Predicting State-Wide Cotton Yield using Geospatial Data

Gregory Gee

3/10/24

Abstract

Forecasting cotton yield has immense interest in the agriculture world considering how valued the crop is world wide. However, current methods of predicting cotton yield remain significantly unreliable since they rely on human assumption. Additionally, these forecasts are significantly impacted by climate change. This paper aimed to use spatial models to better predict cotton yield. Prior papers on predicting cotton production have yielded fair results utilizing non-spatial models [1]. Therefore, the focus of this study was to use Geospatial Linear Regression to predict cotton yield. Through this study, it was found that using climate, temporal, and geographic data cohesively produces the most accurate predictions. Additionally, the Random Forest model, with an MSE of 2.67, performed better than the Geospatial model, with an MSE of 8.22. This concludes that non-spatial models are better for predicting state-wide cotton yield when compared to spatial models. This research provides valuable results to the agriculture setting as well as different aspects to further this research.

1. Introduction

Cotton is an indispensable commodity in the modern world, with it being one of the most important fiber crops world wide. More than 100 countries produce cotton over 33.2 million hectares with an average annual cotton production of 18.9 million tons [3]. The United States is the 3rd biggest cotton producing country, behind China and India, producing nearly 20 million cotton bales in 2019 to 2020 [4].

Cotton is a crop that is particularly sensitive to climate, causing climate change and rising temperature to drastically affect its yield. Optimal cotton production requires a total rainfall of 850-1050 millimeters per year. 70 to 95 degrees fahrenheit produces optimal cotton and higher temperatures for continuous days adversely affects the yield [5]. The sensitivity of cotton to climatic conditions suggests how accurate predictions can help farmers make more informed decisions regarding planting and harvesting.

This research question involves testing if state-wide cotton yield relies on geospatial data using machine learning models and autocorrelation algorithms.

Machine Learning models are vital because these models are what is producing the predictions. We chose to use 5 different models, each contrasting from one another, because some machine learning algorithms work better depending on the problem. Through a process of comparing each model's result with a metric, we can identify which model performs better.

Therefore, the aim of this study was to 1) use geospatial, temporal, and climate data to predict state-wide cotton yield through spatial models, 2) use non-spatial models to compare results, 3) explore the significance of climate versus state data in predicting cotton yield, 4) identifying state cotton yield hotspots and key climate parameters.

Using geospatial data refers to a specific location, presenting a potentially better way to predict cotton yield. However since our study predicts state-wide cotton yield, the coordinates are of the entire state. In this study, our models will incorporate specific coordinates and state geometries to predict cotton yield.

Non-spatial models refers to models that do not take geospatial data into account in their algorithms.

In this paper, we will research if using geospatial data and taking into account neighboring states will build a model that predicts state-wide cotton yield better than an ordinary non spatial model.

2. Background

In the past, papers have used non spatial models with only climate parameters to predict cotton yield in Brazil [1], however the R^2 score was relatively low. Their highest performing model was Random Forest Regressors with an R^2 adjusted score of 0.38, followed by a variety of models such as Multiple linear regression, K Neighbors Regressor, Multi-layer Perceptron and Gradient Boosting. Additionally, it was found that evapotranspiration and water storage were the two parameters that most positively correlated to cotton production. Although they had coordinates of data points in their dataset, they chose not to use geospatial linear regression. To build on to this paper, geospatial linear regression with climate parameters will be used. Additionally, our study will not be restricted to specifically climate data since geospatial, temporal, and state data will be incorporated.

In a paper predicting within-field cotton yields using publicly available datasets [2], similar models were used such as Random Forest Regression and Gradient Boosting Machines. This paper's main focus was using infield data, such as soil properties to observe and predict how cotton yield predictions change depending on the stages of cotton growth (Squaring, Flowering and Boll-fill). However, this paper used both soil property data as well as one climatic parameter of rainfall. Random Forest had a highest R^2 score of 0.48 and GBM had a highest R^2 score of 0.7. The data used were from 2 irrigated paddocks on an Australian farm from 2014 - 2016. Since the amount of data, 2 years on 2 separate farms, was so low, we chose to use a wider broader dataset. Differentiating from this paper, we chose to focus on more climatic and geographic data rather than soil property data.

Autocorrelation, which is used many times for geospatial analysis, refers to the correlation between values of a variable at different locations within a geographic space. Autocorrelation discovers if values with a certain attribute are similar or different than the values around it. Autocorrelation in geospatial analysis can help build and identify spatial patterns. Statistical measures such as Moran's i show if there is autocorrelation.

Finally, choropleth maps are one of the main types of graphs used in this paper. Choropleth mapping uses different shades of colors and geographic boundaries to show a certain quality or value in a graph. For example, choropleth mapping is used to display Local Moran's i hotspots.

3. Dataset

We used a public dataset found on Kaggle with many climate parameters, specific US states, years, and corresponding cotton yield. This dataset was created by Abhirup Saha, a researcher at Penn State.

The dataset had 28 columns with 528 samples. For our models, the data had to be numerical. The majority of the data (climate parameters) is numerical other than the years and the State. To perform geospatial analysis, the state name by itself isn't enough data, so a US Census dataset with state names, geometry of states, and coordinates were merged.

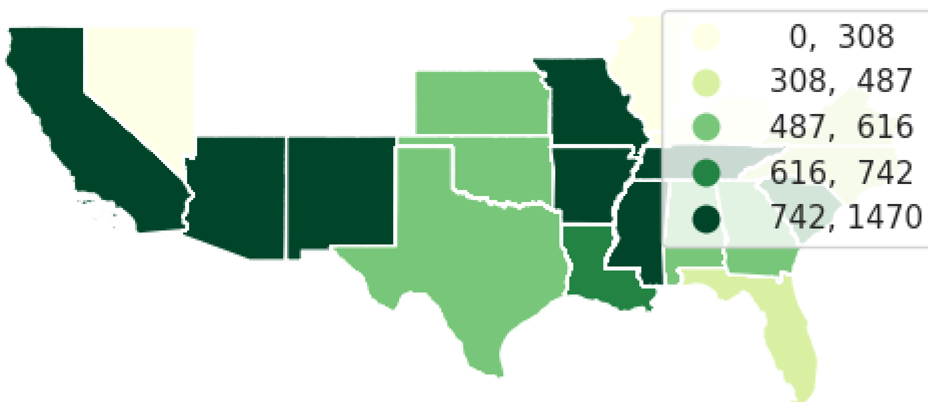
See Appendix A for variable descriptions. Variable Descriptions directly taken from the Cotton Kaggle dataset [6].

Data Pre-processing

With the first step in data-preprocessing, the dataset contained temporal data, ranging from 1970 - 2002. We chose to use all the data across the years and enumerated them into integers (0, 1, 2, 3…) The categorical state names were enumerated using one hot encoding.

Missing data were imputed using the mean values of available data. However, since analyzing geographic patterns was a main focus, the data without states names were dropped. Additionally, a log function was performed to the "Planted", "Harvested", and "Yield" columns because Geospatial Linear Regression outputs results on a logarithmic scale. We performed the research with this dataset, however for the graphs, we chose to show non log graphs for human interpretation.

Figure 1: Choropleth map showing distribution of Cotton Yield by state.



4. Methodology / Models

In terms of our problem, we had 4 different approaches to analyze the relationships between cotton yield and states. The first 3 approaches involved non-spatial models.

Approach 1: Testing with only Climate Data (reduced Model)

Our 1st baseline approach involved dropping all state names, which left all columns with climate data and temporal data (26+ columns). Approach 1 was created to compare it with Approach 2 and 3. This approach was used to ask the questions: 1) Was climate data enough information, or even better information than geographic data for predicting cotton yield? 2)Was state names hurting cotton yield predictions?

Approach 2: Testing with only State Names (reduced model)

Our 2nd baseline approach involved dropping all variables besides Harvested tons, Planted tons, Cotton Yield, and State names to compare with Approach 3. Harvested and Planted tons were kept because it is directly related to cotton yield since not all planted + harvested amounts were the same. Approach 2 was created to compare it with Approach 1 and 3. This approach was used to ask the questions: 1) Was statedata enough information, or even better information than geographic data for predicting cotton yield? 2)Was climatic data hurting cotton yield predictions?

Approach 3: Testing with Climate and State Data (Fully Adjusted Model)
Our 3rd approach was testing nonspatial models with our combined dataset, climate and state names, to compare it with the other approaches. Approach 3 was created to compare it with Approach 1 and 2. This approach was used to ask the questions: 1) Is cotton yield best predicted when using geographic and climatic data cohesively?

Approach 4: Testing with Geospatial data (Climate + State + Geospatial Data)
The 4th approach was our main objective and other approaches were used to compare nonspatial models to spatial models in terms of state-wide cotton yield. The base dataset did not have Geospatial data of certain coordinates and geometries, so the base dataset was merged with a US Census dataset according to state names. This additionally was essential for Geospatial Linear Regression Our coordinates (longitude + latitude) weren't leading to a specific point, but rather the state as a whole. With our final dataset, we had all that was needed to perform geospatial analysis.

Models Used

We chose to use 5 Main models (4 non-spatial + Geospatial Linear Regression); Linear Regression, KNN Regressor, Dense Neural Network, Random Forest Model, and Geospatial lagged Regression.

Linear Regression

Linear regression was chosen because it is a baseline, simple model, which could be the start to compare other models. Linear regression computes a linear relationship between the multiple x features and y value, to predict the cotton yield. This linear regression relationship sets coefficients to each feature. This coefficient is commonly found in many machine learning models and signifies how important each feature is to producing a prediction. Coefficients can be negative, however importance is seen through the magnitude or absolute value of the coefficient. Therefore, this coefficient is a number and the higher the absolute value of the coefficient, the more important the feature is.

Random Forest Regression

Random Forest Regression was chosen because it is a classification and regression model that performs highly in past related research. At the core, Random Forest models are many decision trees. Decision Trees are similar to flow charts, where each numerical data feature is put in a node and travels down computing branches to produce a prediction. Essentially, a random forest model works by merging many decision trees together to produce a prediction.

KNN Regression

KNearest Neighbors (KNN) is normally a classification model using k neighbors to similar data points to make a prediction. Based on how far/what these data points, it makes a prediction. However, KNN Regression is a modified version of this commonly used classification model to apply it to supervised learning problems. This classification model used for regression was chosen to see if "classification models" are better for predicting cotton yield.

Neural Networks

Neural networks are a model algorithm inspired by our brain.Neural Networks are commonly used for deep learning problems, such as detecting things in images. Nonetheless, Neural Networks can also be used for numerical problems. In our case, we wanted to find if these deep learning algorithms would perform well in our rather simple numerical problem. For our Neural Network architecture, we used 5 layers, 3 hidden layers + input + output. Optimizer was "adam_optimizer" and loss function was "mean_squared_error"

Geospatial Lagged Linear Regression

We used PYSALs SPREG (Spatial Regression) model to perform our geospatial Linear Regression. The difference between Geospatial Linear regression vs. Linear Regression is that Geospatial linear regression uses a spatial weights matrix based on the relationship between coordinates, the overall data set, and cotton yield. Essentially, this model assumes that there are stronger relationship features with features closer than features that are farther away.

Spatial Autocorrelation Measures

Multiple spatial autocorrelation methodologies were used to explore spatial relationships in our dataset along with supporting the results from geospatial linear regression, such as global and local moran's i and p and z value. Moran's i is a statistical measure used to interpret autocorrelation. Global Moran's i is a number ranging from -1 to 1 that interprets IF and how strong an autocorrelation is. Negative values show a negative or hurting relationship while positive values show a positive correlation. Local Moran's i is an array of numbers showing how strong an autocorrelation is relating to neighboring states. In other words, Local Moran's i shows hotspots. Local Moran's i can be shown through graphs. Often when using geospatial data, autocorrelation can be of random chance in that specific data set, so p value is a number as well that denotes the chance of the autocorrelation being random. This number ranges from 0-1 and the lower the number, the less of a chance it is by random chance. A number below 0.05 generally shows that autocorrelation is most likely not by chance.

MODEL METRICS:

3 model metrics were chosen to evaluate our models performance, which were Mean Squared Error (MSE), Mean Absolute Error (MAE), and R Squared Score (R2). Essentially, the models used predict a list of values which are then used to evaluate how well it did to the actual results.

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n} \tag{1}$$

Mean Squared Error (1) is a measure of the average squared difference between the predicted and actual values. The measure is squared to penalize large errors for human interpretation. Oftentimes, error isn't evidently shown in the measure, so MSE is used to help humans interpret the error. The lower error shows more accuracy.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{2}$$

Mean Absolute Error (2)is similar to MSE, however it takes the absolute error instead of squaring. MAE shows the actual average error between predicted and actual values. The lower error shows more accuracy.

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

R Squared Score (3), also known as coefficient of determination, is our final metric, which shows how much of the data variance did the model capture. This score ranges from 0 to 1, with 1 indicating a perfect fit or correlation. The higher number shows more data correlation.

5. Results

Model Development:

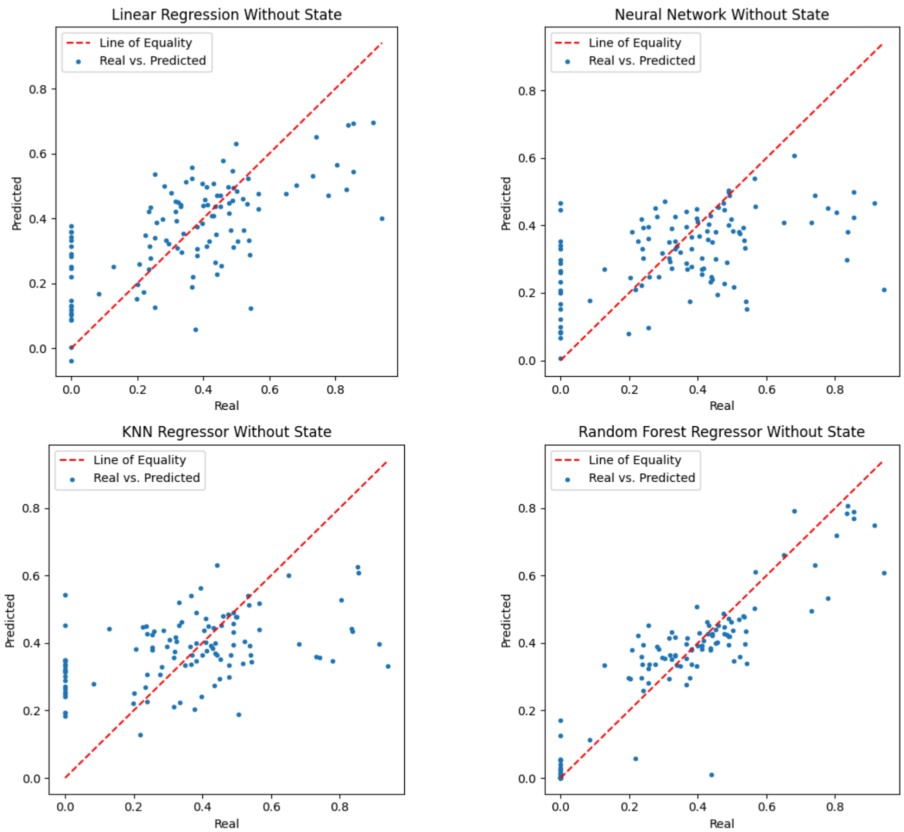
Table 1: Model metrics results for our machine learning models. The columns are the different models: LR is Linear Regression, RF is Random Forest, KNN is K-Nearest Neighbors Regressor, NN is Dense Neural Network, GLR is Geospatial

Linear Regression. The rows represent our different approaches shown through model metrics; MSE is Mean Squared Error, MAE is Mean Absolute Error, R2 is R^2 Score.

Indexes	LR	RF	KNN	NN
With States + Climate MSE: MAE: R2:	 7.47 1.46 0.87	 2.67 0.59 0.95	 18.38 1.94 0.704	 9.17 1.23 0.85
Only Climate MSE: MAE: R2:	 11.69 1.87 0.81	 8.15 0.94 0.86	 16.29 1.72 0.78	 10.49 2.49 0.69
Only States MSE: MAE: R2:	 5.3 1.6 0.75	 7.89 0.87 0.87	 21.38 2.43 0.707	 6.94 1.35 0.89

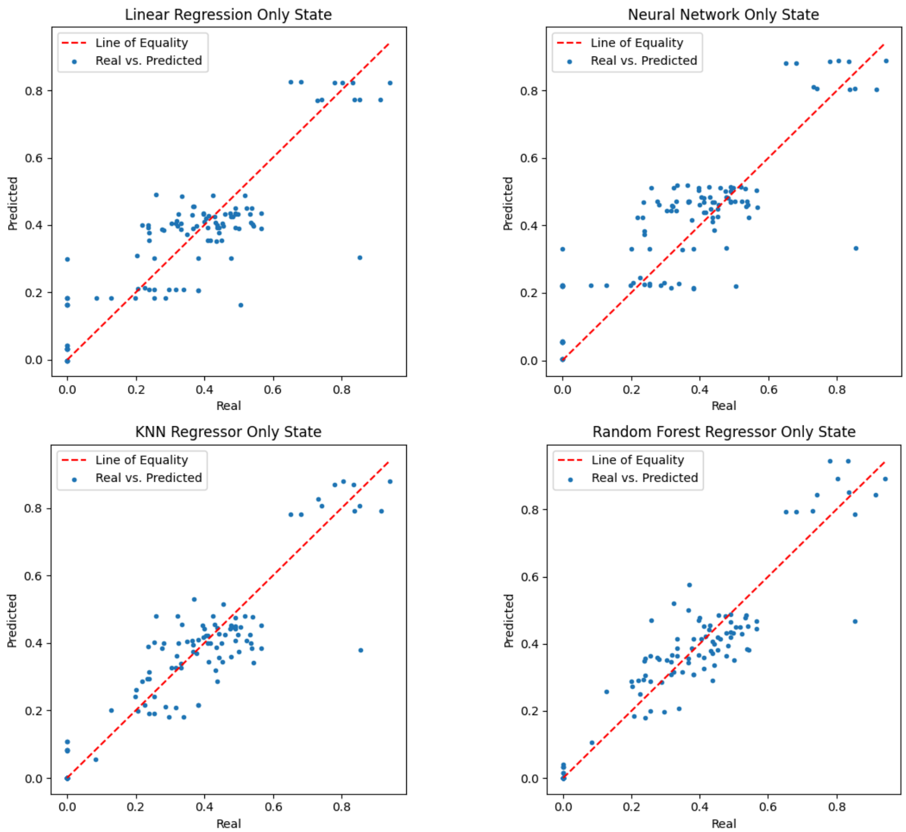
	GLR			
With Geospatial Data MSE: MAE: R2:	 8.22 1.68 0.83			

Figure 2: Scatter Plot graph of Approach 1 results. Title shows which model followed by Without State, meaning the approach with only climate. The red line (line of equality) shows a perfect prediction matching with the result.



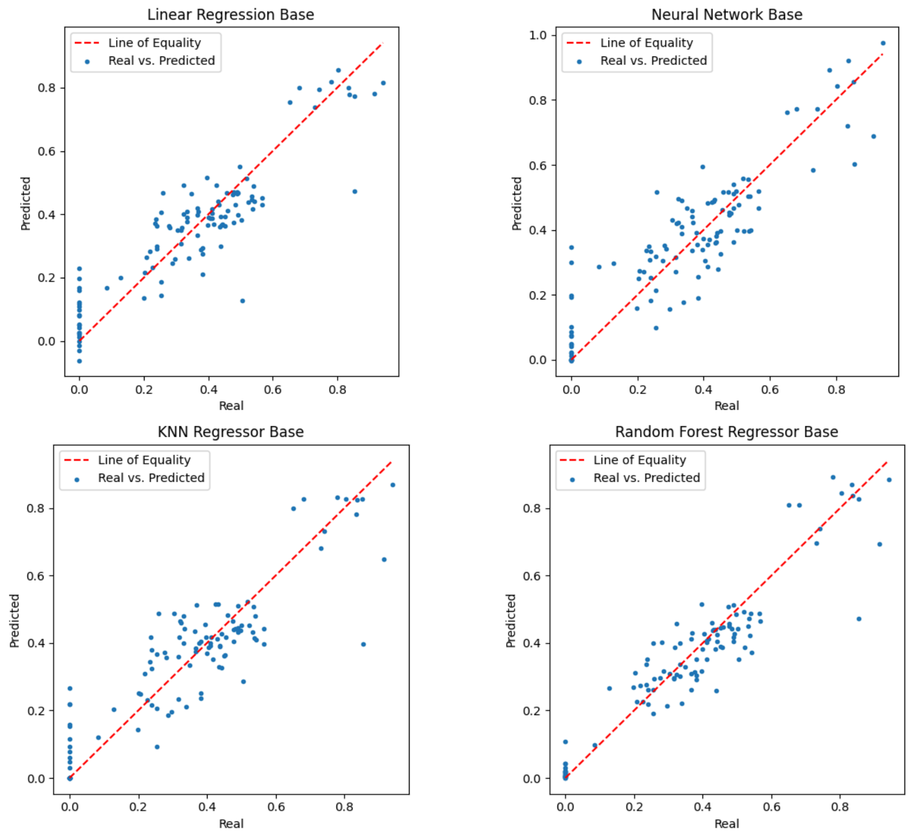
For the first baseline approach (Only Climate Data) the model results from most simple to advanced models were Linear Regression with an MSE of 11.69 and an R^2 score of 0.81. Next is K-nearest Neighbors with an MSE of 16.29 and an R^2 score of 0.78. Following that were Random Forest Regressor with an MSE of 8.15and R^2 score of 0.86. Finally was Neural Networks with an MSE of 10.49 and an R^2 score of 0.81

Figure 3: Scatter Plot Graphs of Approach 2 (Only state names). Only State meaning the approach without climate variables. The red line (line of equality) shows a perfect prediction matching with the result.



For the second baseline approach (Only State Data), the model results from most simple to advanced models were linear regression with an MSE of 5.3 and an R^2 score of 0.75. Next is K-nearest Neighbors with an MSE of 21.38 and an R^2 score of 0.707. Following that were Random Forest Regressor with an MSE of 7.89 and R^2 score of 0.87. Finally was Neural Networks with an MSE of 6.94 and an R^2 score of 0.89.

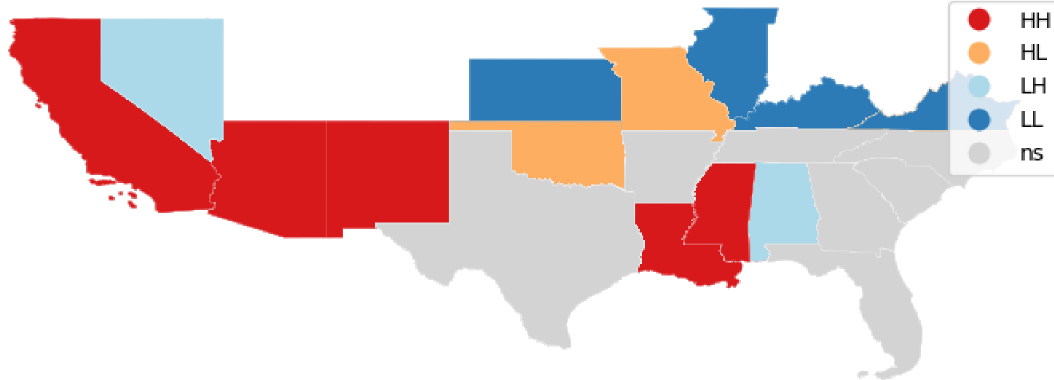
Figure 4: Scatter Plot graphs of our 3rd approach (Fully Adjusted Models) results. Title shows which model was used, followed by "Base", meaning climate + state data (Original Dataset). The red line (line of equality) shows a perfect prediction matching with the result.



With our fully adjusted approach, the model results from most simple to advanced models were linear regression with an MSE of 7.47 and an R^2 score of 0.87. Next is K-nearest Neighbors with an MSE of 18.38 and an R^2 score of 0.704. Following that were Random Forest Regressor with an MSE of 2.67 and R^2 score of 0.95. Finally was Neural Networks with an MSE of 9.17 and an R^2 of 0.85. Scatterplots of the

For geospatial linear regression, the model's results were an MSE of 8.22 and an R^2 score of 0.83.

Figure 5: This is a local moran's i figure, denoting spatial autocorrelation in states. To interpret the legend, HH shows a high spatial correlation surrounded by neighboring high correlation states. HL shows high spatial correlation surrounded by low neighboring states. LH shows low individual correlation surrounded by high states. LL shows low individual correlation surrounded by low states. NS means not significant in terms of detecting autocorrelation.



Feature Importance:

Through Geospatial Linear Regression reports, the values with the highest variable coefficients were Harvested with a variable coefficient of 0.589, Planted with a variable coefficient of 0.336, Precipitation Anomaly with a variable coefficient of -0.22, and Palmer Z-index Anomaly with a coefficient of 0.146.

Autocorrelation results:

Our autocorrelation measures achieved a global moran's i of +0.29 and a local moran's I described in Figure 6. We found cotton hotspots in California, New Mexico, Arizona, Louisiana and Mississippi. We achieved a P-value of 0.01.

6. Discussion

Model Results:

Previously, the model results showed that KNN Regressors performed extremely well in all non-spatial approaches and outperformed Geospatial Linear Regression. In terms of Mean Squared Error, the fully-adjusted approach performed the best, followed by Only Climate approach, Only State approach, and then Geospatial Linear Regression. Based on the result comparison above, we can conclude that using only states or climate is enough to predict cotton yield, meaning climate and geographic data are necessary. Additionally, from these results, it can be concluded that using only climate data has more of an effect to achieve high quality results rather than only states data. With Geospatial Linear Regression Results coming in last, it can be concluded that our spatial approach is not the best for predicting state-wide cotton yield. This can be told based on the higher overall metric scores, R^2, MSE, MAE. From this, we can conclude that climate and geographic data are necessary to produce high quality predictions, while geospatial data is not necessary.

Figure 6: Geospatial Linear Regression Report, stating variable coefficients/importance.

Variable	Coefficient	Std.Error	t-Statis
CONSTANT	7.07712	4.88714	1.44
Year	-0.04553	0.01846	-2.46
Planted (1000 Acres)	0.33590	0.05348	6.28
Harvested (1000 Acres)	0.58936	0.05254	11.
Average Temperature Value	-0.07427	0.07199	
Average Temperature Anomaly	0.13079	0.11266	
Maximum Temperature Value	0.04006	0.06830	
Maximum Temperature Anomaly	-0.04411	0.11469	
Minimum Temperature Value	-0.05489	0.05423	
Minimum Temperature Anomaly	-0.03163	0.10902	
Precipitation Value	0.07680	0.09721	0.79
Precipitation Anomaly	-0.22473	0.11343	-1.9
Cooling Degree Days Value	-0.00199	0.00408	
Cooling Degree Days Anomaly	-0.00294	0.01197	
Heating Degree Days Value	-0.00371	0.00314	
Heating Degree Days Anomaly	0.00579	0.00849	
Palmer Drought Severity Index (PDSI) Value	0.01093		
Palmer Drought Severity Index (PDSI) Anomaly	0.12223		
Palmer Hydrological Drought Index (PHDI) Value	0.05938		
Palmer Hydrological Drought Index (PHDI) Anomaly	-0.10848		
Palmer Modified Drought Index (PMDI) Value	0.01266		
Palmer Modified Drought Index (PMDI) Anomaly	-0.06117		
Palmer Z-Index Value	-0.05701	0.11385	-0.50
Palmer Z-Index Anomaly	0.14612	0.11279	1.

Based on the results, we found the explanatory variables with the highest coefficients were Planted with a coefficient of 0.336, Harvested with a coefficient of 0.589, Precipitation Anomaly with a coefficient of -0.225, and Palmer Z-index Anomaly with a coefficient of 0.146. Additionally, since Planted and Harvested variables are directly related to cotton yield, the most important explanatory variables were palmer z index and precipitation anomaly.

Autocorrelation Discussion:

Regarding spatial statistical measures, we had a global moran's i of +0.29. This value being positive shows there is a positive correlation between States and Cotton yield. However, since Global Moran's i are in a 0-1 scale, the correlation strength is low. To display our Local Moran's i, Figure 6 shows a map of this relationship. From Figure 6, it can be determined that the states, California, New Mexico, Arizona, Louisiana and Misssipi are cotton yield hotspots.

Oftentimes, spatial correlation is by random chance because of an odd sample or data pulling. P-value shows a percentage that the autocorrelation is random. Our p value was 0.01, which shows that there is a 0.1% chance that our spatial correlation is random in comparison to our data.

Our geospatial model, with an MSE of 8.22, performed worse than our baseline models with climate + state data slightly.

From these results, we can conclude that there is spatial correlation from the P-value and Moran's i, however for predicting state-wide cotton yield, it is better to interpret states as individual entities rather than taking into account where it is geographically. Interpreting states as individual entities was concluded since our Fully adjusted approach without geospatial data yielded better results and lower error than a geospatial approach.

Furthering Research:

To further this research, several different approaches to accurately predict cotton yield could be used. For example, rather than predicting state-wide cotton yield, predictions of specific farms using Spatial models could be done. This would include using individual coordinates of farms. Additionally, comparing autocorrelation results of individual farm yields and state-wide yield could further this research as well. Further methods to build on this research could be utilizing different climate parameters to look for feature importance. In addition, data from more recent years affected by climate change could be tested. Finally, different models can be used to compare results with this research.

7. Conclusions

This research demonstrated and showed the results of using geospatial data to predict State-Wide Cotton Yield. A public dataset from Kaggle was used, consisting of various climate parameters, states, and temporal data. A US Census dataset was merged to give geospatial data to the dataset. Linear Regression, KNN Regressor, Random Forest, dense Neural Networks, and Geospatial Linear Regression were the models used. The Random Forest was the best performing model with the base dataset from Kaggle. From our results, we can conclude that there is a relationship between states and cotton yield, however this relationship is stronger when interpreting these states as separate entities, rather than being next to other states. We got to this conclusion based on our geospatial linear regression model having a relatively high R^2 score and MSE, however it was overperformed by a Random Forest model using state names. Additionally, states are better interpreted as separate entities since the model interprets state names as categorical data rather than specific coordinates near each other. Finally, ways to further research may be t changing the data and models used.

Acknowledgments

Extreme thank you to Inspirit AI for providing this excellent opportunity for me to guide my own Research Paper. Additionally, an extreme thank you to Ribhav Gupta, my mentor in this project, for guiding my research, answering questions, as well as helping give comments on this paper.

References

[1] Aparecido, Lucas Eduardo & Meneses, Kamila & Souza, Glauco & Carvalho, Mary Jane & Pereira, Washington & Silva, Paulo & Santos, Tatiana & Moraes, José. (2020). Algorithms for

forecasting cotton yield based on climatic parameters in Brazil. Archives of Agronomy and Soil Science. 68. 10.1080/03650340.2020.1864821.

[2] Leo, S., De Antoni Migliorati, M. and Grace, P.R. (2021) 'Predicting within-field cotton yields using publicly available datasets and machine learning', *Agronomy Journal*, 113(2), pp. 1150–1163. doi:10.1002/agj2.20543.

[3] Hussain, M. *et al.* (no date) *Efficacy of fertilizing method for different potash sources in cotton (gossypium hirsutum L.) nutrition under arid climatic conditions*, *PLOS ONE*. Available at: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0228335 (Accessed: 03 March 2024).

[4] *Cotton sector at a glance (no date) USDA ERS - Cotton Sector at a Glance. Available at: https://www.ers.usda.gov/topics/crops/cotton-and-wool/cotton-sector-at-a-glance/#:~:text=Cotton%20is%20planted%20from%20March,Georgia%2C%20Mississippi%2C%20and%20Arkansas. (Accessed: 03 March 2024).*

[5] *(No date) Ikisan. Available at: https://www.ikisan.com/ap-cotton-climateandsoils.htm#:~:text=Warm%20season%20(tropical)%20crop.,is%20necessary%20for%20higher%20yields. (Accessed: 03 March 2024).*

[5] *SAHA, A. (2023) Effect of climate change on commodity yields, Kaggle. Available at: https://www.kaggle.com/datasets/abhisaha97/effect-of-climate-change-on-commodity-yields?select=USA_test.csv (Accessed: 10 March 2024).*

Appendix A:

Variable Name:	Variable Description:
Year	The year of the data observation.
State	The name of the state where the data was recorded.
Planted (1000 Acres)	The area in thousands of acres that was planted with crops.
Harvested (1000 Acres)	The area in thousands of acres that was harvested for crops.
Yield (Pounds/ Harvested Area):	The average yield of crops in pounds per harvested area.

Average Temperature Value	The average temperature recorded during the specified period.
Average Temperature Anomaly	The deviation from the long-term average temperature for the specified period.
Maximum Temperature Value	The highest temperature recorded during the specified period.
Maximum Temperature Anomaly	The deviation from the long-term average maximum temperature for the specified period.
Minimum Temperature Value	The lowest temperature recorded during the specified period.
Minimum Temperature Anomaly	The deviation from the long-term average minimum temperature for the specified period.
Cooling Degree Days Value	The number of cooling degree days recorded during the specified period.

Cooling Degree Days Anomaly	The deviation from the long-term average cooling degree days for the specified period.
Heating Degree Days Value	The number of heating degree days recorded during the specified period.
Heating Degree Days Anomaly	The deviation from the long-term average heating degree days for the specified period.
Palmer Drought Severity Index (PDSI) Value	The value of the Palmer Drought Severity Index, which indicates the severity of drought conditions.
Palmer Drought Severity Index (PDSI) Anomaly	The deviation from the long-term average Palmer Drought Severity Index for the specified period.
Palmer Hydrological Drought Index (PHDI) Value	The value of the Palmer Hydrological Drought Index, which indicates hydrological drought conditions.
Palmer Hydrological Drought Index (PHDI) Anomaly	The deviation from the long-term average Palmer Hydrological Drought Index for the specified period.
Palmer Modified Drought Index (PMDI)	The value of the Palmer Modified Drought Index, which indicates modified drought conditions.

Value	
Palmer Modified Drought Index (PMDI) Anomaly	The deviation from the long-term average Palmer Modified Drought Index for the specified period.
Palmer Z-Index Value	The value of the Palmer Z-Index, which indicates drought conditions based on soil moisture.
Palmer Z-Index Anomaly	The deviation from the long-term average Palmer Z-Index for the specified period.