Assessing the Impact of Climate Change on Corn Quality in Iowa:
A Comprehensive Data Analysis of Environmental Variables
4000 Level

Abstract and Introduction

This project aims to explore the impact of weather data on the quality of corn produced in the US, specifically in Iowa. In recent years, climate change has accelerated and the environment has been constantly changing. Global temperatures have risen 1.98℉ from 1901 to 2020, but climate change is more than just temperature[1]. Weather patterns that are vital to agriculture production have been changing, such as wind and precipitation. I wanted to examine the effect of some of these shifting weather factors on agriculture in the US. Corn is the most widely produced crop in the US and Iowa is the country's largest producer.[2] The hypothesis of this study is that changes in weather data, including evapotranspiration, air temperature, and precipitation, have a significant negative effect on the quality of corn produced in Iowa. This project aims to improve our understanding of the relationship between weather data and corn quality and also to provide info that can potentially be used to mitigate these effects. Prior research has explored the relationship between the warming climate and corn production, with studies showing that the warming climate can significantly decrease corn yields.[3] However, these studies have focused on the relationship between weather and yields rather than quality. By examining the effect of weather data on corn quality, I hope to contribute to existing research and provide a more complete understanding of the environmental factors that impact corn production.

---

[1]"Climate Change Impacts," National Oceanic and Atmospheric Administration, accessed April 25, 2023, https://www.noaa.gov/education/resource-collections/climate/climate-change-impacts.
[2]"Crops," USDA ERS - Crops, accessed April 25, 2023, https://www.ers.usda.gov/topics/crops/.
[3]Jeff Mulhollem, "Warming Climate to Result in Reduced Corn Production; Irrigation Blunts Effect," Penn State University (Penn State News), accessed April 25, 2023, https://www.psu.edu/news/research/story/warming-climate-result-reduced-corn-production-irrigation-blunts-effect/#:~:text=%E2%80%9CIn%20our%20study%2C%20depending%20on,21.5%25%2C%E2%80%9D%20he%20said.

Data Description and Exploratory Data Analytics

First I looked for a dataset on corn production. The main governance on crop production is by the United States Department of Agriculture. On the USDA website there are data analysis tools such as CropScape and the Cropland Data Layer. These are tools that model the density of certain crops over a map of the united states. Production details were measured in pixels which represented a certain amount of crops per gridded layer. While this was insightful information it was not exactly what I was looking for. I headed over to the National Agriculture Statistics Service website which has a tool called quick stats. Using this tool I was able to select different commodities, categories, states, or data and date ranges. From the NASS quick stats, I was able to query weekly data on the quality of crops in Iowa from 2002 to 2022. The corn data had many columns but the relevant ones were the dates and the classification of crop qualities. Each week had 5 entries, each entry being a crop quality factor(Excellent, Good, Fair, Poor, Very Poor).

Year	Period	Week Ending	Geo Level	State	State ANSI	Commodit	Data Item	Domain	Domain Category	Value
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT EXCELLENT	TOTAL	NOT SPECIFIED	15
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT FAIR	TOTAL	NOT SPECIFIED	13
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT GOOD	TOTAL	NOT SPECIFIED	71
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT POOR	TOTAL	NOT SPECIFIED	1
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT VERY POOR	TOTAL	NOT SPECIFIED	0
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT EXCELLENT	TOTAL	NOT SPECIFIED	18
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT FAIR	TOTAL	NOT SPECIFIED	13
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT GOOD	TOTAL	NOT SPECIFIED	68
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT POOR	TOTAL	NOT SPECIFIED	1
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT VERY POOR	TOTAL	NOT SPECIFIED	0

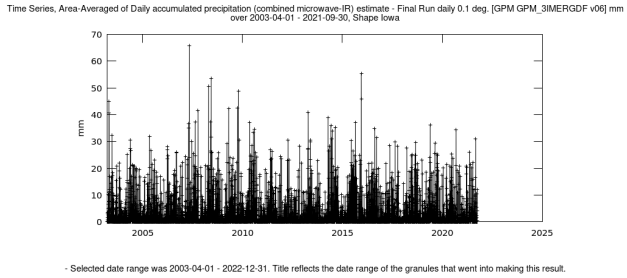
Above are entries for two weeks' worth of data. The import columns are the Week Ending which is currently characters but will be converted to dates later. Additionally, the data item column is technically qualitative, however when combined with the value column gives sort of a hybrid data value - the percentage of 100 of a certain quality. Clearly, I had a lot of data cleaning to do but this dataset gave me the factors I was looking for.

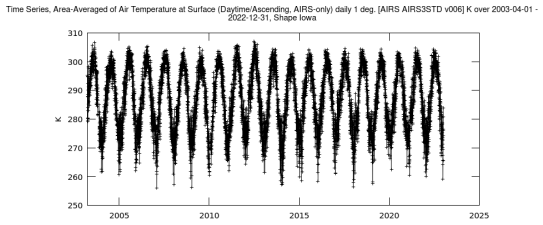
After obtaining corn quality data I needed weather data to see how the two were correlated. At first, I tried using NLDAS which stands for North American Land Data Assimilation System. This massive dataset definitely contained all the environmental factors I wanted to analyze however the sheer scope of data provided proved to be a little challenging for me to handle. I tinkered with a small subset of the data in python testing different methods however in the end I decided to find a different approach. Instead, I decided to use the NASA Giovanni tool. "Giovanni is a NASA Goddard Earth Science Data and Information Services Center (GES DISC) Distributed Active Archive Center (DISC) web application that provides a simple, intuitive way to visualize, analyze, and access Earth science remote sensing data, particularly from satellites[4]". Giovanni could access data from models such as NLDAS, parse, and return relevant info. I used Giovanni to create a bounding box over Iowa and set a custom time range. I then went through and selected daily data only - the corn quality data was weekly and I did not want weird overlaps in differences of week ranges, instead choosing to generate weekly data myself. You are then able to export information by CSV instead of the complicated models that NLDAS offered. Based on the US Geological Survey crop growth is affected by four main factors: terrain, climate, soil properties, and soil water.[5] Using this information I chose 6 different environmental factors from Giovanni that fell under these categories. The six were daily accumulated precipitation(mm) from NASA GPM, air temperature(K) at the surface from NASA AIRS, perceptible water vapor(cm) from NASA MODIS, evapotranspiration (kg/m^2), plant canopy surface water (kg/m^2) and root zone soil moisture (kg/m^2)from NASA GLDAS. To explain some of the factors - Precipitable water vapor is the vertically integrated amount of water

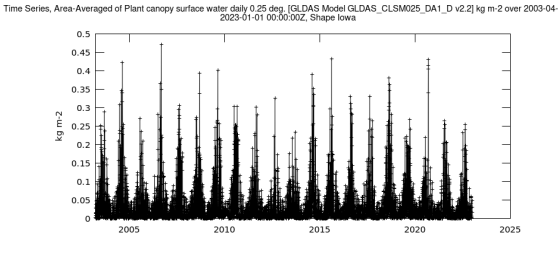
[4]NASA Earth Science Data Systems, "Giovanni," NASA (NASA), accessed April 25, 2023, https://www.earthdata.nasa.gov/technology/giovanni

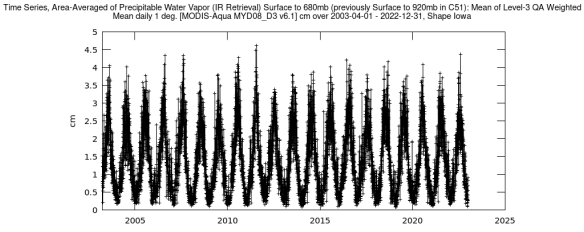
[5]Nancy T. Baker and Paul D. Capel, "Environmental Factors That Influence the Location of Crop Agriculture in the Conterminous United States," United States Geological Survey, accessed April 25, 2023, https://pubs.usgs.gov/sir/2011/5108/.

vapor in the atmosphere[6], Evapotranspiration is the sum of all processes by which water moves from the land surface to the atmosphere via evaporation and transpiration[7], and canopy surface water is the presence of water in a form of dew or interception on the canopy surface[8]. The GPM, AIRS, MODIS, and GLDAS are all different datasets compiled from different measurement device, but mainly satellite data. The GLDAS is a global version of the NLDAS model I mentioned earlier. Below are the time series, and area-averaged plots generated by Giovanni for each environmental factor I am examining. Immediately it is clear that the factors tend to cycle by year, however within the years there is a lot of variance. These time series graphs gave me a pretty good idea of the different ranges and variances within the data even before analysis.





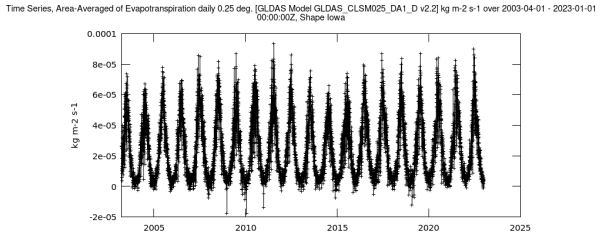


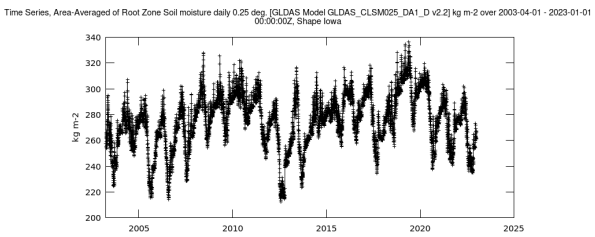


[6]Vicki Kelsey, "Atmospheric Precipitable Water Vapor and Its Correlation with Clear-Sky Infrared Temperature Observations," Atmospheric Measurement Techniques (Copernicus GmbH, March 18, 2022), https://amt.copernicus.org/articles/15/1563/2022/.

[7]"Evapotranspiration and the Water Cycle Completed," Evapotranspiration and the Water Cycle | U.S. Geological https://www.usgs.gov/special-topics/water-science-school/science/evapotranspiration-and-water-cycle.

[8] https://www.sciencedirect.com/science/article/pii/S0034425721005095



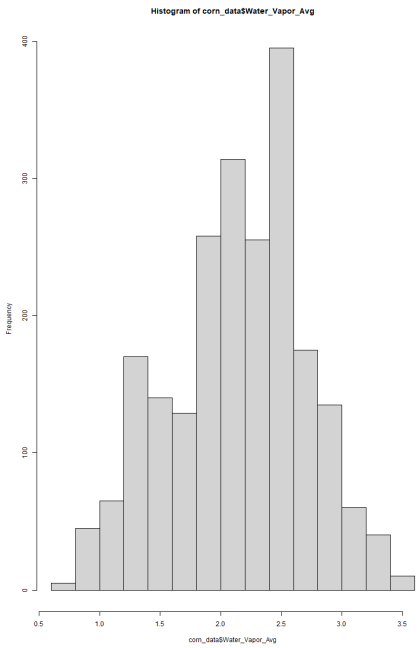


Analysis

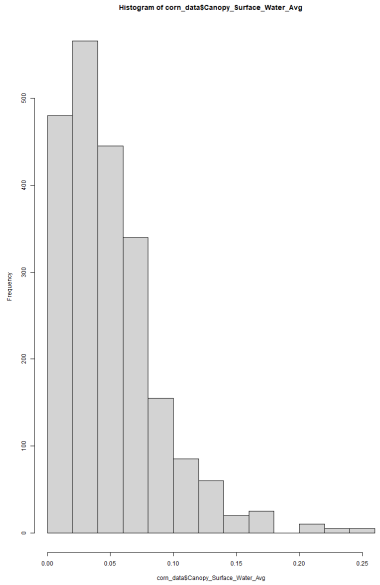
My analysis started with a ton of cleaning and reformatting. To begin, I imported the corn data and the six environmental datasets. While gathering my data I had already determined that I needed to parse daily environmental data to fit the weekly data. I manually deleted the first eight rows of each environmental data, which contained information from Giovanni about urls, the certain queries made to get that dataset, and the source of the data. I had originally tried to do this within R but ran into issues with recreation column headers after row deletion. I then ran summary statistics on all datasets to examine variable types and start looking at distributions.

The corn quality data had a “Week Ending” variable that was in characters so it had to be converted to date format in order to analyze it. The “Data Item” column was the second important column as it had different levels of crop quality and finally, the “Value” column included the actual percentage of each level of crop quality. The six environmental factors had a couple of issues - surface air temperature and water vapor had character values in the data columns so I had to convert them into numeric ones instead. This introduced NA values that I then removed.

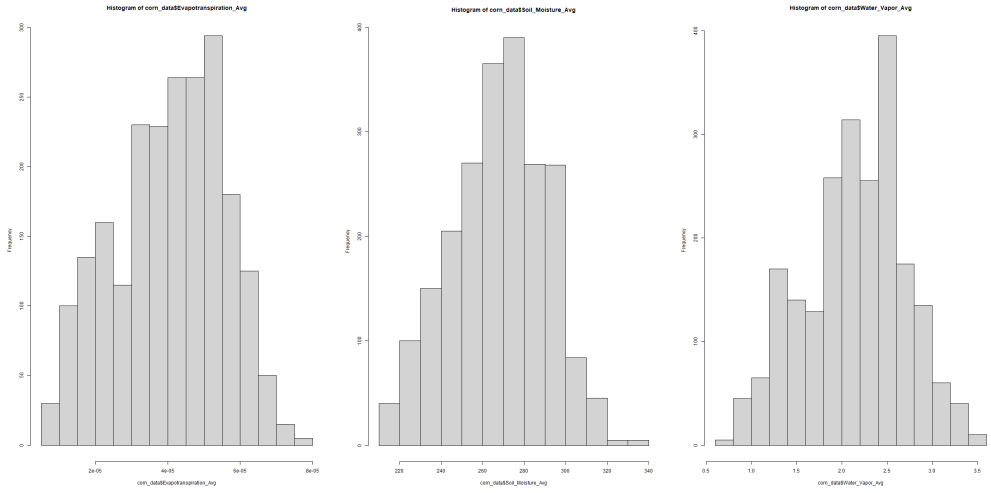
I made the corn data sorted by ascending instead of descending because the 6 environmental factors all had ascending dates. This was relatively easy as I just converted the week ending column into date values instead of characters which let me sort the rows using basic R functions. The corn data included some data that the environmental datasets were not in the date range of so I deleted all rows in corn quality data that was before 4/1/2003. Then to insure correct data formatting the time column in the environmental datasets to actual date values instead of characters as well. As discussed previously, I needed a way to the daily aggregate environmental data on a weekly basis. To do this a week starting and ending column was created in corn quality, this was calculated by subtracting 1 and 7 days from a specified date. In order to assist in my data compilation I created a function called "get_avg_past_7days" that took in a dataframe, a variable name, and a date. Based on these three parameters, it calculated the correct date range for a week before the given date value. If the date value in the dataframe was within this range, it took its variable value and calculated the sum of all seven days. I then looped through corn data and ran this function for each of the environmental variable datasets. This created a new column in the original data for each of the environment variables which had the summed data from the week. I went through many iterations of this function and the loop to generate the right data, there were a lot of issues applying the function on different data frames across 6 different variables. After cleaning and processing the data, histograms were created for each of the six environmental variables. Surface air temperature was observed to be normally distributed with no significant outliers. It has a few local peaks and valleys but generally seemed to be a stable variable.



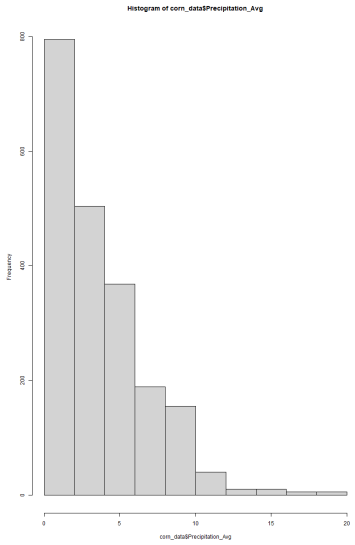
Canopy surface water, on the other hand, was heavily skewed to the right, indicating a significant difference between the highest and lowest values, but with enough data values in the extremes such that the skew was not caused by a couple of outliers. While not disqualifying this data from modeling it is definitely a factor to watch out for during data analysis.



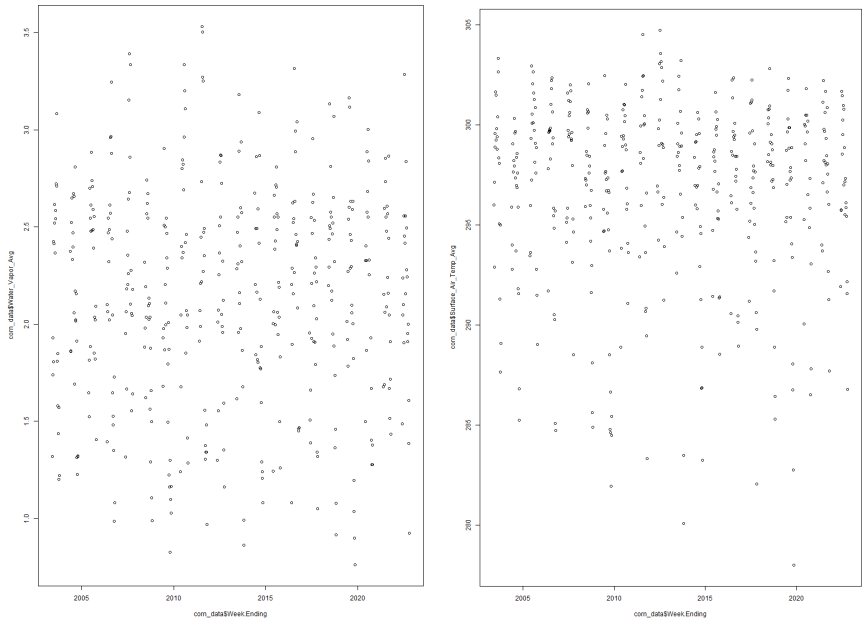
Evapotranspiration, soil moisture, and water vapor were all mostly normally distributed, with a slight left skew for evapotranspiration but no outliers. Soil moisture and water vapor both looked similar to evapotranspiration in terms of their distributions



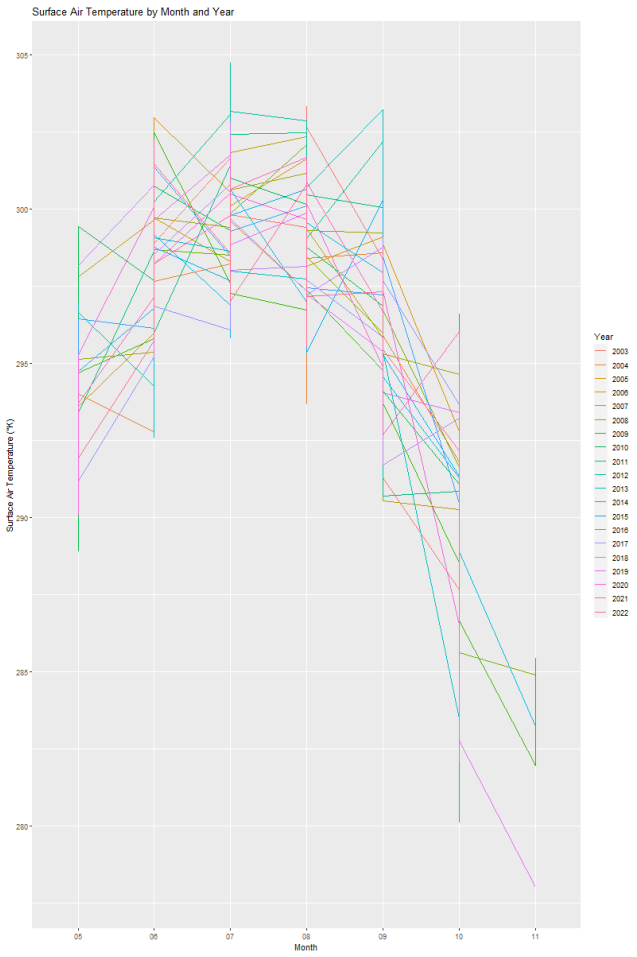
However, average precipitation was heavily skewed to the right, indicating that there were many days without rainfall. This skew was not due to outliers and the variable was still viable for use in data analysis.



I plotted each variable in relation to the week-ending date, however, it was pretty clear from these graphs I would have to analyze data year by year not just overall. Here are two examples of such graphs. Just looking at these, it is hard to make any analysis or assumptions about the data.



Using the observations drawn during the previous graphs, I wanted to have my x-axis be the different months of the year, where each year is graphed as a different colored line. I then made 6 different graphs where the y value was the 6 different environmental factors. This was achieved first for surface air temperature by grouping the data by year and month and calculating the mean surface air temperature. These separate year groups were then plotted. The plot showed a glaring problem in my data analytics.



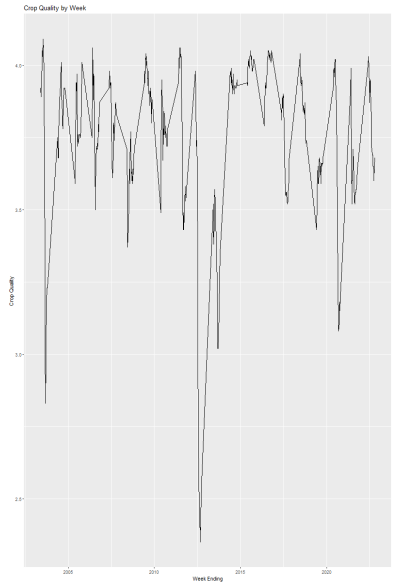
Due to the nature of the corn quality dataset, each date value had 5 entries. Each of these 5 entries showed up on each plot 5 times and any models developed would be nullified by these "repeat" entries.

Year	Period	Week Ending	Geo Level	State	State ANSI	Commodit	Data Item	Domain	Domain Category	Value
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT EXCELLENT	TOTAL	NOT SPECIFIED	15
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT FAIR	TOTAL	NOT SPECIFIED	13
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT GOOD	TOTAL	NOT SPECIFIED	71
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT POOR	TOTAL	NOT SPECIFIED	1
2022	WEEK #21	5/29/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT VERY POOR	TOTAL	NOT SPECIFIED	0
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT EXCELLENT	TOTAL	NOT SPECIFIED	18
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT FAIR	TOTAL	NOT SPECIFIED	13
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT GOOD	TOTAL	NOT SPECIFIED	68
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT POOR	TOTAL	NOT SPECIFIED	1
2022	WEEK #22	6/5/2022	STATE	IOWA	19	CORN	CORN - CONDITION, MEASURED IN PCT VERY POOR	TOTAL	NOT SPECIFIED	0

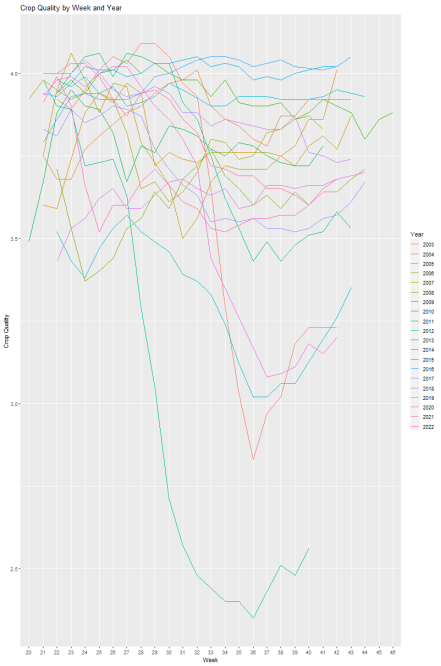
I decided to combine these columns into one clear “quality score”. This was calculated by assigning numerical representations to each quality where excellent was 5 ranging to very poor which was 1. The quality percentages were then multiplied by their numerical value. In the example shown above week 21’s quality score would be calculated like so:

$$(5 * 0.15) + (4 * 0.71) + (3 * 0.13) + (2 * 0.1) + (1 * 0.0) = 4.18$$

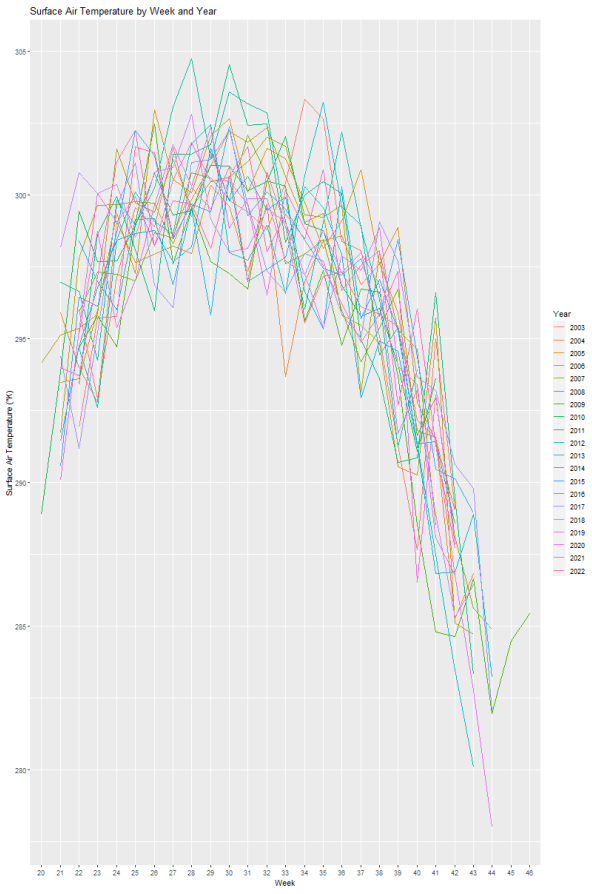
Hypothetically this sounded like an easy change, I would just create a new data frame with these new scores and carry over any other relevant information. However in practice, actually implementing this took a ton of trial and error. In the end, I settled on creating a quality_scores vector that contained(excellent, good, fair, poor, and very poor). I then used dplyr to create a new data frame called corn_data_new by grouping by week ending. I then summarized to create summary statistics for each group. For each crop quality variable I used the sum function to reference the vector I created and then ad up quality scores for the week. The ifelse function proved to be very useful when assigning these values. The scores are multiplied by the value and then divided by 100, then added together to get the total quality score for the week. Finally, I ungroup and return a flat data frame. I plotted the new data frame receiving the following plot.



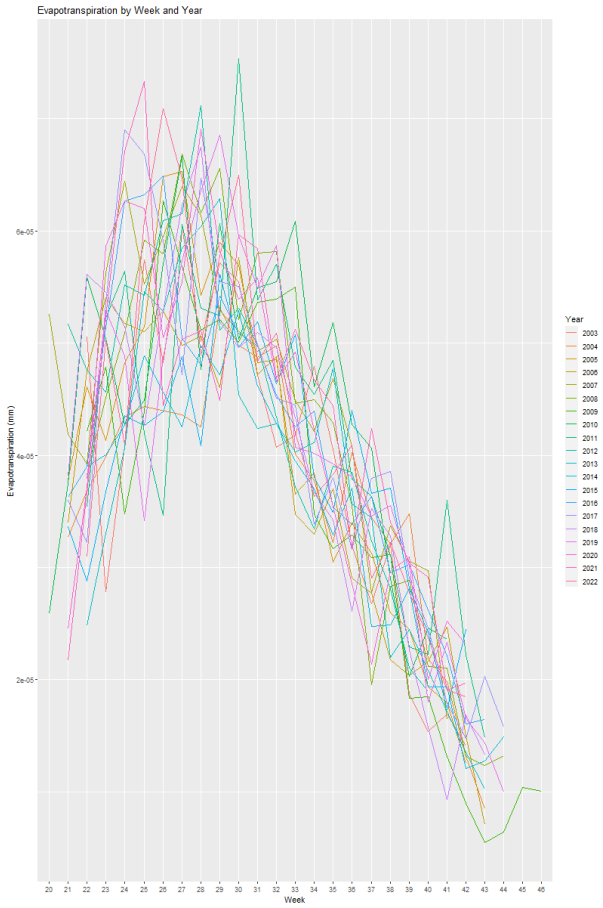
From this graphical representation, it was observed that crop quality can vary greatly even within the same year. However, this graph of crop quality changes over time did not provide much insight into my hypothesis. To address this issue I grouped the data by year and week and displayed them as separate lines.



This new graph showed that the quality of corn in recent years has been on the downturn. As discussed in my introduction, climate change affects a lot of factors related to crop growth and this was a promising observation supporting my hypothesis. Now to take a look at the different environmental factors year to year. Some of these factors made sense graphically and I will be discussing those.



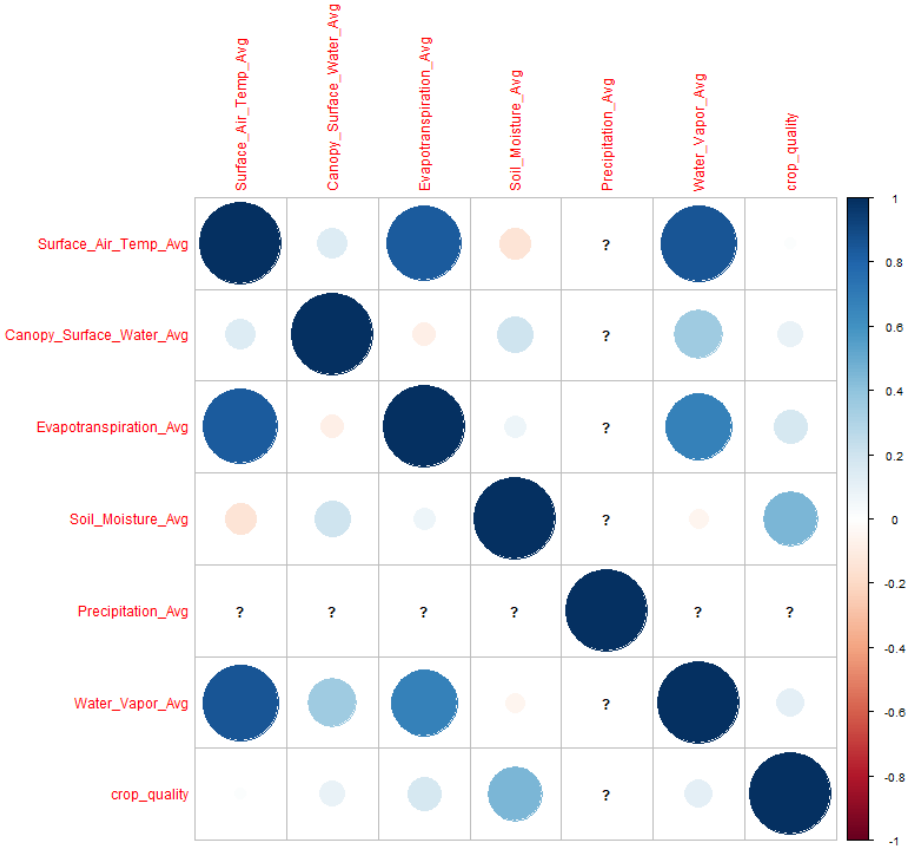
Surface air temperature on average has increased from 2003 to 2022, the lighter blue and green lines represent more recent years while the orange and burgundy lines represent past years. Most of the recent years have been peaks in surface air temperature within this graph.



It is evident from this graph, evapotranspiration seems to be on the climb - This makes sense as while average temperature increases evaporation also increased as well. While it is impossible to determine the direct effect of this increase on crop quality it is definitely worth noting.

I conducted chi-squared tests to determine the impact of various environmental factors on crop quality, including soil moisture, precipitation average, water vapor, evapotranspiration, canopy surface water, and surface air temperature. The p-values for most of these factors were not statistically significant, with values around 0.37. However, evapotranspiration had a p-value of 0.02371, indicating a higher correlation with crop quality compared to the other factors. Despite the low p-values for the other factors, I decided to keep them in the data analysis for further

examination. I also generated a correlation plot to visualize the correlations between crop quality and environmental factors. The plot indicated that soil moisture and evapotranspiration had the highest correlation with crop quality.



Models

Based on the mostly normal distribution of the environmental variables and the values calculated from the chi-squared test I decided to conduct a linear regression analysis to determine the effect of environmental variables on corn quality. The first linear regression model I ran had crop quality as the response variable and the six environmental columns(Soil_Moisture_Avg, Precipitation_Avg, Water_Vapor_Avg, Evapotranspiration_Avg, Canopy_Surface_Water_Avg, and Surface_Air_Temp_Avg) as predictor variables.

```
Residuals:
    Min      1Q  Median      3Q     Max
-1.0413 -0.1388  0.0500  0.1868  0.5024

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.043e+01  2.421e+00   4.307 2.07e-05 ***
Soil_Moisture_Avg        5.640e-03  7.309e-04   7.716 9.22e-14 ***
Precipitation_Avg       -8.537e-03  5.849e-03  -1.460 0.145154
Water_Vapor_Avg          1.995e-01  5.344e-02   3.733 0.000216 ***
Evapotranspiration_Avg   5.544e+03  1.991e+03   2.784 0.005610 **
Canopy_Surface_Water_Avg 2.795e-01  4.877e-01   0.573 0.566947
Surface_Air_Temp_Avg    -2.981e-02  8.294e-03  -3.595 0.000364 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2697 on 410 degrees of freedom
  (23 observations deleted due to missingness)
Multiple R-squared:  0.2618,    Adjusted R-squared:  0.251
F-statistic: 24.23 on 6 and 410 DF,  p-value: < 2.2e-16
```

The regression results showed that soil moisture, water vapor, surface air temperature, and evapotranspiration were significant predictors of corn quality with all their p-values less than 0.05. These four results made a lot of sense as each of these factors affects some important growth factors for plants. High evapotranspiration means that there is a high demand for water which might affect water distribution to each crop. However, when evapotranspiration is low it means there is too much water, potentially stunting crop growth from waterlogging. The amount of water vapor, which affects humidity and the air temperature, in either extreme can also

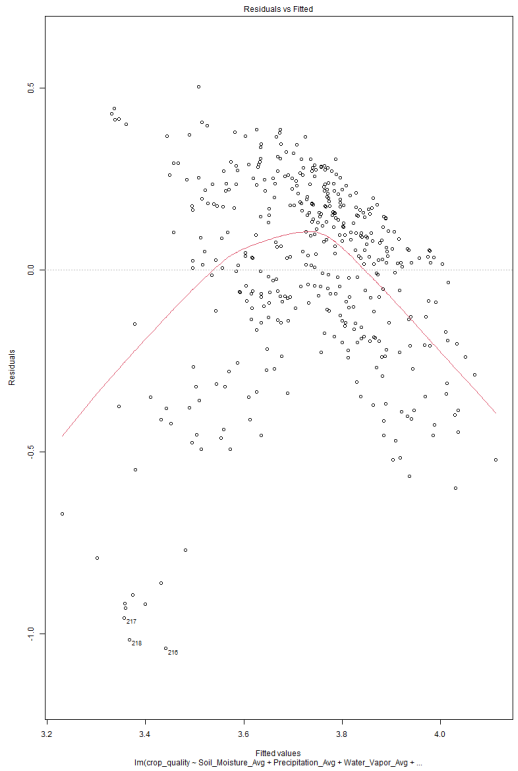
negatively affect plant growth. Surface air temperature is affected by both of these previous factors and is therefore important to crop growth and quality as well. Lastly soil moisture having the lowest p value makes sense as well. It determines factors such as nutrient uptake, water intake, and amount of photosynthesis. The surprising result was seeing that the precipitation average and surface water were not significant($p>0.05$). When discussing environmental factors in my introduction I was sure one of the main factors was rainfall. A possible explanation for this is the skewdness of the rain data - a majority of days it does not rain so when it does the data is skewed heavily to the right. In the future, I would try to find another rainfall variable, not just one that measures daily average precipitation. The adjusted R squared value for the first model was 0.251, indicating that around 25% of the variation in corn quality could be explained by the six environmental variables. While not very high, this r-squared value was still significant and supported the analysis and the original hypothesis. To further investigate the relationship between corn quality and environmental variables I added the year and week number to the data. Corn quality and environmental variables tend to vary throughout each year. I thought by adding these two I would be able to increase the accuracy of my regression model. I did a preliminary check by running chi-squared tests for both year and week and received promising results. The p-value for year was 2.2e-16 which meant it had a significant relationship to the crop quality. The second linear regression model I ran had all the predictor variables just with year and the week number added.

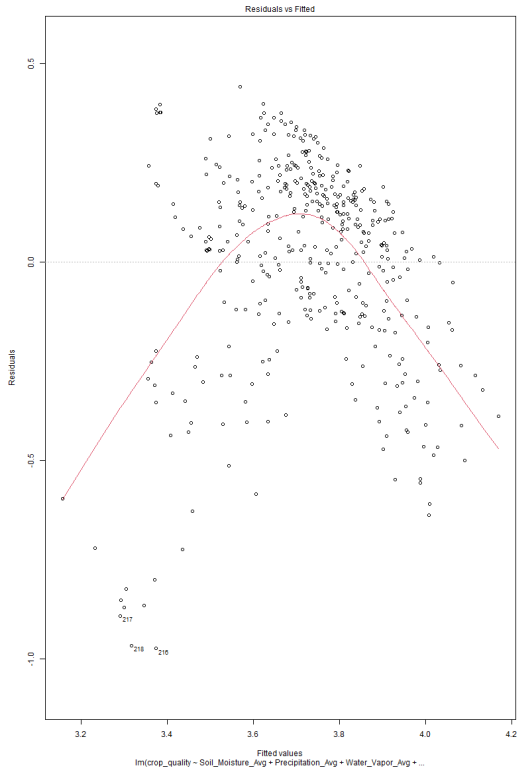
```
Residuals:
     Min       1Q   Median       3Q      Max
-0.97466 -0.13071  0.05826  0.17822  0.44067

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.806e+01  5.584e+00   6.816 3.38e-11 ***
Soil_Moisture_Avg         7.153e-03  8.142e-04   8.785  < 2e-16 ***
Precipitation_Avg        -1.276e-02  5.902e-03  -2.161 0.031241 *
Water_Vapor_Avg           2.235e-01  5.232e-02   4.271 2.42e-05 ***
Evapotranspiration_Avg    5.253e+03  2.114e+03   2.485 0.013348 *
Canopy_Surface_Water_Avg  2.606e-01  4.855e-01   0.537 0.591698
Surface_Air_Temp_Avg     -2.940e-02  8.089e-03  -3.635 0.000314 ***
year                     -1.402e-02  2.596e-03  -5.401 1.13e-07 ***
week                      8.243e-04  3.234e-03   0.255 0.798954
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.261 on 408 degrees of freedom
  (23 observations deleted due to missingness)
Multiple R-squared:  0.312,     Adjusted R-squared:  0.2985
F-statistic: 23.13 on 8 and 408 DF,  p-value: < 2.2e-16
```

This model was very similar to the first one and it seems that adding the year and week did not have as significant of a change as I would have wanted. However, the R-squared value did go up to .2985 which shows a slight improvement in the accuracy of the model.

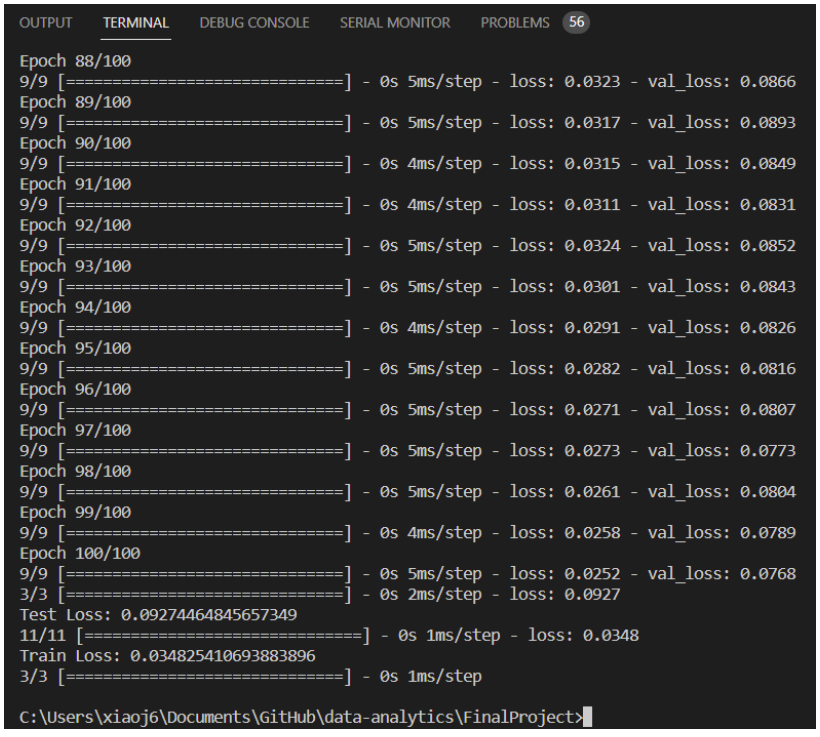


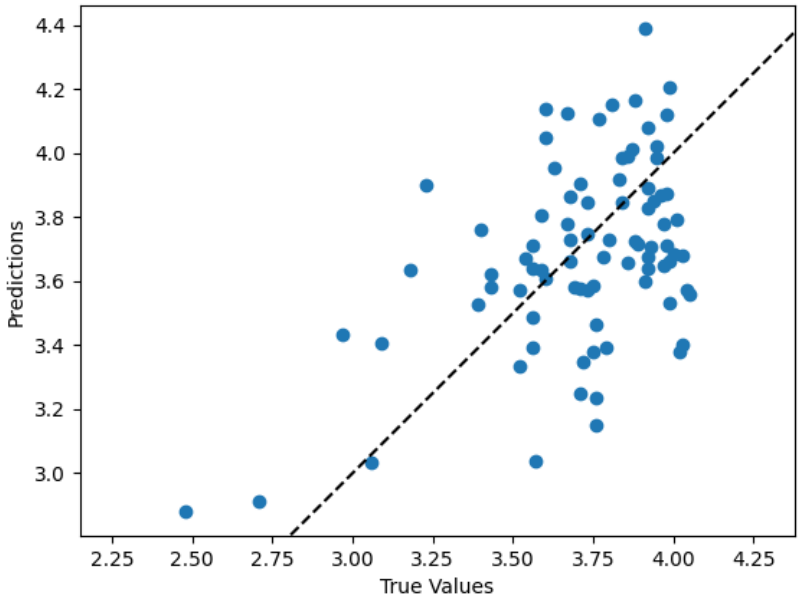


Examining the plotted residuals where the first model is on the left and the second on the right, you can observe a slightly better fit in the latter model. The fitted line for the residuals has a more pronounced curve in the second, representing the model trying to adapt to the upper range of values.

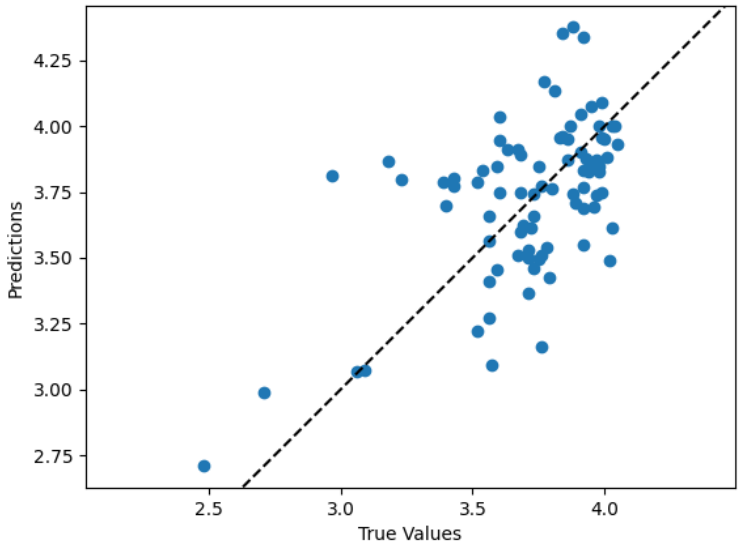
The second model I ran was a simple neural network. This model is very popular within the field of machine learning because they excel at fitting data. This does not only apply to linear relationships - neural nets are very good at identifying and modeling nonlinear relationships as well. When examining multiple environmental factors and their complex interactions with corn quality a neural network takes into account the different relationships they have. Second a neural network can identify which relationships are important even when not specified by the user. If something like evapotranspiration is extremely important but is not immediately visible, the neural net will determine this importance and weight by itself. While neural nets can be very useful it is also important to keep in mind some of their downsides, the main one being overfitting. I was not too worried about this as my training and testing datasets were made from the same set of data so there were no worries about it performing poorly on “new data”.
To create this neural network I first clear the data by dropping categorical data such as week ending and state. A neural network can only be created with numerical values. I then split the data into training and testing sets - 80% of the data will be used for training and 20% will be used for testing. Due to the variance in variable values I wanted to standardize my input values to combat any issues or biases encountered during training. I do this using the standard scaler from sci kits which is a package that provides machine-learning tools in Python. I build the actual neural network model using Keras, a high-level neural network API that runs on top of

TensorFlow which is a free and open-source software library for machine learning and artificial intelligence. My model consists of three layers the input layer with 64 neurons and ReLU, a hidden layer also with 64 neurons and ReLU, and finally an output layer with just one neuron and a linear activation function. Activation functions determine whether a neuron should activate or not and ReLU is simply an activation function that outputs the input directly if it is positive, otherwise, it will output zero. I then compile the model with the Adam optimizer and mean squared loss error function. Initially, I train it for 100 epochs and store the history of the training process in the history variable. This is done so relevant info can be accessed later for analysis and graphing. The performance of the model is calculated by computing the mean squared error loss. One of the last steps is to use the trained model on the previously made test set and the predicted values are flattened. Finally, I graph the performance and loss of the model using the matplotlib library.

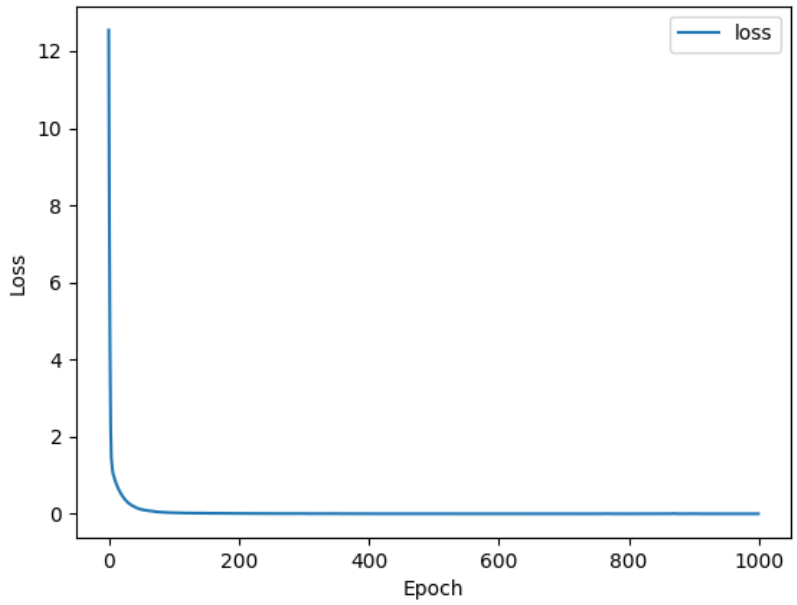




These were the results of the first neural network model I implemented. As shown in the terminal outputs above for 100 epochs the test loss was 0.0927 and the training loss is 0.0348. It is safe to say that the neural network greatly outperforms the linear regression model. The chart above displays the predicted values vs the true values.



Even with such a low training error, I wanted to see the performance when increasing the epochs. For 1000 epochs the test loss was 0.0879 and the train loss was 0.0151. The chart above graphs the predicted vs actual values for 1000 epochs. The change is not drastic bus is definitely visible, the points are close to the true fit line. For 10000 epochs the test loss was 0.0761 and the train loss was 0.0150. As epochs increase the training error decreases. With more time I would have liked to run 100,000 and 1,000,000 epochs. However, it is important to note that as epochs increase the rate at which loss decreases is exponential. This can be demonstrated by this graph of loss in 1000 epochs



Conclusion

The goal of this project was to create models that predicted the quality score of corn based on six environmental factors: surface air temperature, canopy surface water, evapotranspiration, soil moisture, water vapor, and average precipitation. I cleaned and processed the data to create weekly aggregated data for each of the environmental factors. Surface air temperature and evapotranspiration were mostly normally distributed with no significant outliers, while canopy surface water, soil moisture, water vapor, and average precipitation were all skewed. When attempting to graph data a glaring error was found: each date value in the corn quality dataset had five entries, which would nullify any models developed. To overcome this issue, I created a new quality score column by assigning numerical values to each quality level and multiplied these values by their corresponding percentage. This score was calculated for each week of data and combined into a single quality score column for each week. From the models used I concluded that environmental variables such as soil moisture, water vapor, surface air temperature, and evapotranspiration have a significant effect on corn quality. While both the linear regression model and the neural network demonstrated this, the neural network was more convincing with a significantly lower loss. Overall, the analysis supports the hypothesis that changes in weather data, including evapotranspiration, air temperature, and precipitation, have a significant negative effect on the quality of corn produced in Iowa. For subsequent explorations I would stick to the neural network and explore a larger variety of environmental factors and expand the scope of crops I am looking at.

References

"Climate Change Impacts." National Oceanic and Atmospheric Administration. Accessed April 25, 2023. https://www.noaa.gov/education/resource-collections/climate/climate-change-impacts.

"Crops." USDA ERS - Crops. Accessed April 25, 2023. https://www.ers.usda.gov/topics/crops/. Earth Science Data Systems, NASA. "Giovanni." NASA. NASA. Accessed April 25, 2023. https://www.earthdata.nasa.gov/technology/giovanni#:~:text=Giovanni%20is%20a%20NASA%20Goddard,having%20to%20download%20the%20data.

"Evapotranspiration and the Water Cycle Completed." Evapotranspiration and the Water Cycle | U.S. Geological Survey. Accessed April 25, 2023. https://www.usgs.gov/special-topics/water-science-school/science/evapotranspiration-and-water-cycle.

Kelsey, Vicki, Spencer Riley, and Kenneth Minschwaner. "Atmospheric Precipitable Water Vapor and Its Correlation with Clear-Sky Infrared Temperature Observations." Atmospheric Measurement Techniques. Copernicus GmbH, March 18, 2022. https://amt.copernicus.org/articles/15/1563/2022/.

Mulhollem, Jeff. "Warming Climate to Result in Reduced Corn Production; Irrigation Blunts Effect." Penn State University. Penn State News. Accessed April 25, 2023. https://www.psu.edu/news/research/story/warming-climate-result-reduced-corn-production-irrigation-blunts-effect/#:~:text=%E2%80%9CIn%20our%20study%2C%20depending%20on,21.5%25%2C%E2%80%9D%20he%20said.

Nancy T. Baker and Paul D. Capel. "Environmental Factors That Influence the Location of Crop Agriculture in the Conterminous United States." United States Geological Survey. Accessed April 25, 2023. https://pubs.usgs.gov/sir/2011/5108/.