Spring 5-3-2018

Creating a Predictive Model for Flowering of Virginia Orchid, Cypripedium pubescens

Emily Horton
horton_e@lynchburg.edu

Creating a Predictive Model for Flowering of Virginia Orchid, *Cypripedium Pubescens*

Emily Horton

Senior Honors Project

Submitted in partial fulfillment of the graduation requirements

of the Westover Honors Program

April, 2018

Dr. Doug Thomasey

Dr. Samrat Thapa

Dr. Nancy Cowden

Abstract

This research investigates the native Virginia orchid, *Cyprepedium pubescens*, or the large yellow lady-slipper. Researchers at Lynchburg College, located in central Virginia, collected orchid activity data over the span of nine years from 2006-2014. This data included the following information: when the first sprout appeared, the number of leaves, the number of flowers, and the number of flowers per each plant. Using collected data about nine orchids on the campus of Lynchburg College as a basis, we wanted create a model that would predict when we might see flowering on a yearly basis. Flowering is important because researchers are investigating the pollination of these flowers, and pollination happens when the flowers are in full bloom.

Using a multiple linear regression, we created a model which used the number of flowers as the dependent variable, and the independent variables of: temperature, yearly rainfall, average humidity, and the number of days since the last freeze. It was found that the most significant factors in this model are the number of days since the last freeze, rainfall, and temperature. The model provided insight towards the insignificant role that humidity played in predictive measures.

Introduction

Large yellow lady-slippers, or *Cypripedium pubescens,* are an annually-blooming plant native to central Virginia. These flowers are perennials, meaning that they may flower each year of their lifespan, once they are mature enough to flower. Given that orchid species typically reach flowering maturity seven years after establishment from a seed, *Cypripedium pubescens* may flower for at least 13 years and produce multiple millions of seeds over that span (Case & Bradford, 2009). Plants have been documented to live for up to 20 years after reaching adulthood. However, individual seeds have a very low likelihood of producing viable plants, contributing to the rarity of this orchid species. For decades, scientists have struggled to determine how these plants attract pollinators (Walsh, Arnold & Michaels, 2014).

Many plant ecology models study the population dynamics of the plants. However, there are several different obstacles in studying plant ecology, and we had to make a few decisions about how we would approach the data available to us. The first obstacle is that plants are sessile and tend to interact primarily with nearby individuals (Pacala & Silander, 1985). The second challenge is that plants exhibit plastic growth such that the sizes or reproductive abilities of adults may vary by several orders of magnitude (Silander & Pacala, 1985). Since the relative sizes of plants and the size of the area where the plants grow were not measured in the original research, we were not able to create a model in which we focused on population dynamics. Due to the variables available to us, the model we created instead focused on individual plants themselves.

Researchers at Lynchburg College have worked for years to collect data about these plants, to try to answer the question of why pollinators are attracted to these plants. Dr. Nancy Cowen and Dr. Priscilla Gannicott of the Lynchburg College Biology and Chemistry departments, respectively, have worked for the last twelve years to answer questions about these elusive orchids. Of utmost importance to them is elucidating the chemical scent profile of these plants.

The original researchers generated quite a bit of data, which opens the possibility to explore a number of different research options. The focus of this paper is to construct an appropriate model which will predict, based on several variables, when the organisms in the study will move from one life stage to another. Life stages have been determined to be the following: first sprout, appearance of leaves, and full flowering. For the original researchers' purposes, information was also collected about the fruits of the flowers. These are the seed-bearing structures of the plants, and their appearance signifies that the flower is in its last life stage, and that it has been successful. "Success" indicates that the flower will be able to pass its genes along to the next generation, through its seeds. The fruit data did not prove to be as useful as first thought since many of the flowers were missing fruit data. What *will* be assessed in this model is whether several independent variables influence when the first *flower* of the plant can be seen.

Aside from the plant data itself, we did not have any other measured variables available to us. We decided to focus on what seemed to obviously affect plant growth. Therefore, the variables which we focused on in this study were yearly rainfall, daily temperature, humidity, and the number of days since the last freeze. These factors were chosen because of their connection to the growth of the flowers. It seems easy to imagine that any of these factors would cause a change in the flowering of the plants, but we also hoped that these variables would work in conjunction to cause a different overall change. Therefore, we tested whether each variable influences the flowering of the plant either on its own or in combination with the other variables.

The model that was used to analyze this data is a multiple regression model, which examines the possible relationship that multiple predictor variables have on a dependent variable. According to relevant research, a multiple-regression model can be particularly useful with this type of data because of the great number of variables (Seymour, 2002). This type of statistical model was of use to us because of the number of independent variables we were interested in studying. Based on the data accrued, a predictive model was able to be modified or created to help the original researchers better plan their pollinator research for subsequent years.

Research questions:

1. Are we able to determine when *Cypripedium pubescens* will show a first flower on a yearly basis?
2. Among the variables being tested (yearly rainfall, humidity, daily temperature, and the number of days since the last freeze), which has the greatest effect on the timing of the flowering of the plants?

Background & Purpose

Many orchids, up to one-third of all known species, offer no reward to pollinators (Case & Bradford, 2009). These orchids are referred to as non-rewarding species, or as food-deceptive orchids (Case & Bradford, 2009). *Cypripedium pubescens* is, in fact, considered a food deceptive orchid: its pouch-shaped structure acts as a trap to temporarily imprison pollinators (Case & Bradford, 2009). Interestingly, it has also been shown that the deceptive pollination strategy employed by many orchids may result in high levels of pollen limitation (Walsh, 2014). Pollen limitation occurs when a plant is unable to be pollinated, and this is dangerous because pollen is required for reproduction. In short, this means that researchers are still unable to say with certainty what attracts pollinators to this plant. With a predictive model, researchers will be better able to plan their time in the field by pinpointing the specific time periods when orchids will be flowering. This will assist in the efforts to answer this important question.

Existing evidence by field botanists shows that an annual fluctuation in flower production is particularly noticeable in the native orchids of the temperate zones and is especially apparent in members of the genus *Cypripedium* (Curtis, 1954). Much of the data that has been collected by researchers at Lynchburg College has yielded evidence to support this claim. Interestingly, even though these researchers were looking to collect information about the pollination of the flowers, they gathered additional data which can be used to predict when the flowers will bloom from year to year. However, as has been shown by Curtis (1954) and others, these plants do not always flower predictably. This is important to the original researchers because, if they were able to predict when the flowers would bloom in normal conditions, they might be able to collect data more efficiently.

In many terrestrial plants, flowering is an important measure of the maturity of the plant, as well as a predictor for its reproductive success. Commonly, flowering is associated with attracting pollinators to the plant. In the normal plant growth cycle, flowering plants must exchange pollen (gametophytes) between the male part (stamen) of the plant and the female part (pistil) of the plant. Cross-pollination between plants is also important to create a greater amount of diversity in the future generations of the plant. In our model, flowering was used as determinants of the flower’s potential for reproductive success. This is tangentially important to our model because the original researchers were interested in the reproductive success in the population—this is why they are trying to determine why pollinators are attracted to the plants.

Cypripedium pubescens are rare plants, therefore it is important that conservation efforts focus on the reproductive success of these plants. As shown by previous researchers, multiple factors limit reproduction (Walsh, Arnold & Michaels, 2014). These factors include rainfall, frost, and temperature, which all may indeed cause an effect on these organisms. Lapenis et al. (2013) offers additional support and discusses how plant sensitivity to warming is a real concern. Furthermore, the authors discussed how plant sensitivity to warming is greater for plants flowering in the spring than plants flowering in fall or summer. The same study suggests that sensitivity to warming may interact with other factors such as changes in precipitation and increased temperature variability. For this reason, we decided to investigate the impact that the last frost has on the plants. Using the average last date of the last freeze as a baseline, it is a well-supported hypothesis that the weather will continue to get warmer, yet the variability in the temperature may continue (Lapenis et al., 2013).

Figure 1: Portrayal of the pouch-like structure of the *Cypripedium pubescens.*

Other factors such as plant height, number of flowers, number of stems, pollen limitation, and seed predation can also help determine reproductive success of these organisms (Walsh, 2014). This is likely since pollinators may only be able to reach flowers that are a certain height or shape or size. Flowering stem height can be an important factor: pollinators are better able to see these plants (Walsh, 2014). While these factors could be important, we have decided to focus on other variables, such as number of leaves, which better align to our research question. The number of leaves may or may not correlate with flower visibility or recruitment of pollinators, but it can represent increased photosynthetic resources available for subsequent fruit maturation (Walsh, 2014). This is important because the plant is entirely unable to recruit pollinators if it is unable to make energy for itself through photosynthesis. In addition to this data, we also utilized data concerning the day when a sprout first appeared, the number of stems, the number of flowers per stem (and the day that they first appeared), and the number of fruits (or reproductive structures) per stem. The date of the last freeze will serve as a baseline for which this data will be based off. We decided to study the effect of temperature because, as can be seen in nature, flowers generally bloom in spring and summer, and warmer—if you are in the Northern hemisphere—months. As discussed previously, the number of days since the last freeze is important to consider because flowers are more likely to bloom after the last freeze. Yearly rainfall is important to our model because plants require a certain amount of water to live.

On a larger scale, it is important to understand how and when these plants are pollinated to better understand the entire ecosystem in which they live. There may come a day in which the original researchers or others may wish to investigate the entire population and the impact of surrounding plants and animals on the population. Because this is an endangered species, we could expect that with its loss could come a greater impact on the surrounding ecosystem. A model such as this would be able to predict flowering, but with some modifications, it may be able to tell more about the ecosystems in which the plants live.

Methods

Between the years of 2006-2014, the original researchers collected the following data related to the flowers: first sprout, first leaves (and count of leaves), first bud, and first date of flowering. An example of this data can be found in Appendix A. The researchers collected this data each time the population was visited, which averaged out to be three or four days each month each year.

For each plant, we had to transfer the field notes into an electronic form: for our purposes, Google Sheets. To supplement the size of the dataset and increase the validity of our results, ten extra data points were added to each year's dataset. This data allowed us to incorporate archived data starting from the average date of the last freeze for this region. Based on evidence from previous research, this date was calculated to be March 1st (Case & Bradford, 2009; Suckling, 1988). Similarly, we found that the first flowering of the plants most likely occurred 30 days after the last freeze (Hunt, 1982; Trevezas, Malefaki & Cournede, 2014). It is known that there were no flowers yet present, however, these added data points help us better understand the effects of spring temperature, humidity levels, and rainfall leading up to the first bloom. Since much of the data collection did not begin until late March or early April, this did not interfere with, or replace, the original collected data. We based these dates on the actual first freeze dates for ten years, starting in 2006. For the ten years, we averaged the day of the first freeze to find that the average first freeze date for our region is March 1st. The added data points looked like this: 0 days since the last freeze, 6 days since the last freeze, 12 days since the last freeze, and so on (these were all based on the average first freeze date). The only numbers associated with the added data was the date and the number of days since the last freeze. The stems, leaves, flowers, and fruits data all remained zero because there would be no estimated growth of plants that close to the last freeze. The added data included stem, leaves, flowers, and fruit data, which were all assumed to be 0, because of the added data with the zero-points for stems, leaves, and flowers. Ten data points were added before the first data point for 2008, which was 4/24/2008. Each of these points were added regarding how many days since the last freeze, in intervals of 6 days.

We were able to obtain historical weather records online. Similarly, we used the same website to find the average daily humidity, and the precipitation to date for each day which was sampled. All this data was compiled and stored in the Google sheet. We were able to do this because we saw from the data that plants did not begin to show stems until late March or early April. Overall, this helped improve the results of the statistical tests: multiple regressions performed on the data set with zero-data added yielded results much closer to the estimated outputs. This will be further discussed in the results section below.

For our research, we used a step function to determine when each flower would bloom. Instead of determining that a value of one would represent a *whole flower*, we decided that a bloom would be any value greater than 0. This is important to note as we further discuss the results.

Sample

For the original research, data was collected about *Cypripedium pubescens* currently growing at Lynchburg College. The Lynchburg College researchers began collecting in data in 2004. The largest amount of data in our study comes from these plants which were first seen in

2004. The researchers subsequently discovered other plants, each labeled with the year in which they were discovered (Appendix B explains the appropriate numbering of flowers). We are not able to pinpoint the year in which each flower began growing, but instead use the year in which it was discovered as the starting point. Plant 012 was discovered in 2007, plant 013 was discovered in 2008, and plant 014 was discovered in 2009.

Although there was data recorded for 11 plants, we used plants 001-009 because of the completeness of the data for these plants. In general, the data collected from these plants included stems, leaves, and flowers. As stated before, there was little to no data collected on the fruits produced by the plants. Again, fruits refer to the reproductive structures of the plants. We assume in this study that fruits arise after the flowering of the plants, and that those are the structures that are visited by pollinators. For plants 010-013, there was about half of the data collected as compared to the other plants because these plants were discovered in later years (2007 and beyond). These plants in particular were exempted from our statistical tests because of the relatively small amount of data they produced.

The population is located on the campus of Lynchburg College in Lynchburg, Virginia. The plants grew in a temperate zone during each year of data collected, near a man-made lake (College Lake). Orchids such as *Cypripedium pubescens* grow best under dappled sunlight, which is what they experienced while growing under oak, maple, and poplar trees native to the region. The traditional growing season of the plant is March until May-- therefore data was collected during these months.

The data was analyzed using SPSS. To create a rudimentary model, each variable was investigated to determine which variable(s) had the largest effect on the number of flowers that would grow during a certain year. As we were trying to create a predictive model for flowers, (F) was the dependent variable, and the number of days since the last freeze (DSF), temperature (T), humidity (H), and yearly rainfall up to that date (R) were treated as the independent variables, in differing arrays. In SPSS, we ran 15 different multiple regression analyses (as laid out in Analysis and Results), each with an emphasis on a different variable or group of variables. We used the data in the Google sheet to do this.

Results

To create a rudimentary model, each variable was investigated to determine which variable(s) had the largest effect on the number of flowers that would grow during a certain year. As we were trying to create a predictive model for flowers, (F) was the response variable, and the number of days since the last freeze (DSF), temperature (T), humidity (H), and yearly rainfall up to that date (R) were treated as the predictor variables, in differing arrays. In SPSS, we investigated 15 different multiple regression analyses, each with an emphasis on a different variable or group of variables.

In order to determine the best fit model, it was decided that each variable should be analyzed by itself and in relation to all other possible combinations with the other predictor variables. For example, one combination was to investigate each of the independent variables separately, which resulted in four possible models. Another combination was to pair two predictor variables together, which resulted in six combinations. The same was conducted for groups of three predictor variables and then one model for all four predictor variables. Below is a listing of all fifteen different possible combinations:

- Flowers by DSF
- Flowers by T
- Flowers by H
- Flowers by R
- Flowers by H, R
- Flowers by R, T
- Flowers by DSF, H, T
- Flowers by DSF, H, R

- Flowers by DSF, H
- Flowers by DSF, T
- Flowers by DSF, R
- Flowers by H, T
- Flowers by DSF, R, T
- Flowers by R, T, H
- Flowers by DSF, R, T, H

For each different combination, a linear regression model was created in SPSS. The output from SPSS included coefficients for each of the different predictor variables. An example of the SPSS output, where a simple predictor, temperature, was used, is listed in Appendix C. We focused on these coefficients in each of the fifteen models we created, in an attempt to find a best fit model. A summary of outputs for each coefficient associated with all the different combinations can be found in Appendix D. It was important for us to analyze each of these different combinations in order to determine which model(s) and variables would be considered the most significant. Specifically, we wanted to include only the most significant predictors from the general model which takes on the following form:

$$\mathbf{F} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{DSF}) + \boldsymbol{\beta_2}(\boldsymbol{T}) + \boldsymbol{\beta_3}(\boldsymbol{R}) + \boldsymbol{\beta_4}(\boldsymbol{H})$$

In this model, the flowers (F) are reliant on all of the different factors; β_0 represents the overall constant, or how many flowers we would expect to see if all other predictors were zero. In other words, β_0 represents the initial number of flowers assuming the value of all other factors were zero. As expected, in most models this value was negative, indicating that no flowers would yet be present. The coefficients associated with the predictor variables influence the number of flowers with varying degrees in order to predict the number of flowers. The constants β_1, β_2, β_3, β_4, represent the coefficients for each individual variable. In each model, the predictor variables have an effect on the number of flowers, so these values represent the amount of the effect on the flowers, with regard to their specific variable (DSF, T, R, or H). These constants were calculated by using SPSS.

For a two-variable model, model (j) was found to be the best-fit model with all p<0.005 for all constants. Model (j) can be expressed as:

$$\mathbf{F} = \mathbf{-0.186 + 0.004(\boldsymbol{T}) + 0.014(\boldsymbol{R})}$$

For a three-variable model, the number of days since the last freeze (p<0.000), rainfall (p=0.018) and temperature (p<0.000) have the greatest effect on the growth of the flowers. Model (m) can be expressed as:

$$\mathbf{F} = \mathbf{-0.190 + 0.008(\boldsymbol{DSF}) + 0.003(\boldsymbol{T}) - 0.035(\boldsymbol{R})}$$

The reason that models (f) and (m) are designated with two asterisks (**) in the Appendix is because we found *all* the p-values associated with the constants to be significant. As seen in model (f), which considers the days since last freeze and the temperature, both p-values for the variables are below 0.05. This is important because it shows that both variables are significant to the model—or that they are influencing the growth of these flowers significantly. Similarly, with model (m), both p-values are below 0.05.

Discussion

For both successful models outlined in the previous sections, we were expecting β_0 to be a negative number because we were not expecting to see a flower bloom on the very first day since last freeze. This was an important aspect of all the models. In fact, we did not see a single model which did not have a negative $\beta_{0,}$ which is interesting to note. We decided to use a step function to determine the accuracy of our model. Instead of trying to solve the model for where the flower value is equal to one, it makes sense for any nonnegative, nonzero value less than or equal to one to be equal to one in this model. To this end, any sign of a flower would be counted as a whole flower. This way, when the model returns any value greater than zero, we assume that one flower is present. Any value greater than one will be assumed to be two flowers, and so on. This allows us to fine-tune our model to see the exact conditions in which one or more flowers would be present.

Confirming Evidence

Referring to our first research question, we wanted to know if it was possible to predict when *Cypripedium pubescens* will show a first flower on a yearly basis. We began by looking at model (j), to determine when F>0. Using computational software to create an inequality plot, we saw that F>0 for R >= 20, and for T>=70. Looking at model (m), we saw that F>0 for DSF>=40, T>=60, and R>=8. Comparing these results with actual data is important to our research because we need to be able to demonstrate how well the model fits actual data. To do so, we found an average of the important factors when the original researchers observed one flower present. These are: DSF=50, T=63.125, and R=5.12. Interestingly, using the above values, model (m) is much closer to the actual values than model (j).

To answer our second research question, we wanted to know which variable(s) have the greatest effect on the timing of the flowering of the plants. As we can see from models (j) and (m), it would appear that DSF, R, and T have the greatest effect overall. Using model (m), we wanted to find out when a flower would be expected to bloom and compare that to the date that it actually bloomed. We investigated three different flowers: LC 001, LC 003, and LC 005 for two years each. Not every flower showed blooming twice over the research period, so we did our best to represent a wide variety of cases in the following analysis.

Case (1) is flower LC 001 in 2008. The actual-data values for when this plant flowered are: DSF=56, T=51, and R=5.01. Plugging these into the model, we get F=0.296. In this case, the model does an accurate job of predicting when the flower will bloom. Case (2) is flower LC 001 in 2010. The flower did not bloom this year, but we can predict when it should have using the model. Using the same values for 2008, we can see that F=0.296 at 56 days after the last freeze.

Case (3) is LC 003 in 2011. The actual-data values for when this plant flowered are: DSF=62, T=68, and R=7.49. Plugging these into the model, we get F=0.248. In this case, the model does an accurate job of predicting when the flower will bloom. Case (4) is flower LC 003 in 2016. The actual-data values for when this plant flowered are: DSF=43, T=63, and R=4.05. Plugging these into the model, we get F=0.201. However, the model shows that there were actually *two flowers* present at this time, so this model may be off in this regard.

Case (5) is flower LC 005 in 2008. The actual-data values for when this plant flowered are: DSF=56, T=51, and R=5.01. Plugging these into the model, we get F=0.244. In this case, the model does an accurate job of predicting when the flower will bloom. Case (6) is LC 005 in 2011. The actual-data values for when this plant flowered are: DSF=56, T=51, and R=5.01. Plugging these into the model, we get F=0.542. In this case, the model does an accurate job of predicting when the flower will bloom.

Flowers Which Did Not Bloom

As stated above, there were several flowers which did not bloom during several years, and one flower which did not bloom at all. LC 004 did not bloom at all—one hypothesis for this is that the flower did not receive enough sunlight, or it was possible that other environmental factors we did not measure were affecting this flower.

What is interesting is that some flowers bloomed one year and did not bloom the next. This was the case for LC 001 and LC 007 in the years 2008 (bloomed) and 2009 (did not bloom). In 2008, the average temperature was 53.94 during all days in which data was collected, while in 2009 it was 4 degrees cooler on average, at 49.60 degrees. In 2008, the total rainfall during all data-collecting days was 8 inches, while the total rainfall in 2009 was 5.78. The combined effects of a lower average temperature and a lower cumulative rainfall could have cause these flowers not to bloom during the particular year.

Different Trials

Using model (m), we wanted to know what would happen to the flowering of the plants if different weather conditions occurred. Using the same average values, we tried to determine when flowering would occur during a rainy year, a really hot year, an early last freeze, and a late last freeze in which the weather gets warm very quickly.

Case (a) is a year in which there is a lot of rain. The average rain value is 5.12, so we decided to double this value to be 10.24 inches. In the model, this would account for a value of -0.35875. The other values would have to compensate for this negative value, in addition to the negative constant. In a year where there is an average of 10.24 inches, we would still expect to see the plants flowering at 50 days after the last freeze, and an average of 63 degrees. In a year where there are 20 inches of rain, the flowers would not be expected to flower until about 87 days after the last freeze.

Case (b) is a year in which the average temperature is much higher than usual. The average temperature value for flowering is 63 degrees. During a year where the temperature might be 85 degrees, when would we expect the flowers to bloom? The temperature value becomes 0.255 when the average temperature is 85 degrees, meaning that the DSF and R constants could be zero, and the flowers would still be expected to bloom.

Case (c) is a year in which there is a late last freeze, but the weather gets warm very quickly. Suppose that the DSF value for this year would be equal 60 days (greater than the observed average of 50 days), and let the T value equal 70, which is greater than the observed average during the time of flowering. This would make the DSF value equal to 0.48, and the T value equal to 0.21. These together overcome the overall constant of -0.190, meaning that rainfall would not need to compensate for that negative value, however, with the 5.12 observed rainfall average, we would see a flower value of 0.3208.

Humidity

As we saw from both most successful models, humidity had no significant impact on the growth of the flowers. This is an interesting discovery because one could assume that rainfall would contribute to humidity, making the H factor an important one to study. One study found that orchids grow best at controlled humidity, but of course controlling the humidity is not an option for us (Liao, 2017). In fact, this same study found that the optimal humidity for orchid growth is between 60-80% relative humidity (Liao 2017). Interestingly, we found that the average overall humidity for all data-collection days was 60%. Overall, humidity may not have a large effect on the flowers' growth in this model because the humidity was in the optimal range.

Limitations of These Studies

As has been shown in other studies, plants of the genus *Cypripedium* vary greatly in their reproductive capacity each year (Curtis, 1954). This may be a reason for the great variance in the number of flowers that were seen from year to year. In the original dataset, some plants did show flowering each year, and some did not: plant 002 in particular. Because of this, flowering may not have been the best variable to examine—however, it was the variable that was most accessible to us.

In addition, it is difficult to predict the growth of a flower based on a few factors; any type of growth is much more complex than we assume in this paper. If time would permit, it would be interesting to examine the soil temperature, photosynthetically active sunlight, and other factors.

Data limitations

Ideally, we would be able to do this study again, with more data from the field. It would be the best-case scenario for many more plants with many more flowers to be considered—perhaps a larger, older population which has flowering each year. As discussed previously, these plants are perennials, and can live up to 20 years past maturation, and according to research, these plants can take between 10 and 16 years to mature (Curtis, 1954). It is important to note that, by these standards, the Lynchburg College orchids are still quite young, with the oldest of the sample being first seen in 2004.

Because of all of the zero-data entered for each plant, the data could be skewed to yield different results than if all of the data had been collected exactly from the field. As with any research, human error should be considered with the research presented—both with the data collection, and analysis.

Conclusions and Future Study

We created a model for these researchers by using a multiple regression model, which used the number of flowers as the dependent variable, and the independent variables of: temperature, yearly rainfall, average humidity, and the number of days since the last freeze. The most influential variables were found to be the number of days since the last freeze, rainfall, and temperature. Humidity had little to no effect on when flowers could be expected to bloom each year.

In general, we were able to find two models which, upon testing, matched the data quite accurately. Based on a limited amount of data which was not collected for the intent of creating a model, we have been able to construct a model with moderate success. It is our hope that, by doing more work with the models we have, we may come up with a more complete understanding of the life stages of these plants.

After performing this research, it would be interesting to continue with looking at life-stage data. Since many of the plants have been living since 2004, it would be fascinating to keep checking on them as they reach maturation well into the year 2020. Furthermore, this amount of data gathered from this amount of time would give us a better idea of the accuracy of our model.

The original researchers are also involved in studying the same plant at a larger facility further from the Lynchburg College campus. This larger population may give more information as to the growth cycles of the plant. In future study, I would be interested in testing our model against the larger population, or against a wild population of *Cypripedium pubescens.* We are hopeful that this model may, after more testing in a larger population, be able to predict flowering of different species of the genus *Cypripedium*. In addition, this model may be able to predict flowering or growth for species in central Virginia, or with similar climates to the area of study.

Acknowledgments

I would like to thank Dr. Doug Thomasey for his guidance and assistance with developing this research. I would also like to extend gratitude to Dr. Nancy Cowden for allowing me to have access to her datasets, and for teaching me everything I know about these beautiful plants. In addition, I would like to thank Dr. Samrat Thapa for his help during my writing process. In closing, I would like to extend my sincere gratitude to the Westover Honors Program at Lynchburg College for supporting this project.

Literature

Case, M., and Bradford, Z. (2009). Enhancing the trap of lady's slippers: a new technique for discovering pollinators yields new data from *Cypripedium parviflourm. Biological Journal of the Linnean Society, 160.* 1-10.

Curtis, J. (1954). Annual fluctuation in rate of flower production by native cyprediums during two decades. *Bulletin of the Torrey Botanical Club, 81(4)*, 340-352.

Hunt, R. (1982). *Plant growth curves: the functional approach to plant growth analysis*. Edward Arnold Ltd., 248.

Lapenis, A., Henry, H., Vuille, M., & Mower, J. (2014). Climatic factors controlling plant sensitivity to warming. *Climatic Change*, 122(4), 723-734.

Liao, Min-Sheng. "On Precisely Relating the Growth of Phalaenopsis Leaves to Greenhouse Environmental Factors by Using an IoT-Based Monitoring System." *Computers and Electronics in Agriculture*, Elsevier, 17 Mar. 2017,

Pacala, S. W., & Silander Jr, J. A. (1985). Neighborhood models of plant population dynamics. I. Single-species models of annuals. *The American Naturalist*, *125*(3), 385-411.

Seymour, Lynne A. (2002). Modeling plant growth using regression with periodically correlated errors. *Journal Of Agricultural, Biological, And Environmental Statistics*, (2), 191.

Suckling, P. W. (1988). Freeze dates and growing season length for the southern piedmont, 1911-1986. *Southeastern Geographer, 28(1),* 34-45.

Trevezas, S., Malefaki, S., & Cournède, P. (2014). Parameter estimation via stochastic variants of the ECM algorithm with applications to plant growth modeling. *Computational Statistics and Data Analysis*, 7882-99.

Walsh, R., Arnold, P., Michaels, H. (2014). Effects of pollination limitation and seed predation on female reproductive success of a deceptive orchid. *Annals of Botany*.

Appendix A

(example data sheet)

Identity	Date	DSF	Stems	Leaves	Flowers	Fruits	T	H	R
LC 04 001	3/1/2008	0	0	0	0	0	32	49	3.22
LC 04 001	3/7/2008	6	0	0	0	0	43	85	4.71
LC 04 001	3/13/2008	12	0	0	0	0	53	45	5.4
LC 04 001	3/19/2008	18	0	0	0	0	59	80	6.02
LC 04 001	3/25/2008	24	0	0	0	0	39	64	6.11
LC 04 001	3/31/2008	30	0	0	0	0	43	97	6.83
LC 04 001	4/6/2008	36	0	0	0	0	47	93	8.18
LC 04 001	4/12/2008	42	0	0	0	0	65	47	8.26
LC 04 001	4/18/2008	48	0	0	0	0	62	46	8.26
LC 04 001	4/24/2008	54	0	0	0	0	64	70	9.45
LC 04 001	4/26/2008	56	2	10	1	0	71	70	9.72
LC 04 001	4/29/2008	59	2	10	1	0	48	71	11.22
LC 04 001	4/30/2008	60	2	9	1	0	48	67	11.22
LC 04 001	5/1/2008	61	2	12	1	0	59	61	11.22
LC 04 001	5/2/2008	62	2	11	1	0	68	61	11.22
LC 04 001	5/5/2008	65	2	11	1	0	62	63	11.22
LC 04 001	3/1/2009	0	0	0	0	0	33	95	4.59
LC 04 001	3/7/2009	6	0	0	0	0	59	66	4.99
LC 04 001	3/13/2009	12	0	0	0	0	36	80	5.06
LC 04 001	3/19/2009	18	0	0	0	0	54	69	6.74
LC 04 001	3/25/2009	24	0	0	0	0	41	61	6.82
LC 04 001	3/31/2009	30	0	0	0	0	48	54	7.5
LC 04 001	4/6/2009	36	0	0	0	0	54	66	8.21
LC 04 001	4/12/2009	42	0	0	0	0	49	51	8.52
LC 04 001	4/18/2009	48	0	0	0	0	59	52	9.22
LC 04 001	4/24/2009	54	0	0	0	0	63	48	10.37

Appendix B

Table 1: Identification of Variables

LC # $	
#	Represents the year in which the flower was discovered
$	Represents the number assigned to the plant in the population (for identification purposes)
Leaves a,b,c	
a	Represents the first identified leaf
b	Represents the second identified leaf
c	Represents the third identified leaf
Stems	
*	Denotes new sprout
Flowers	
Denoted in digits (i.e., 1-10)	Refers to colorful bloom of the plant; Will not be measured by length, width, etc.
Fruits	
Denoted in digits (i.e., 1-10)	Refers to seed-bearing structures in the plant
Temperature	
All temperature data from historical records	Reported in degrees Fahrenheit

Days since last freeze	
Last Freeze data from historical records	Reported in degrees Fahrenheit
Rainfall to date	
Rainfall data from historical records	Recorded in in/year for each year represented

Appendix C

(Sample statistical findings - SPSS)

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	T[b]	.	Enter

a. Dependent Variable: Flower

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.193[a]	.037	.036	.31541

a. Predictors: (Constant), T

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.061	1	4.061	40.826	.000[b]
	Residual	105.350	1059	.099		
	Total	109.412	1060			

a. Dependent Variable: Flower

b. Predictors: (Constant), T

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.235	.049		-4.786	.000
	T	.006	.001	.193	6.390	.000

a. Dependent Variable: Flower

Appendix D: Statistical Results

	Model	Constant	P-val	DSF	p-val2	H	p-val3	T	p-val4	R	p-val5	Model fit (ANOVA)
	1 choice											
a	DSF	-0.069	0.018	0.004	0.000	-	-	-	-	-	-	
b	H	0.087	0.044	-	-	0.000	0.047	-	-	-	-	0.410
c	T	-0.217	0.000	-	-	-	-	0.005	0.000	-	-	0.000
d	R	-0.015	0.369	-	-	-	-	-	-	0.022	0.000	0.000
	2 choices											
e	DSF, H	-0.066	0.146	0.004	0.000	-3.896E-5	0.955	-	-	-	-	0.000
f	DSF, T	-0.168	0.000	0.004	0.000	-	-	0.002	0.020	-	-	0.000
g	DSF, R	-0.069	0.000	0.008	0.000	-	-	-	-	-0.033	0.000	0.000
h	H,T	-0.213	0.001	-	-	-6.11E-5	0.930	0.005	0.000	-	-	
i	H,R	-0.004	0.938	-	-	0.000	0.792	-	-	0.022	0.004	0.000
j **	R,T	-0.186	0.047	-	-	-	-	0.004	0.000	0.014	0.000	0.000
	3 choices											
k	DSF,H,T	-0.169	0.008	0.004	0.020	3.01E-5	0.965	0.002	0.008	-	-	0.000
l	DSF,H,R	-0.070	0.123	0.008	0.000	1.176E-5	0.986		-	-0.033	0.000	0.000

m **	DSF,R,T	-0.190	0.000	0.008	0.000	-	-	0.003	0.005	-0.035	0.000	0.000
n	R,T,H	-0.183	0.005	-	-	-4.84E-5	0.945	0.004	0.000	0.014	0.001	
	4 choices											
o	DSF,H,T,R	-0.196	0.002	0.008	0.000	9.999E-5	0.883	0.003	0.005	-0.035	0.000	0.000