hypericum (height at the beginning of the year), we examined both a power model, (1), and a logarithm model, (2):

$$H_f = b \ast (H_o)^v \quad (1)$$
\[ H_f = a \ln(H_o) + b \quad (2) \]

where \( a \) and \( b \) are constants to be determined. A least squares analysis was used to find the parameters that minimized the average deviation for both models in the RM5 region. The following values were found to minimize the average deviation, where \( \varepsilon^2 \) is the square of the average deviations:

Table 2: Initial model comparison

<table>
<thead>
<tr>
<th>Equation</th>
<th>( a )</th>
<th>( b )</th>
<th>( \varepsilon^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.30</td>
<td>15.8</td>
<td>86.46</td>
</tr>
<tr>
<td>(2)</td>
<td>9.9</td>
<td>10.1</td>
<td>85.43</td>
</tr>
</tbody>
</table>

Note that the above least square fits were obtained only for the plants in the region RM5.

In Figures 3 and 4 below, the best fit for each model is shown superimposed on all plants in region RM5.
Figure 3: The model given by equation (1) superimposed on the data from region RM5.

Figure 4: The model given by equation (2) superimposed on the data from region RM5.
As one can see, the two fits are similar except near the origin where the second fit dips somewhat lower than the first. Because of this, it appears that the second fit more closely represents the actual data in this region. The average deviations for equations (1) and (2) are essentially equivalent since if we use $\frac{1}{\sqrt{N}} (N = 293)$ as an estimate of the relative expected variation ($\delta R$) in the observed average deviations, we find that the numerical differences between the average deviations in the two models is only about 10% of the expected variation in the average deviations. However, since the average deviation for equation (2) was slightly smaller and the dip in the predicted heights for small heights appears to more closely resemble the actual data, this model was chosen for our background model.

Now that a backbone model has been chose, let us see how well it can be fitted to the other regions. Using equation (2), a least squares analysis was carried out to find $a$ and $b$ for each of the 4 remaining distance regions, along with their expected variations, $\delta a$ and $\delta b$. The purpose of this was to look for possible distance dependencies in these coefficients. The results of this analysis are given in Table 3 and are shown in Figures 5 and 6. What one observes in Table 3 is that the average deviations ($\epsilon^2$) themselves demonstrate a definite dependence on the distance to the nearest rosemary because there are larger deviations the closer the rosemary is to the hypericum. The increase in this spread in the average deviations as one goes toward the smaller distance regions signifies that there is very likely some influence which is being exerted by the rosemary on the hypericums at smaller distances which is absent in the RM5 region and which is not being accounted for in the current model.
Table 3: $a$, $b$, their expected variations ($\delta a$ and $\delta b$), the square of the average deviation ($\varepsilon^2$), and the average distance between the hypericum and its closest rosemary for the 5 distance regions.

<table>
<thead>
<tr>
<th>Category</th>
<th>$a$</th>
<th>$B$</th>
<th>$\delta a$</th>
<th>$\delta b$</th>
<th>$\varepsilon^2$</th>
<th>Average Distance (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM1</td>
<td>10.2</td>
<td>-0.3</td>
<td>2.07</td>
<td>6.13</td>
<td>139.22</td>
<td>31</td>
</tr>
<tr>
<td>RM2</td>
<td>11.5</td>
<td>-2.3</td>
<td>1.67</td>
<td>5.19</td>
<td>140.28</td>
<td>75</td>
</tr>
<tr>
<td>RM3</td>
<td>9.8</td>
<td>5.4</td>
<td>1.14</td>
<td>3.63</td>
<td>116.36</td>
<td>152</td>
</tr>
<tr>
<td>RM4</td>
<td>8.6</td>
<td>12.6</td>
<td>1.58</td>
<td>4.98</td>
<td>100.13</td>
<td>244</td>
</tr>
<tr>
<td>RM5</td>
<td>9.9</td>
<td>10.1</td>
<td>0.56</td>
<td>1.46</td>
<td>85.43</td>
<td>549</td>
</tr>
</tbody>
</table>

Figure 5: The average value of $a$ and its expected variation in each region versus the average distance for each region.

From Figure 5, it appears that there is probably little change in the parameter $a$ as the distance changes. First, one observes that all the error bars enclose a common set of values.
around 10. Thus the variations in \( a \) that seem to be here could also easily be explained to be due to the stochastic nature of the data. It is also possible that the parameter \( a \) could actually be larger for distances closer to the rosemary. However based on the data in Figure 5 and due to the large deviations observed in the smaller distance regions, the best that one could say would be that its value appears to be relatively insensitive to the distance to the nearest rosemary.

Next, we consider the parameter, \( b \), as shown in Figure 6. This figure illustrates that the parameter \( b \) is probably significantly affected by the nearest rosemary plant because here, the changes in its value as a function of \( d \) are beyond the expected variations.

![Graph showing \( b \) vs. Average Distance](image)

Figure 6: The average value of \( b \) and its expected variation in each category versus the average distance of each region.

Clearly Figure 6 suggests that \( b \) is some function of \( d \). A general function for the shape seen in Figure 6 would be
\[ b(d) = \alpha + \frac{\beta d}{1 + \frac{d}{\gamma}} \]  \hspace{1cm} (3)

where \( d \) is the distance to the nearest rosemary and \( \alpha, \beta, \) and \( \gamma \) are constants to be determined.

The parameter, \( \alpha \), represents an average starting value while \( \beta \)'s purpose is to increase the function as \( d \) increases and \( \gamma \)'s purpose is to flatten out the function once \( d \) is sufficiently large.

Using least squares to determine the best values for these coefficients and also the coefficient \( a \) which appears in Equation (2), and using all plants from all categories, the best fit was found to be \( a = 9.8, \alpha = -6.28, \beta = 0.16, \) and \( \gamma = 143.18 \) and is shown as the smooth curve in Figure 7. For this fitted curve, the growth equation (2) becomes

\[ H_f = 9.8 \ln(H_o) - 6.28 + \frac{0.16d}{1 + \frac{d}{143.18}} \]  \hspace{1cm} (4)

Two other functions were also tested to model the dependence of \( b \) on \( d \): a stepwise function and a linear piece-wise continuous function, both of which are also shown in Figure 7 and will be described shortly.
As one can see, the data suggests that there may well be a peak of some form in the curve around 200-300cm. To test out the possibility that the data might support the presence of a slight peak in this range, two other models were tried, each of which would have a slight peak in this range already built in. The piecewise-continuous and stepwise functions were both examined to explore the possibility of a peak in the RM4 region. However, recall that Figure 6 was obtained by breaking the data into the regions RM1 – RM5, and finding the best fit inside each region. Therefore the presence of the peak may be misleading because the different $b$ values correspond to differing $a$’s. In particular, the value of $a$ determined by the fit at the peak of $b$ ($d=250cm$) is lower in value than the other values of $a$. Thus the apparent peak in $b$ could well simply be a result of attempting to compensate for the lower value of $a$. On the other hand, the smooth curve shown in Figure 7, and given in Equation (4), was obtained by finding the best fit of the assumed curve, Equation (3), over the fitted values from all regions. Thus this curve should be expected
to have some universal validity. On the other hand, one could design a function for a curve
which would have such a slight peak, but only at the cost of additional parameters. Since the
data appears to be equivocal in this matter at this moment, the best choice appeared to be to test
out the following two models instead: a linear piece-wise continuous curve and a step-wise
curve based on the fitted values of $a$ and $b$ shown in Figures 5 and 6. The linear piece-wise
continuous curve shown in Figure 7 was constructed from the data points in Figure 6 and
consists of the three lines,

$$
b(d) = \begin{cases} 
-0.45d + 1.11 & 0 \leq d \leq 75 \\
0.088d - 8.56 & 76 \leq d \leq 244 \\
-0.008d + 14.6 & 245 \leq d 
\end{cases} \quad (5)$$

In this case, with $b(d)$ already determined, there was only the parameter $a$ to be determined, and
carrying out a least squares fit over all the data gave $a = 10.0$. So the second model for $b$ as a
function of $d$, is

$$
H_f = 10 \ln(H_o) + b(d) \\
$$

where $b(d) = \begin{cases} 
-0.45d + 1.11 & 0 \leq d \leq 75 \\
0.088d - 8.56 & 76 \leq d \leq 244 \\
-0.008d + 14.6 & 245 \leq d 
\end{cases} \quad (6)$

Another function that would test for a peak was the stepwise fit shown in Figure 7.
Carrying out another least squares fit as above, it was found that the growth function becomes:
\[ H_f = 10 \ln(H_o) + b(d) \]

where \( b(d) = \begin{cases} 
-0.3 & 0 \leq d \leq 53 \\
-2.3 & 54 \leq d \leq 114 \\
5.4 & 115 \leq d \leq 198 \\
12.6 & 199 \leq d \leq 397 \\
10.1 & 398 \leq d 
\end{cases} \) \hspace{1cm} (7)

where the least squares value for \( a \) for this model was found to also be equal to 10.0 as in the previous case.

Lastly, as a base line comparison, a least squares fit was made by taking Equation (2) with both \( a \) and \( b \) as constants and independent of \( d \). In this case, \( a \) was found to be 9.2 and \( b \) was found to be 8.9, making the fourth growth model to be

\[ H_f = 9.2 \ln(H_o) + 8.9 \] \hspace{1cm} (8)

Below in Table 4, the results are summarized for these four models, wherein the average deviation squared, the expected variation in the average deviation, and the expected variation in \( a \), were calculated. Equation (4) has the smallest \( \varepsilon^2 \), so this is the best model to fit the data given. The following four figures show the expected final height (black) and the actual final height (color) versus the initial height for each of the four models. What one observes in Figure 8 is that the simplest model is just a single curve passing through the data. As we add features to the model, there are more sections to this curve and it becomes more accurate since the predicted
values (black) can begin to spread throughout the data, thus becoming more successful in reducing the average deviations.

Table 4: Average deviations and expected variations for the four models considered.

<table>
<thead>
<tr>
<th>Equation</th>
<th>$\varepsilon^2$</th>
<th>$\varepsilon$</th>
<th>$\delta \varepsilon$</th>
<th>$a$</th>
<th>$\delta a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4)</td>
<td>107.83</td>
<td>10.38</td>
<td>0.41</td>
<td>9.8</td>
<td>0.44</td>
</tr>
<tr>
<td>(6)</td>
<td>112.28</td>
<td>10.60</td>
<td>0.42</td>
<td>10</td>
<td>0.45</td>
</tr>
<tr>
<td>(7)</td>
<td>115.81</td>
<td>10.76</td>
<td>0.42</td>
<td>10</td>
<td>0.46</td>
</tr>
<tr>
<td>(8)</td>
<td>116.92</td>
<td>10.81</td>
<td>0.42</td>
<td>9.2</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Figure 8: Equation (8) Final Height versus Initial Height.
Figure 9: Equation (7) Final height versus Initial height.

Figure 10: Equation (6) Final Height versus Initial Height.
The model with the smallest average deviations is shown above in Figure 11 and appears to be the best functional representation of the growth of the hypericum at the present time.

The next step was to see if there is anything which can indicate possible parametric regimes not covered in the current model. One way to see how well the model covers the functional dependence on the distance to the nearest rosemary and the rosemary height would be to plot the deviations (the difference in the actual final height and the predicted final height) from the model predictions as a function of these variables. For a good model, one would expect a uniformity in the deviations as a function of these variables.

In Figures 12 and 13, the deviations in the five categories for Equations (6) and (4) are shown as a function of the distance to the nearest rosemary. For each of these plots, a linear fit to the deviations was made to determine the average slope and its expected variation, which are tabulated in Table 5.
Figure 12: Equation (6) $dH_f$ versus distance to the nearest rosemary.
Figure 13: Equation (4) $dH_f$ versus distance to the nearest rosemary.
Table 5: The slopes of the deviations versus the distance to the nearest rosemary and their expected variations for Equations (6) and (4) in each of the 5 regions.

<table>
<thead>
<tr>
<th>Category</th>
<th>Slope and expected variation for Equation (6)</th>
<th>Slope and expected variation for Equation (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM1</td>
<td>0.092 ± 0.135</td>
<td>-0.065 ± 0.139</td>
</tr>
<tr>
<td>RM2</td>
<td>0.071 ± 0.106</td>
<td>0.031 ± 0.102</td>
</tr>
<tr>
<td>RM3</td>
<td>-0.129 ± 0.033</td>
<td>-0.078 ± 0.032</td>
</tr>
<tr>
<td>RM4</td>
<td>0.043 ± 0.048</td>
<td>0.052 ± 0.046</td>
</tr>
<tr>
<td>RM5</td>
<td>0.028 ± 0.007</td>
<td>0.010 ± 0.007</td>
</tr>
</tbody>
</table>

Looking at Table 5, one notes that to within the expected variation, the slopes found for RM1 and RM2 are equivalent to zero. This means that the data does not indicate that any further corrections are needed to model the dependence on $d$ in these regions. However, one notices in RM3 that there is a definite tendency for the slope to be nonzero and for RM4 to be slightly positive; this indicates that in the region between RM3 and RM4, the dependence of $b$ on $d$ could possibly be more positive than that in the two models. In regions RM3 and RM4 there is a slight depression around 200; this may have something to do with the possible peak in the parameter, $b$, noticed earlier in Figure 6. However, the average slopes for Equation (4) are no more than two standard deviations beyond their expected variations, so the best that one could say is that it is likely that some additional structure may be present in $b(d)$. Note that even though the RM5 region is listed in the table, care should be used in considering it because if the distance to the nearest rosemary was greater than 400cm, the distance was not recorded (for plotting purposes, all data with distances greater than 400cm were arbitrarily placed at 575 cm).
Let us now consider the possibility of effects due to the heights of the rosemary plants. For this purpose, in Figures 14 and 15, we have plotted the deviations in each of the five regions for Equations (6) and (4) as a function of the height of the nearest rosemary. Then as before, we have made a linear fit to these deviations to determine the slope of that fit and its expected variation which are now tabulated in Table 6.
Figure 14: Equation (6) $dH_f$ versus rosemary height.
Figure 15: Equation (4) $dH_f$ versus rosemary height.
Table 6: The slopes of the deviations versus the height of the nearest rosemary and their expected variations for Equations (6) and (4) in each of the 5 regions.

<table>
<thead>
<tr>
<th>Category</th>
<th>Slope and expected variation for Equation (6)</th>
<th>Slope and expected variation for Equation (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM1</td>
<td>0.018 ± 0.031</td>
<td>0.018 ± 0.032</td>
</tr>
<tr>
<td>RM2</td>
<td>0.014 ± 0.029</td>
<td>0.014 ± 0.03</td>
</tr>
<tr>
<td>RM3</td>
<td>0.021 ± 0.017</td>
<td>0.022 ± 0.016</td>
</tr>
<tr>
<td>RM4</td>
<td>0.013 ± 0.023</td>
<td>0.012 ± 0.021</td>
</tr>
<tr>
<td>RM5</td>
<td>-0.047 ± 0.016</td>
<td>-0.015 ± 0.016</td>
</tr>
</tbody>
</table>

What one observes here is that there is a possibly significant nonzero slope only in categories RM3 and RM5 (once again, care should be used in the consideration of the category RM5 because the height of the nearest rosemary was not recorded if the plant was more than 400cm away). The positive slope for category RM3 indicates that in RM3 (where the average distance to the nearest rosemary was 152 cm), \( b \) could be decreasing somewhat as a function of the rosemary’s height. However the statistics are such that a definite conclusion cannot be reached since the magnitudes of the slopes are less than about 1.5 times the expected variation.
CHAPTER FIVE: SUMMARY AND DISCUSSION

With the goal of developing an equation to model the growth of hypericum, the initial heights and final heights after one year for multiple hypericums were analyzed to come up with an initial structure for a growth equation. Each of the five distance regions was initially analyzed separately and then their parameters and average deviations compared. From the comparison of the differences in the average deviations in each of the five regions, it was clear that there is a definite dependence on the distance to the nearest rosemary because there are larger deviations the closer the rosemary is to the hypericum. From the comparisons of the parameters, it was clear that there was a definite influence on the parameter, \( b \), and a possible influence on parameter, \( a \), by the distance to the nearest rosemary. Based on the statistics, it was decided to consider the parameter, \( a \), to be unaffected by the distance to the nearest rosemary and to incorporate the influence of the distance entirely into parameter \( b \). Through this action, it was intended that the model for the growth would be improved, as was the case. The distance, \( d \), was incorporated into the equation by making \( b \) a function of \( d \). Four different ways of constructing \( b \) as a function of \( d \) were examined. The average deviations for each of the models were calculated and it was determined that Equation (4) was the best model with \( \varepsilon^2 = 107.83 \) for that model. The other models were not that significantly different, but since the \( b \) function for Equation (4) seemed to more accurately model the actual data and what was actually happening with the rosemary’s influence on hypericum, this model is probably the best of those considered.

In the initial analysis of \( b \) from Figure 5, it was noted that there was possibly a peak in the \( b \) parameter (and therefore, a possible peak in growth) in region RM4. However, in comparing the average deviations for the different models, the models incorporating the peak had
higher average deviations than the model which did not incorporate the peak. Further analysis of this model indicated that there were possibly additional distances-influences around a distance of 200 cm between hypericum and its nearest rosemary neighbor. These influences quite probably may be related with the observed peak in $b$ mentioned earlier and further work could be done in examining this possibility.

In exploring the parameters, $a$ and $b$, it was clear that each of them affect the final expected height of a hypericum. However, whereas a change in $a$ would produce different changes in the expected final height depending on the initial height, a change in $b$ would produce the same change regardless of the initial height. Thus in comparing Figures 5 and 6, this is suggesting that the change in the expected final height of any hypericum due to the presence of a rosemary plant would be the same amount regardless of the initial height of the hypericum. Note that if $b$ was constant and $a$ changed, the effect of a given rosemary plant on the growth would be different for different heights. We have no possible explanation for this observation except that the data support it.

As was mentioned before, there are larger average deviations in the parameters for smaller distances between the hypericum and the closest rosemary. This indicates that there may be some additional factors concerning the rosemary which may have additional influences on the growth of the hypericum. There are several directions that could be explored here. An important avenue to explore would be the number and the various heights of different rosemary plants that are within a certain distance to a hypericum plant. As was noted early, only the distance from the hypericum to the nearest rosemary was recorded. The exclusion of other nearby rosemary plants which would be within some certain distance to the hypericum could account for the
larger average deviations present in the regions RM1 and RM2. Thus the collection of additional data of this nature could possibly shed some light on these larger deviations.

The model which has been developed here, Equation (4), predicts the height of hypericum after one year given its initial height and the distance to the nearest rosemary. However, these two variables are probably not the only factors influencing the growth of hypericum. This possibility is indicated in Figures 12 and 13 in which the differences in the predicted final heights and actual final heights were plotted versus the distance resulting in regions where the slope errors were slightly nonzero. The models can be improved through further analysis of more variables which may have an influence on the growth of hypericum, such as the number of conspecific neighbors and the growth of the nearest rosemary, and the possible inclusion of these parameters into the model.
APPENDIX A: ESTIMATING ERRORS IN LINEAR FITS
Let us now describe how the expected error in the slope and intercept of a linear fit can be calculated. Let \((x_i, y_i)\) be the actual data values. Let \(y(x_i)\) be the expected predicted value by the linear line:

\[ y(x_i) = mx_i + r \]

Let \(N\) be the number of data points. Then the square of the average deviation, \(\varepsilon\), is given by:

\[ \varepsilon^2 = \frac{1}{N} \sum_{i} (y_i - mx_i - r)^2 \]

In the method of least squares, one wants to adjust the coefficients \(m\) and \(r\) so that the average deviation is a minimal. In order for this to be so, it must be required that:

\[ \partial_m \left( \frac{1}{N} \sum_i (y_i - mx_i - r)^2 \right) = \partial_r \varepsilon^2 = 0 \]

and

\[ \partial_r \left( \frac{1}{N} \sum_i (y_i - mx_i - r)^2 \right) = \partial_m \varepsilon^2 = 0 \]

32
which result in (A1) and (A2)

\[ \frac{1}{N} \sum_i (x_i y_i - mx_i^2 - rx_i) = 0 \quad (A1) \]

\[ \frac{1}{N} \sum_i (y_i - mx_i - r) = 0 \quad (A2) \]

The following notation is used for averages of both \( x \) and \( y \).

\[ \bar{x} = \frac{1}{N} \sum_i x_i \quad , \quad \bar{x}^2 = \frac{1}{N} \sum_i x_i^2 \quad , \quad \bar{xy} = \frac{1}{N} \sum_i x_i y_i \quad , \text{etc.} \]

Solving (A2) and (A1) for \( r \) and \( m \) yields.

\[ m = \frac{\bar{xy} - \bar{x} \bar{y}}{\bar{x}^2 - \bar{x}^2} \quad (A3) \quad , \quad r = \frac{\bar{x^2 y} - \bar{x} \bar{xy}}{\bar{x}^2 - \bar{x}^2} \quad (A4) \]

Let us now consider how to estimate the probable uncertainty in \( m \). This is done by considering the following. Let all the values of \( x_i \) be called the “data points”. Let each data point have a distribution of possible \( y \)-values. At each data point assume that the same standard deviation of these \( y \)-values is assumed to be independent of \( x \). Furthermore assume that there is no correlation between the variations in \( y_i \) and \( y_j \) at different data points. For terminology, let us note that the collection of all possible \( y \)-values at all data points will be referred to as an “ensemble”. A data set is defined to be one set of \( y \)-values at all data points which can be drawn
from the ensemble. Thus the ensemble is a union of all possible data sets. Now it is clear that for each data set drawn from the ensemble, one would generally calculate a different value for $m$. The question is: “what is the standard deviation of the collection of $m$’s that can be calculated from the data sets drawn from this ensemble?” To find this, one needs to calculate the deviation in $m$, $\delta m$, that will occur on the average when one considers all possible deviations due to the variations in the individual $y$-values of a data set drawn from the ensemble. For a single data set, this deviation is:

$$\delta m = \sum_i \left( \frac{\partial m}{\partial y_i} \right) \langle \delta y_i \rangle$$

where from (A3) it follows that

$$\frac{\partial m}{\partial y_j} = \frac{x_i - \bar{x}}{N(x^2 - \bar{x}^2)} \quad (A5)$$

In the above, it has also been assumed that the deviations in the $y$-values are sufficiently small so that a linear approximation of a Taylor series is valid.

Now consider an ensemble average over all data sets in the square of $\delta m$. (The ensemble average of $\delta m$ vanishes since the average of the deviations of the $y$-values vanish by assumption.) Thus we have

$$\left\langle (\delta m)^2 \right\rangle = \sum_{i,j} \left( \frac{\partial m}{\partial y_j} \right) \left( \frac{\partial m}{\partial y_j} \right) \langle \delta y_i, \delta y_j \rangle$$
The ensemble average of $\delta y_i \delta y_k$ is zero unless $k = i$, since the signs of the $\delta y$’s can arbitrarily be positive and/or negative. However, when $k = i$ the sign is always positive and only these terms will contribute to the above sum. Furthermore, as stated earlier, the ensemble average of the square of the $\delta y_i$’s was assumed to be independent of $x_i$. This value will be designated by

$$\langle \delta y_i \rangle^2 = \varepsilon_y^2 \quad \text{(A6)}$$

Therefore $\varepsilon_y$ can be taken outside the summation, giving

$$\langle (\delta m)^2 \rangle = \varepsilon_y^2 \sum_i \left( \frac{\partial m}{\partial y_i} \right)^2$$

From Equation (A5) it is easy to perform the final sum, and upon taking a square root, one obtains

$$\sqrt{\langle (\delta m)^2 \rangle} = \frac{\varepsilon_y}{\sqrt{N(x^2 - \bar{x}^2)}}$$

Similarly, one can show that

$$\sqrt{\langle (\delta x)^2 \rangle} = \varepsilon \sqrt{x^2} \sqrt{N} \sqrt{\frac{x^2 - \bar{x}^2}{N(x^2 - \bar{x}^2)}}$$
APPENDIX B: MODEL ERROR
The model error from Chapter 5 is calculated using the following equation.

\[ e^2 = \frac{1}{M} \left( \sum_{i} \frac{1}{N_{RM1}} (y_i - H_{i, RM1})^2 + \ldots + \sum_{i} \frac{1}{N_{RM5}} (y_i - H_{i, RM5})^2 \right) \]

, where \( M \) is the total number of plants (652), \( N_{RM1} \) is the number of plants in the RM1 category, etc., and \((y_i - H_i)_{RM1}\) is the actual final height minus the calculated final height in the RM1 category, etc.
LIST OF REFERENCES


