MEMORANDUM

To: Miriam Mendez

From: Team B (Kevin Thai, Andrew Harlow, Anushka Patnaik, Jessica Eng)

Date: April 27, 2022

Re: Effect of Weather Conditions on Resistance to Macrophomina Root Rot Statistical Analysis Recommendation

---

The purpose of this memo is to describe the statistical methods and findings from an analysis of your strawberry cultivar trials. I hope that this information helps you address your research question:

> *"What are some ways that we can characterize environmental conditions that may be explanatory of disease incidence?"*

This memo is organized into five sections.

I. The first section, "**Abstract of Key Findings**," presents an overview of key results from our analysis. (page 2)

II. The second section, "**Background and Data**," includes a description of our understanding of your data and of your main statistical question, along with a description of some statistical issues that are important in selecting and then explaining a statistical analysis of your data. (page 3-4)

III. The third section, " **Statistical Methods**," describes an analysis approach for your consideration. (page 5)

IV. The fourth section, "**Results and Discussion**," describes the results from an analysis for your consideration. (pages 6-10)

V. The fifth section, "**Summary of Key Findings**," summarizes and discusses the key findings based on the statistical analyses. (page 11)

VI. The last section, "**Technical Output,**" includes various computer outputs for reference that were not included in our analysis. (page 12-14)

If you have any additional questions about this work following our consulting meeting today, feel free to contact us through Professor Smith at hsmith@calpoly.edu so that we may set up another meeting to discuss your questions.

I. **Abstract of Key Findings**

- Out of all the explanatory variables from the dataset, dew point and vapor pressure were the most correlated with mortality rates of strawberry plants.

- As the soil/air temperature threshold increases, the correlation between degree-days and mortality rate becomes more negative.

- Warmer weather conditions (higher air temperatures, soil temperatures, vapor pressure etc.) are generally associated with higher mortality rates of strawberry plants.

II. Background and Data

Our understanding is that you want assistance in answering your research question: "What are some ways that we can characterize environmental conditions that may be explanatory of disease incidence?" After cleaning and combining the data sets provided, we have listed below the variables that will be included in this analysis.

- **Mortality Rate** (quantitative): proportion of strawberry plants that died from Macrophomina Root Rot
- **Precipitation** (quantitative): total amount of precipitation during one cultivation season from November - July (mm)
- **Soil Temperature** (quantitative)**:** sum of monthly average soil temperatures during one cultivation season (°C)
- **Air Temperature** (quantitative): sum of monthly average air temperatures during one cultivation season (°C)
- **Dew Point** (quantitative): sum of temperatures at which air is saturated with water vapor (°C)
- **Vapor Pressure** (quantitative): sum of partial pressure exerted by atmospheric water vapor during one cultivation season (kPa)
- **Degree-days for Soil Temperature** (quantitative): Accumulated sum of degree-days for soil temperature using lower thresholds of 0°C, 10°C, 20°C, 25°C, 30°C (°D)
- **Degree-days for Air Temperature** (quantitative)**:** Accumulated sum of degree-days for air temperature using lower thresholds of 0°C, 10°C, 20°C, 25°C, 30°C (°D)

Data Cleaning

In regards to the data provided, we decided to utilize the monthly data set because it provided the most information on environmental conditions, and we were able to combine it with the mortality data by Year. When we combined the datasets, we calculated the sum of the values for all explanatory variables for each year. We gathered data for degree-days on air temperature and soil temperature for each cultivation season using the UC-IPM degree-days calculator (http://ipm.ucanr.edu/WEATHER/index.html) with the single sine/horizontal method for five lower thresholds at 0°C, 10°C, 20°C, 25°C, and 30°C. In the analyses that follow, we investigated the ten degree-days measurements to determine which degree-day above some lower threshold has a relationship with mortality rate.

Sampling Methods

Next, we want to document our understanding of the sampling method used to collect mortality rates on the following eight strawberry varieties: bg 6.3024, ps 9271, Big Sur, Fronteras, Monterey, pe 7.2059, Sweet Ann, and Ruby June. Twenty plants of each strawberry variety were planted at four different inoculated locations at Cal Poly Field 35b for each cultivation season between November 1st to July 31st from 2016 to 2021. Final mortality rate was recorded at the end of each cultivation season which resulted in four measurements for each year.

Scope of Inference

Finally, we want to review the ability to infer about the relationship between environmental conditions and mortality rate. Because the mortality rates used in the following analysis are for eight common strawberry varieties, it is important that you are cautious about stating your conclusions. We would recommend that you limit your inference to those eight strawberry varieties that are grown on Cal Poly fields that experience environmental conditions in San Luis Obispo, CA.

Limitations

When approaching this research question we first looked at using an ANOVA (Analysis of Variance) or a Multiple Regression model to analyze the relationship between environmental conditions and mortality rate. However, as we ran our tests, we unfortunately were not able to produce any statistically significant results with these models. The reason for this problem was that we would lose degrees of freedom whenever we introduced additional predictors which we attribute to a low number of replicates and mutlicollinearity issues as seen in Table 4. We then looked at different statistical methods such as Generalized Regression and Partial Least Square Regression, and unfortunately we encountered the same problems. This caused us to settle on an exploratory analysis to answer the research question. We also want to acknowledge that the data used to calculate the accumulated degree-days for soil temperature contained missing measurements for different days for each year as seen in Table 5. Therefore, the accumulated degree-days for soil temperature is not a true accumulation for the entire cultivation season of 274 days.

III. **Statistical Methods**

To answer the question of how weather conditions affect mortality rate, we constructed an exploratory analysis which contains many visualizations that display the trends of mortality rate and most of the explanatory variables mentioned in Background and Data over time. The goal of finding important variables that may be explanatory of mortality rate caused us to think about the correlations between the explanatory variables and mortality rate. By looking at the correlations, we were able to make the following interpretation: Because of a high correlation between variable A and mortality rate, as the value for variable A increases, we expect mortality rate to also increase. After reporting these correlations, we then shifted our attention toward finding the best degree-day lower threshold for both air temperature and soil temperature in order to minimize mortality rate. We compared the correlation of these degree-days to mortality rate. Our main goal was to find strong correlations with mortality rate as we would not recommend using the degree-day calculations with strong positive correlations. This indicates that the mortality rate is higher for the corresponding degree-day.

IV. Results and Discussion

We split this section into three subsections to cover multiple aspects of your research question. In the three subsections, we will explore the correlations between the predictor variables and mortality rate, different degree-day thresholds, and how these relationships have changed over time.

Correlations between Environmental Conditions and Mortality Rate

Variables	Correlation
Dew Point	0.4402
Vapor Pressure	0.4217
Soil Temperature	0.4145
Air Temperature	0.4061

Table 1: Table of variables (not including degree-day variables) used in the model and their respective correlations with mortality rate

These four variables, dew point, vapor pressure, soil temperature, and air temperatures, were the most correlated with mortality rate. While these correlations are weak, we would advise you to take note of them regardless. We can infer that as dew point, vapor pressure, soil temperature, and air temperature increase, mortality rate will most likely also increase. Since you mentioned during the initial meeting that you would like the mortality rate to be as low as possible, we would advise you to minimize these four environmental conditions as much as possible. Because *Macrophomina phaseolina* favors high temperatures and low soil moisture, this recommendation corresponds to limiting the growth of the fungus and decreasing mortality rate.

Degree-days for Air and Soil Temperature

We reviewed ten degree-days measurement for air and soil temperature at five lower developmental thresholds of 0°C, 10°C, 20°C, 25°C, and 35°C. As shown in Table 2 and Table 3, as the lower threshold increases, the correlation between degree-days and mortality rate becomes more negative.

Degree-days for Air Temperature	Correlation
Lower Threshold of 0°C	0.4455
Lower Threshold of 10°C	0.4325
Lower Threshold of 20°C	0.4128
Lower Threshold of 25°C	0.4196
Lower Threshold of 30°C	0.4218

Table 2. Correlation table between degree-days for air temperature at five lower thresholds and mortality rate

Degree-days for Soil Temperature	Correlation
Lower Threshold of 0°C	0.4463
Lower Threshold of 10°C	0.4451
Lower Threshold of 20°C	-0.3034
Lower Threshold of 25°C	0
Lower Threshold of 30°C	0

Table 3. Correlation table between degree-days for soil temperature at five lower thresholds and mortality rate

The accumulated degree-days for soil temperature with a lower threshold of 25°C and 30°C for all five cultivation years were 0°D, which indicates that the threshold is too high for *Macrophomina phaseolina* to begin development. Therefore, the correlation between degree-days for soil temperature with a lower threshold of 25°C and 30°C and mortality rate is 0.

Graphs of Trends Over Time for Mortality Rate, Soil Temperature, and Air Temperature

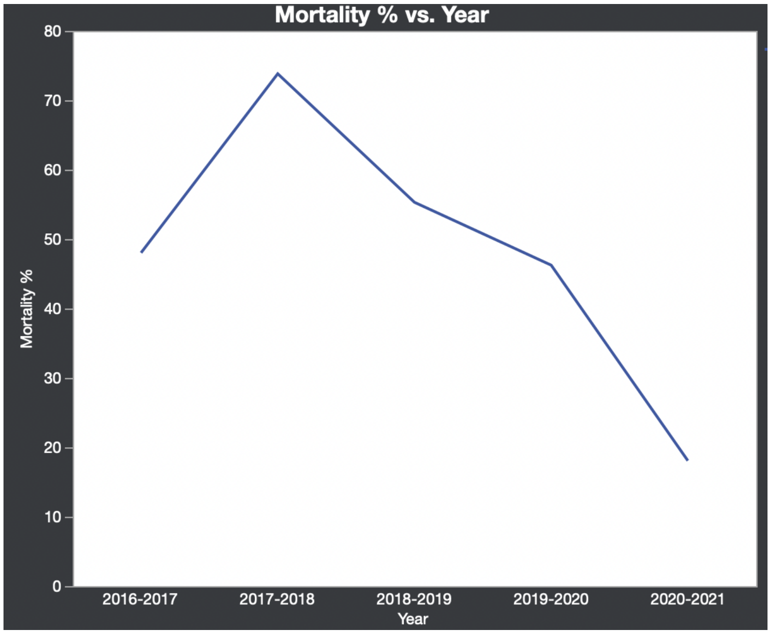

Figure 1. Line Graph of the relationship of Mortality Rate to Year

In Figure 1, we noticed that between 2017 and 2018 there was an increase in mortality rate. After 2018, however, there is a decrease through 2021. We attributed the decrease to the COVID-19 pandemic and the installation of the additional drip line in each plot and we believe the spike between 2018 and 2019 could be due to the multiple forest fires and wildfires occurring throughout California, as well as the lack of the additional drip line.

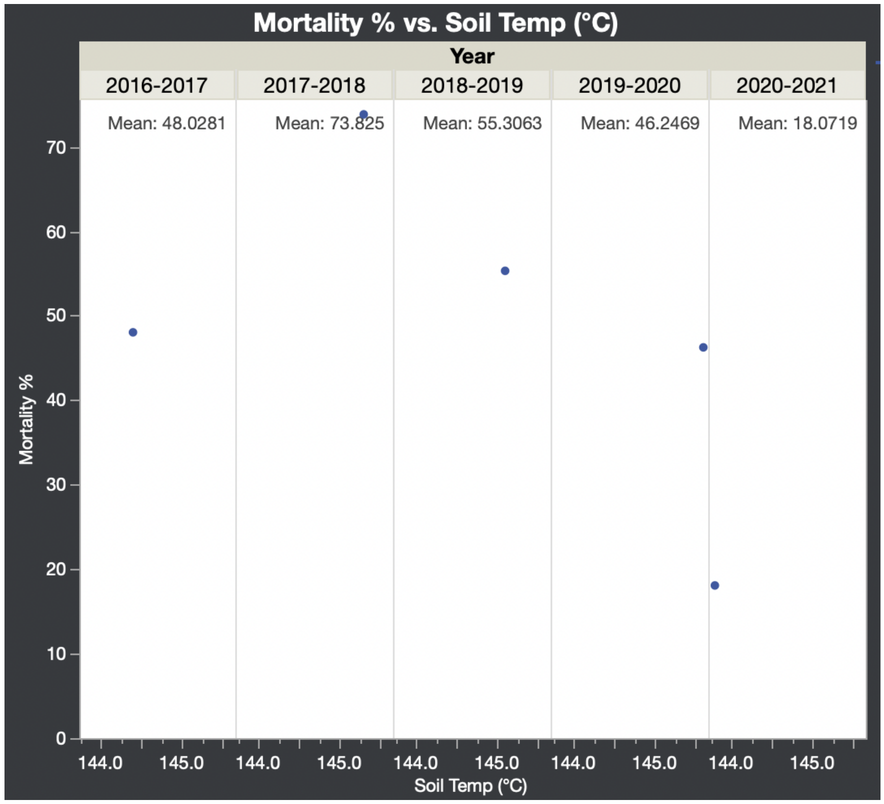

Figure 2. Line Graph of the relationship of Soil Temperature to Year

Figure 2 shows a sharp spike in mortality rate between 2017-2018 and 2018-2019, which we attribute to hot weather and the wildfires. The decrease in mortality rate between 2019-2020 and 2020-2021 may be due to the additional drip line and the COVID-19 pandemic.

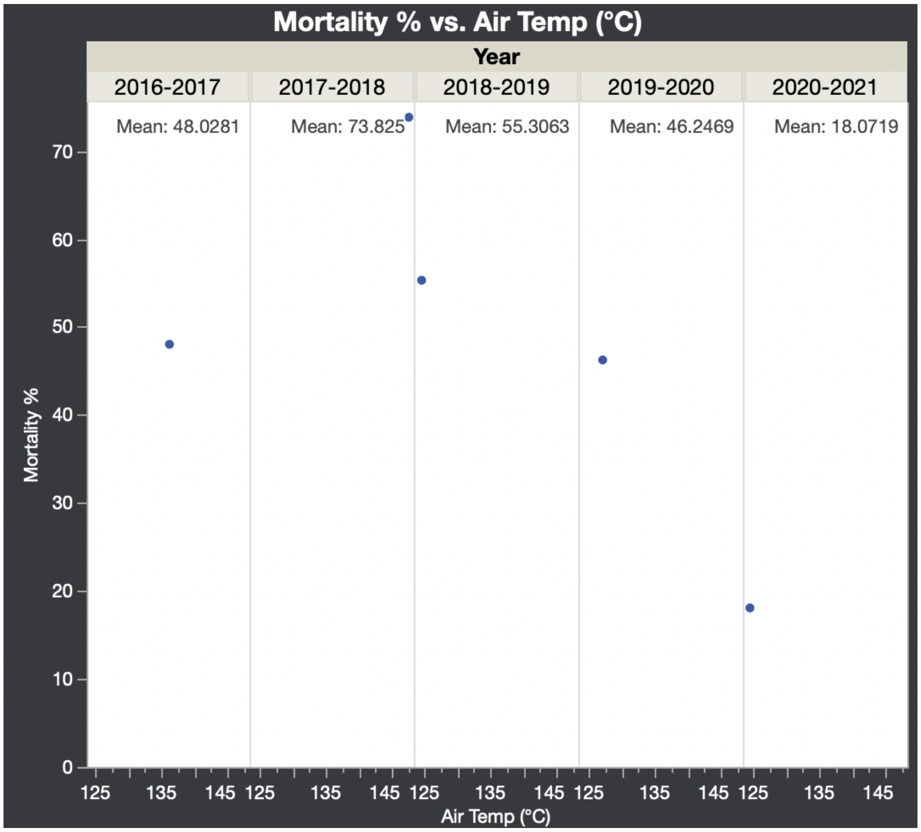

Figure 3. Line Graph of the relationship of Air Temperature to Year

Figure 3 shows that 2017-2018 had an extremely high mortality rate as well as a very high accumulated average monthly air temperature. We attributed this increase to the wildfires spread throughout California. Even though they were not in San Luis Obispo, the air throughout the state was heavily impacted. Additionally, the last five years have been the hottest in the last 140 year, according to analyses done by NASA and the National Oceanic and Atmospheric Administration (NOAA). We noticed decreases in mortality rate between 2018-2019 and 2020-2021, which is when the COVID-19 pandemic occurred. During this time, we experienced a period of lockdown, and deduced that as people remained home, tractors were not being driven, and people were not tracking soil around, so it was harder to spread *Macrophomina phaseolina* since it is a soil-borne disease.

V. Summary of Key Findings

After our analysis of the data, we found useful information that answered your research question. Even though we could not produce a well-structured model, we conducted an exploratory analysis to search for answers. First, we noticed an increasing trend in the important explanatory variables (dew point, vapor pressure, air temperature, and soil temperature) during 2018-2019. We believe that due to 2019 being one of the hottest years since modern record keeping began in 1880, the mortality rate increased. We also noticed a sharp decrease in mortality rate in 2020, which caused us to believe COVID-19 was the reason for this decline. The reason we attribute these results to the COVID-19 is because during lockdown, many people were home and not working in person. Therefore we deduced that tractors and workers were not tracking soil between fields as much as they would have in previous years. As mentioned in the initial meeting, *Macrophomina phaseolina* is a soil-borne disease, so COVID-19 could have minimized the spread. Another aspect we took into consideration was the addition of the drip line in 2020 and how that might have impacted our results. The additional drip line may contribute to lower soil temperatures and higher soil moisture, which may have caused a sharp decrease in mortality rate after it was installed during later cultivation seasons.

VI. Technical Output

Parameter Estimates

Term		Estimate	Std. Error	T Ratio	Probs > |t|
Intercept	Biased	192715.76	81894.67	2.35	0.0199*
Total Precipitation	Biased	- 21.83582	9.230618	-2.37	0.0192*
Average Soil Temperature	Biased	-11330.98	4815.583	-2.35	0.0199*
Average Air Temperature	Biased	-11119.03	4663.349	-2.38	0.0183*
Average Wind Speed	Biased	644.68948	265.886	2.42	0.0165*
Acc. Degree Days 0 (°C)	Zeroed	0	0	.	.
Acc. Degree Days 20 (°C)	Zeroed	0	0	.	.
Acc. Degree Days 25 (°C)	Zeroed	0	0	.	.

Table 4. Statistical Output from JMP of an ANOVA model for the Strawberry Disease data (0's represent lost Degrees of Freedom)

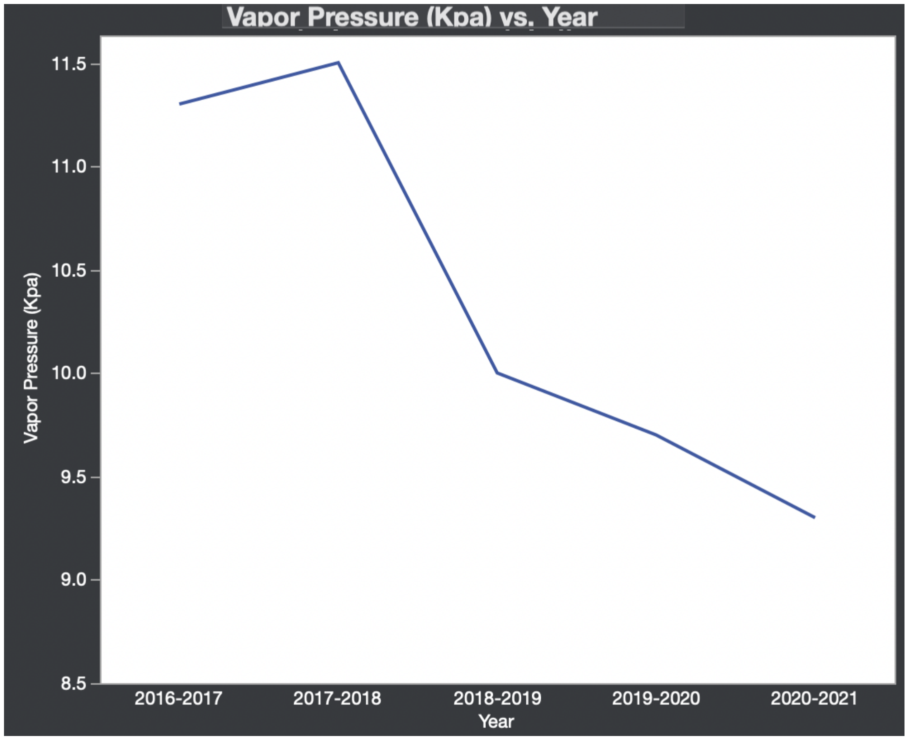

Figure 5. Line Graph of the relationship of Vapor Pressure to Year

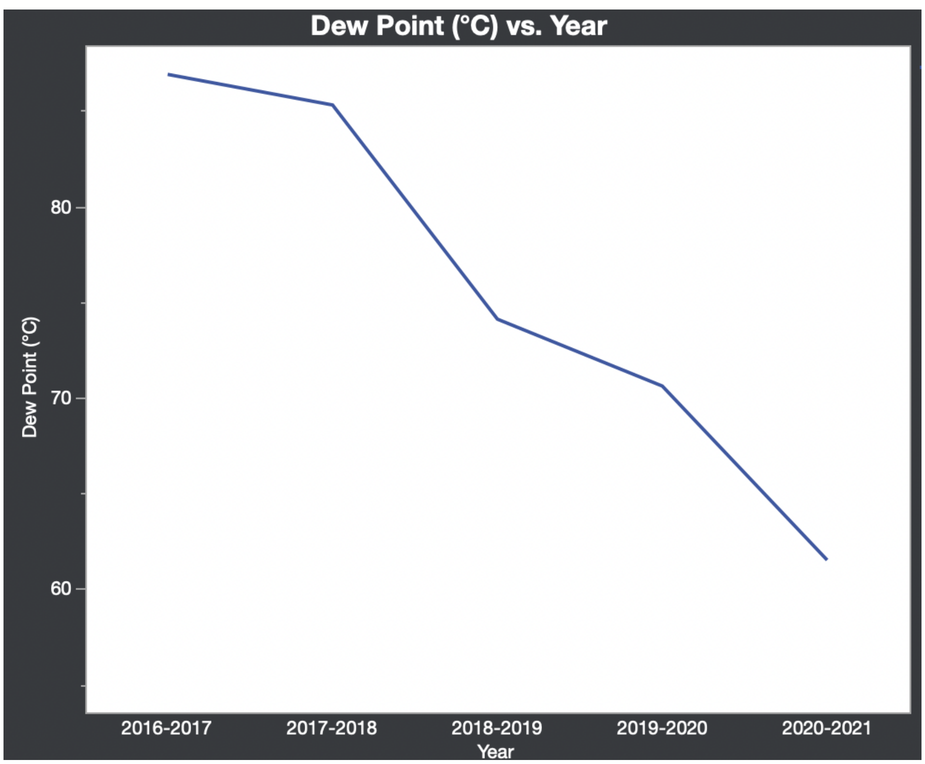

Figure 6. Line Graph of the relationship of Dew Point to Year

Cultivation Year	Number of Missing Values for Degree-days
2016-2017	4
2017-2018	1
2018-2019	5
2019-2020	28
2020-2021	29

Table 5. Number of missing values for the Degree-Days for soil temperature variables for each lower threshold