A Study of Nebraska Weather Trends, do Farmers Need to Take Early Action?

Sean Kerr

Bellevue University

DSC630: Predictive Analytics

Professor Andrew Hua

11/15/2022

Project Background

For this project I have chosen to take a look at data provided by the National Weather Service to determine if changes in future weather patterns necessitate earlier action by farmers in regards to drought. The NWS provides daily historical weather data broken down into multiple reporting stations which comprise a region. The region in particular that will be discussed in this project will be the Eastern Nebraska region. Not every station within the region has equipment to measure each of the categories the will be studied. It is my belief that by using the larger region as a whole, as opposed to just the Omaha area, we can better project future temperatures and rates of snowfall in a meaningful way. The original plan for this project was to use the previous ten years of historical weather data to predict our future temperature values and future snowfall rates. The plan has remained mostly intact with a change to the number of years being used, from 10 down to 6, this change was necessitated by the file size limiter placed by the NWS maxing out at around 6 years and a day. In order to account for this change the window of study has been changed from trying to predict the next 10 years of data to predicting the next 5 years of data.

My initial impression of the data was that due to the linear nature of time that this sort of model would be the most appropriate, which was not the case. After feedback and a bit of trial and error, it was determined that some sort of time series model would be more appropriate for this type of data. The models most considered were the ARIMA and SARIMAX models. It was ultimately decided that SARIMAX would be the most likely candidate due to the addition of seasons to the model which falls more in line with the cyclical nature of the weather data.

I believe that the best way to evaluate outcomes for this project is to compare our output with other studies that have conducted similar tests. The World Meteorological Organization or WMO has conducted a number of studies on the topic and predicts an increase in global temperature of about 1.5 degrees Celsius (*New climate predictions* 2021). This topic is something that effects the human population on a global scale and many groups are likely to have vested interest in the accuracies and availability of predictions and models for future weather. By comparing my outputs to studies of similar timeframes we can see if there is any major deviations from the more global consensus. If there are major issues with the results of my study I will need to determine why these major shifts occurred, outliers, bad data, incorrect model parameters or other issues could be a contributing factor.

Through this project I hope to gain a better understanding of our future weather patterns based on climate data that is sourced from a more recent time period instead of the more long term historical projections. I believe that these long term projections can often skew our ability to predict future weather data simply by the fact of changes in global climate that have already occurred. By using data from the last 6 years, I hope to find projections that better reflect the more recent trends in weather. It is my hope that this information will give farmers within the region a better idea of the possibility of drought and the possibility of limited access to our aquifer. By better understanding more current trends within data it is possible to ease the strain on the already overstretched water resources within the Eastern Nebraska Region.

There is inherent risk in any project. In this case it could be possible that the data used does not provide any evidence for major shifts in climate in future years. It could also indicate that the shift is so major there is not much we can do to account for it. In either case the risk to the original premise still provides useful information. There could also be risk in an inability to find a model the provides accurate results given the data we are using, although this is unlikely. It is also possible that our access to the data may be revoked, but as it is coming from a government source I believe that this is also very unlikely.

If this project does not work out I would like the contingency to fall in a similar vein in order to lessen the loss of progress and allow me to repurposed previous structures. I think that something like creating future projections for rain and snowfall may be a target. This still provides a similar goal to the original premise but narrows the scope somewhat. I would allow me to better understand climate related issues like snowfall related flooding or drought. In that same train of thought, this information would still provide useful projections for farmers in the region to better prepare for the upcoming seasons based on previous cycles of weather. If it is not possible to repurpose any of the data I think that an interesting topic would be to look at data surrounding the flow of traffic within the city of Omaha or a city of comparable size. This kind of shift would take a lot more time and effort to achieve so I would like to avoid this if at all possible.

As an additional note I would like to include the future possibility of additional data sets to supplement the information provided by the National Weather Service. I think that this addition would provide a boost to both the volume and accuracy of our data if used correctly and can improve the accuracy of our results if time permits me to do so. In addition the amount of factors recorded in these datasets will likely need to be trimmed due to the sheer scale at which they are recorded. This may have an impact on the overall integrity of the data so multiple iterations will likely need to be used to determine which information provides the greatest impact on our data.

Preliminary Analysis

With the initial data retrieved from the previously mentioned sources I do believe I will have a sufficient amount of data to answer the questions in mind. I think that I would like to incorporate additional data sources as time goes on to further refine that overall output of my chosen models. While more data is not always correct/necessary to find the results of a question, I believe that the sources that I would be incorporating in the future will provide high quality data to the, all ready, high quality data being used. The data in question is very linear in its structure, I believe that visualizations that complement this structure would be best in this case. I think that tools like the stacked bar or area charts work well to showcase trends in the data and give a better sense of the time scale in question. In addition I think similar tools will better highlight the rates at which extreme weather is occurring giving the end users a more in depth look at whether or not this rate is increasing a at glance.

As far as adjusting my data/questions goes, I think my original topic my have been too broad in its scope. I think the looking at a more narrow field of extremes in the weather would better suit my time frame while still falling in line with my original goals. In addition I believe that this change would better showcase the changes in weather patterns compared to the original premise. Initial results from my chosen models seem to indicate that there is a good chance that they will work well for the updated outlook on the project. I would like to include additional models in the future to more easily validate the findings in my data and try to uncover any unseen errors or interactions that may provide an incomplete or incorrect view of the data. I think that with the changes mentioned above the original expectations for the project are mostly still feasible. This may need to be adjusted again to any further changes in direction or scope, but for now I am satisfied with the current plan of action.

Model/Results

As mentioned previously the model that was chosen for this project was the SARIMAX model. This came as a result of trial and error, with other candidates including ARIMA and SARIMA. In order to get our data into a workable state there was some cleansing necessary from the original flat file provided. There was quite a few extraneous categories of data that were not relevant to the study such as latitude or longitude that were dropped from the table. In addition as stated above, many of the reporting stations did not provide full readouts of every data category so it was necessary to remove instances with null values in the needed categories. There were a couple of factors that I had overlooked initially such as leap years and the accidental inclusion of the first day of 2022. While these dates may have provided additional data to our model, they ultimately provided more trouble than they were worth with how they affected the shape of our dataset. Finally in order to more easily access our data, the various station readouts were grouped together by date.

The first iterations of the model provided less than satisfactory results with the visualizations showing an almost completely flat line for our predictions. It took a considerable amount of trial and error but the order and seasonal order that best fit the model, without providing results that indicated either and increase or decrease in temperature that would coincide with Earth becoming the Sun or the Sun disappearing all together, were 1,1,1 for our basic ARIMA order and 1,1,1,12 for our seasonal order. The biggest factor in choosing these numbers for our order was the issue of over differencing, after running some tests with the ADFULLER tool it was indicated that any number beyond our 1 in the second position would lead to a considerable amount of over differencing, greatly skewing the results of our predictions. In addition the 12 within our seasonal order was a key component that required the creation of a secondary table in which information was grouped and averaged by the month. This allowed the model to run on a 12 month cycle more closely factoring in the seasons.

These changes and decisions have provided results for a number of categories with enough statistical significance to be worth mentioning. The results of the model being show that the average temperatures, maximum temperature, and minimum temperature seem to fall relatively in line with WMO predictions. There seems to be a gradual increase in temperature over time with average and maximum temperatures seeing an increase between 2 and 6 degrees over the next 5 years. Minimum temperatures seem to show and inverse correlation, with the lowest predicted temperature varying drastically from previous years. It is important to note that 2019 and 2021 seemed to have significantly colder winters than previous years which could be skewing the overall trend somewhat. There has been some difficulty with the precipitation and snowfall categories within this model. Precipitation on its own is measured in such small values (1/10 of a millimeter) it was difficult for the model to provide readouts that really showed any change at all. Snowfall on its own provided an interesting piece of visualization due to 1 significant outlier year, 2019. In 2019 massive snowfall in both Nebraska and the Rocky mountains provided conditions for a devastating season of flooding and on its own greatly skewed future predictions. This is likely due to the size of our dataset and could likely be accounted for with the additional 4 years of historic data originally intended.

Conclusions/Recommendations

The results of the model seem to indicate that there are some factors that farmers will need to consider. With how closely the results seem to align with the generally accept consensus of a 4 degree rise in average temperature over the course of the next 5 years, it is likely that growing seasons will be impacted. While seemingly minor in the grand scale of things, a 4 degree increase is actually a major swing in temperature, this could greatly impact the amount of water needed for satisfy crops. With corn and soybeans being the staple crops for Nebraskan farmers, it is important to note the corn requires a significant amount of water per acre to sustain a healthy plant. Increases in temperature, like the ones predicted, will likely necessitate the need for increased irrigation to maintain the levels of moisture in the soil to sustain the needs of the crop. The result of this increase in water usage will very likely put additional strain on water resources in the area such the Ogallala aquifer or the Missouri river. Nebraska itself is one of the leading states in the total acreage of irrigated land and has rapidly begun turning to groundwater to make up the difference in required water allotment (Johnson et al., 2011). While relatively efficient when compared to other major agricultural producers, the user of this water can only be stretched so far. The increased use of finite water resources can only be sustainable for so long without proper safeguards in place.

One of the major contributors to the replenishment of groundwater resources is the amount of snowfall during the winter months. The aquifers themselves are very fragile and rely on underground pockets of very soft rock, if too much of the water is removed it is possible that the pockets will collapse rendering the area completely unusable for future water extraction. It is for this reason that snowfall and other forms of replenishment are so important to maintain the level of water needed within groundwater sources to keep up with farmer requirements. With the predicted downturn the amount of snowfall in future years this could become a critical issue with the sustainability of current water consumption practices. Farmers will likely have to ration water and begin storing water earlier in the season to account for the possible decline in water resources. On the other side of the coin, the increase in temperature earlier on in the season could indicate possibly earlier periods of growth for crops. This may allow farmers to reduce the impact of increased temperature issues later on in the season by getting a head start. In any case, with such dramatic changes in environmental conditions looking exceedingly likely the agricultural industry will need to keep a close eye on the expenditure of finite resources.

References

MIT. (n.d.). Groundwater. Mission 2012 : Clean Water. Retrieved November 19, 2022, from <http://web.mit.edu/12.000/www/m2012/finalwebsite/problem/groundwater.shtml>

US Department of Commerce, N. O. A. A. (2022, March 3). Climate. Retrieved November 19, 2022, from <https://www.weather.gov/wrh/climate?wfo=oax>

New climate predictions increase likelihood of temporarily reaching 1.5 °C in next 5 years. World Meteorological Organization. (2021, May 27). Retrieved November 19, 2022, from <https://public.wmo.int/en/media/press-release/new-climate-predictions-increase-likelihood-of-temporarily-reaching-15-%C2%B0c-next-5>

Artley, B. (n.d.). Time Series Forecasting with Arima , Sarima and SARIMAX. Towards Data Science. Retrieved November 19, 2022, from <https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6>

Johnson, B., Thompson, C., Giri, A., & Van NewKirk, S. (2011). (rep.). *Nebraska Irrigation Fact Sheet* (pp. 1–6). Lincoln, NE: University of Nebraska Lincoln.

Study shows how melting snow affects underground water reserves. Water Quality Products. (n.d.). Retrieved November 19, 2022, from https://www.wqpmag.com/groundwater-treatment/news/10972865/study-shows-how-melting-snow-affects-underground-water-reserves