Climate Change Relates to Hurricane Activity

DS 4002

Group 15: Nathan Wan (leader), Alex DeStephanis, & Jillian Haig

11/1/24

Outline

1. Project Details
2. Data Acquisition & Explanation
3. Analysis Plan & Justification
4. Tricky Analysis Decision
5. Bias & Uncertainty Validation
6. Results & Conclusion
7. Next Steps
8. References
9. Closing

PROJECT DETAILS

- **Motivation:** Understand if climate change variables relate to hurricane activity over time, since many people feel that hurricanes have become more common

- **Research Question:** Does the number of days a hurricane is present in the Atlantic and Pacific oceans rise over time with climate change?

- **Hypothesis:** The number of days per year a hurricane is present in the Atlantic and Pacific oceans increases alongside climate change factors such as sea surface temperature, indicating correlation

- **Modeling Approach:** Confirm relationships with Granger's and Cointegration tests, and then determine if the analysis can go further and use the explanatory series to model the number of days a hurricane is present yearly using a VAR model

- **Goal:** Determine if there is conclusive evidence that climate change indicators relate to an increase in hurricane days through the usage of these tests and modeling procedures

Variable	Description	Potential Responses
Year	Integer variable representing the year (1971-2021)	1971, 2003, 2021
Atlantic_Hurricane_Days	Float variable representing the number of days a hurricane was present over the Atlantic	28.75, 10.0, 49.25
Pacific_Hurricane_Days	Float variable representing the number of days a hurricane was present over the Pacific	38.5, 5.25, 21.5
Total_Hurricane_Days	Float variable representing the total number of days a hurricane was present	67.25, 71.0, 49.25
CO2_emissions	Float variable representing the world CO2 emissions, annually	15502787000.0, 24330945000.0
Population	Integer variable representing the annual human population on Earth	3920805030, 7954448405
CSIRO_Adjusted_SeaLevel_Inch	Float variable measured in inches representing amount the sea level has increased since 1971	4.881889759, 9.594329444
Surface_Temperature_Change	Float variable representing the change in Earth';s surface temperature over time	-0.093, 0.728, 1.711
Surface_SeaTemp_Annual_Anomaly	Float variable representing the change in Earth's sea surface temperature over time	-0.519, 0.41, 0.858

DATA ACQUISITION & EXPLANATION

- Acquired individual data in csv format from multiple online sources (excluding Kaggle), no licensing or ethical concerns
- Individual data merged on the year and cleaned to form final dataset, with unnecessary variables being dropped
- Final dataset in csv format; contains 8 columns and 51 rows

ANALYSIS PLAN & JUSTIFICATION

1. Data Collection and Preprocessing
    a. Hurricane data gathered from different sources (in references) and merged to create the dataset

2. EDA and Data Visualization
    a. Discovered initial trends and how the climate change variables relate to the “Total_Hurricane_Days” variable

3. Modeling Approach
    a. Conduct Granger’s Causality Test with null hypothesis coefficients are zero in order to confirm relationships
    b. Cointegration test on linear combinations of the group’s series to validate joint stationarity and to meet VAR model assumption
    c. Transform data if needed to meet stationarity requirements
    d. Train models and compare using AIC statistics to select optimal order
    e. Calculate residual correlation and accuracy metrics

TRICKY ANALYSIS DECISION

- Data gathered from different sources → different units of time measurements

- The group had to make the decision on how to organize the different time periods, as some of the data was measured monthly or even daily

  - Ultimately, it was decided to use years because the main variable in the investigation, Total_Hurricane_Days, was grouped in yearly increments

    - This was chosen because the group did not want to use any averages for the main variable if the group chose to do a smaller unit of measurement, which could result in unfavorable skewing

    - As a result, could simply sum up the other variables that weren't in year increments instead of trying break down the main variable into smaller parts
      - This maintained efficiency and accuracy in the data

BIAS AND UNCERTAINTY VALIDATION

- There could be potential bias coming from the fact that the group only had the ability to gather this type of data recently relative to Earth's atmospheric history, and there is not a lot of historical data to base the study off of

- Weather station readings may not have been accurate and hurricanes could have easily been missed in the earlier time periods as well due to technological limitations

- There is also the uncertainty of whether two series are merely correlated or if they can actually be used to predict one another, Granger's causality test is used to account for this issue

RELATIONSHIP ANALYSIS RESULTS

	Total_Hurricane_Days_x	CO2_emissions_x	Population_x	CSIRO_Adjusted_SeaLevel_Inch_x	Surface_Temperature_Change_x	SurfaceSeaTemp_Annual_Anomaly_x
Total_Hurricane_Days_y	1.0000	0.0244	0.0	0.0000	0.0246	0.0000
CO2_emissions_y	0.0058	1.0000	0.0	0.0003	0.0000	0.0000
Population_y	0.0000	0.0000	1.0	0.0000	0.0000	0.0000
CSIRO_Adjusted_SeaLevel_Inch_y	0.0003	0.0011	0.0	1.0000	0.0008	0.0173
Surface_Temperature_Change_y	0.0404	0.0000	0.0	0.0000	1.0000	0.0000
SurfaceSeaTemp_Annual_Anomaly_y	0.0057	0.0000	0.0	0.0000	0.0000	1.0000

- Granger's Causality Test, there is evidence that the coefficients are not 0 in the regression equation of x on total hurricane days
- Cointegration test supports that there are statistically significant relationships

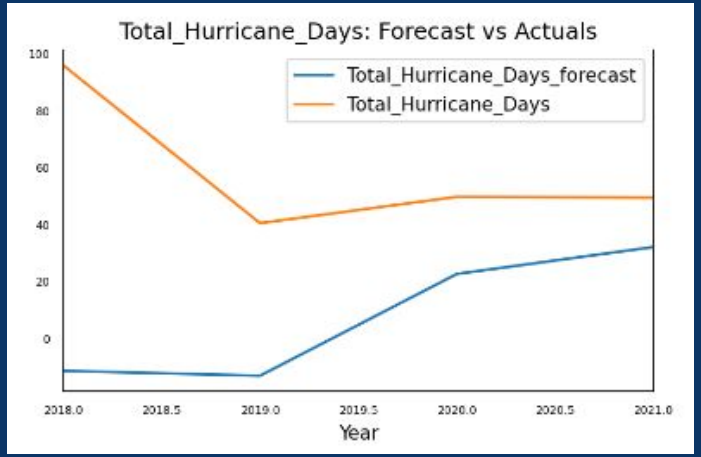
```
Name   ::  Test Stat > C(95%)    =>   Signif
 ----------------------------------------
Total_Hurricane_Days ::  336.89    > 83.9383   =>   True
CO2_emissions ::  136.04    > 60.0627   =>   True
Population ::  83.27    > 40.1749   =>   True
CSIRO_Adjusted_SeaLevel_Inch ::  49.49    > 24.2761   =>   True
Surface_Temperature_Change ::  26.17    > 12.3212   =>   True
SurfaceSeaTemp_Annual_Anomaly ::  10.57    > 4.1296    =>   True
```

VAR MODEL RESULTS

Taking it one step further, since there is a relationship between these variables, can they be used to predict total hurricane days?

- Had to difference data to meet stationarity assumptions
- After fitting increasing orders of the VAR model, the AIC start to increase at lag 10 and slowly decrease again
- An order 9 model is then selected based on this

```
Lag Order = 1
AIC :  64.34527874140083
Lag Order = 2
AIC :  63.97265671082543
Lag Order = 3
AIC :  63.832699742286685
Lag Order = 4
AIC :  61.852795034802014
Lag Order = 5
AIC :  60.124752870630104
Lag Order = 6
AIC :  -17.940473885313835
Lag Order = 7
AIC :  -99.09009990883786
Lag Order = 8
AIC :  -107.17766391518896
Lag Order = 9
AIC :  -108.2040859780341
Lag Order = 10
AIC :  -93.43037404830098
Lag Order = 11
AIC :  -102.12975271303502
```




```
Forecast Accuracy of: Total_Hurricane_Days
rmse   :  62.1703
```

- While the pattern is roughly captured, the residuals are very large, so the model is not very good; could be due to lack of data

CONCLUSION

The group successfully validated the relationship between the explanatory variables and the variable of interest; it can be confidently said that there is a relationship between CO2 emissions, Sea Level, Population, Surface Temperature, and Surface Sea Temperature on Total Hurricane Days

HOWEVER

The model could not utilize these relationships to predict total hurricane days accurately

POTENTIAL ISSUES

A source of error could potentially be from the variable selection. In this study, the variables picked were variables that the group expected had some sort of direct or indirect effect on hurricane activity, but more research could be conducted on the validity of variables for further analysis

There is also a lack of data points; the data had only about 50 years of data, giving the group only 50 data points to work with. Since it was differenced and split into training and testing sets, there is even less data for the model to go off of

NEXT STEPS

New Lines of Exploration

Expansion into analyzing other natural disasters (wildfires, heatwaves, droughts, etc.) due to climate change

Improvements

Include analysis on all oceans in addition to the Atlantic and Pacific

Select variables that have more scientific reasoning to be included

Gather more yearly data to expand timeframe of the analysis (cover more than 50 years)

New Questions

What will hurricane activity look like in future years due to climate change?

What are other potential factors (besides weather patterns) that may cause an increase in total hurricane days or change in activity?

REFERENCES

[1] Machine Learning Plus, “Vector Autoregression (VAR) - Comprehensive Guide withExamples in Python,” Machine Learning Plus, 2019.https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/

[2] Environmental Defense Fund, “How climate change makes hurricanes more destructive,” Environmental Defense Fund, 2023. https://www.edf.org/climate/how-climate-change-makes-hurricanes-more-destructive

[3] US EPA, “Climate Change Indicators: Sea Surface Temperature | US EPA,” US EPA, Dec. 18, 2016. https://www.epa.gov/climate-indicators/climate-change-indicators-sea-surface-temperature

[4] International Monetary Fund, “Climate Change Data,” climatedata.imf.org, 2023. https://climatedata.imf.org/pages/climatechange-data

[5] “Population,” Our World in Data. https://ourworldindata.org/grapher/population?tab=table&showSelectionOnlyInTable=1&country=~OWID_WRL

[6] H. Ritchie and M. Roser, “CO2 emissions,” Our World in Data, Jan. 2024. https://ourworldindata.org/co2-emissions

[7] P. W. Team, “Download Climate Timeseries: Hurricane_DAYS: NOAA Physical Sciences Laboratory,” Noaa.gov, 2022. https://psl.noaa.gov/data/timeseries/month/HURRICANE_ATL_DAYS/
(accessed Oct. 17, 2024).

[8] O. US EPA, “Climate Change Indicators in the United States,” www.epa.gov, Nov. 06, 2015. http://www.epa.gov/climate-indicators

GitHub Repository: https://github.com/jillianhaig/Project2_DS4002/tree/main

QUESTIONS?