9-25-2019

Machine Learning-Based Models for Assessing Impacts Before, During and After Hurricane Events

Julie L. Harvey
Coastal Carolina University

MACHINE LEARNING-BASED MODELS FOR ASSESSING IMPACTS BEFORE, DURING AND AFTER HURRICANE EVENTS

By

Julie Harvey

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Systems Technology with a Concentration in Information Security and Data Analytics in the College of Science

2019

Dr. Sathish Kumar
Major Professor

Dr. Shaowu Bao
Committee Member

Dr. William M. Jones
Committee Member

Dr. Jean French
Chair, Department of CS

Dr. Michael Roberts
Dean, College of Science

Dedication

I would like to thank my family for all of their support as I completed my Master's Thesis.

Acknowledgements

I would like to acknowledge all of my professors in the Computer Science Departments at Coastal Carolina University. Beginning the program as a career-change student, I had limited knowledge of computer science before starting the graduate program at Coastal Carolina University. The professors all provided a breadth of knowledge in many areas of the computer science field. Every professor gave me knowledge that I can take with me to my future career within the field of computer science. Dr. Sathish Kumar, my academic advisor and professor, has guided me through the M.S. program, as well as through the process of this thesis. His abundant support and knowledge has positively impacted my learning throughout the program. I would also like to thank Dr. Shaowu Bao and Dr. William Jones for helping to edit my thesis. I would also like to thank Dr. Shaowu Bao and Dr. Zhenlong Li for providing datasets with respect to the research. I would also like to thank all of the professors that I have had in past at other universities. They have all shaped my knowledge and have given me a framework for life-long learning. Previous undergraduate and graduate degree programs that I have completed, combined with the graduate program at Coastal Carolina University, have helped me to develop my ability for implementing research, analyzing, recording, and reporting data scientifically.

Abstract

Social media provides an abundant amount of real-time information that can be used before, during, and after extreme weather events. Government officials, emergency managers, and other decision makers can use social media data for decision-making, preparation, and assistance. Machine learning-based models can be used to analyze data collected from social media. Social media data and cloud cover temperature as physical sensor data was analyzed in this study using machine learning techniques. Data was collected from Twitter regarding Hurricane Florence from September 11, 2018 through September 20, 2018 and Hurricane Michael from October 1, 2018 through October 18, 2018. Natural language processing models were developed to demonstrate sentiment among the data. Forecasting models for future events were developed for better emergency management during extreme weather events. Relationships among data were explored using social media data and physical sensor data to analyze extreme weather events as these events become more prevalent in our lives. In this study, social media sentiment analysis was performed that can be used by emergency managers, government officials, and decision makers. Different machine learning algorithms and natural language processing techniques were used to examine sentiment classification. The approach is multi-modal, which will help stakeholders develop a more comprehensive understanding of the social impacts of a storm and how to help prepare for future storms. Of all the classification algorithms used in this study to analyze sentiment, the naive Bayes classifier displayed the highest accuracy for this data. The results demonstrate that machine learning and natural language processing techniques, using Twitter data, are a practical method for sentiment analysis. The data can be used for correlation analysis between social sentiment and

physical data and can be used by decision makers for better emergency management decisions.

List of Figures

1.0 Introduction

Several extreme weather events have had a direct impact on the southeastern United States. Four recent severe hurricanes, including Hurricane Matthew in 2016, Hurricane Irma in 2017, Hurricanes Michael and Florence in 2018, have impacted the southeast. In particular, the two major hurricanes in 2018 brought wind and water damage and devastated the southeastern United States, both of which were recognized by the National Oceanic and Atmospheric Administration as "Billion-Dollar Weather and Climate Disasters" (NOAA, 2018). Impacts were felt physically and emotionally. Physical impacts have been well documented by private and public agencies, but sentiment analysis of people impacted by severe weather storms is not widely prevalent. There are a limited number of machine learning models that evaluate the impact of storms on sentiment and that can be used to make predictions for future events.

Sentiment analysis of people impacted by severe weather events has been conducted by phone or mail interviews in the past. This process was tedious and time-consuming for those conducting the surveys, as well as those trying to recover from a devastating storm. With the emergence of social media in the late 1990s, sentiment analysis can be conducted without having to contact people individually. People are connected to each other around the world and information is relayed much faster. The collection of this information has become more efficient as well. There are massive amounts of data collected from multiple social media sites in real-time. This data can be used to examine impacts of severe weather events. The research becomes more efficient with the increased usage of online resources (Bik and Goldstein, 2013). Millions of people use Twitter for social media, which provides real-time and historical data to researchers.

Automated data analysis can give more insight for the government officials and decision makers that can be used to assist those impacted by a natural disaster. Historical data can then be used in forecasting sentiment and physical impacts of future events. Social media has incorporated geotags into their platform that gives the users the option to attach geographical location when making a post. These geotags can be used by researchers to analyze where users are and what users are talking about in relation to the days and times of severe weather events.

Social media responses before, during and after the extreme weather can be used to help government officials, emergency management teams, and decision-makers plan for and respond to extreme weather events. Much research has been conducted analyzing the physical impacts of storms, as explained in Section 4 of this thesis, but the research is limited when focusing on social impacts. Using machine learning for extreme weather events is time consuming, as described in some of the research examined in Section 4, and varies with the data from state to state and from event to event. When using social media, for example, there can be an abundance of data. Data needs to be cleaned to be specific for the problem being addressed by the machine learning. Cleaning and processing the data can be time-consuming depending on the amount of data being used.

The purpose of this study is to examine Twitter text before, during, and after the extreme weather events of Hurricane Florence and Hurricane Michael. In addition, from a data science perspective, an effective model for identifying factors affecting social and physical impacts was studied that will be effective for decision-makers. Different classifiers were evaluated to determine the best classifier for the selected features. The goal of all those involved before, during, and after a storm is to help in preparation and response to future storms. Identifying the features that have the greatest influence socially will help decision-makers to focus on features that will bring the greatest benefits to being prepared for a storm and recovering after a storm. This study demonstrates analysis of models used to identify features that have the greatest social and physical impacts based on social media and cloud cover temperature datasets to support the conclusion of the study.

Analyzing the emotions or sentiments of social network users within the impact area of an extreme weather event can help forecast emotions or sentiments of future events. When emotions and sentiments can be forecasted for future events decisions can be made sooner to prepare for the effects of storms and to identify where assistance will most be needed after the event. This allows for emergency responses to get aid to those in need much sooner. Sentiments are analyzed using text-based posts to determine sentiments as positive, neutral, or negative. Analysis also assigns a sentiment score to each post. These measurements can then be used to forecast sentiment of future storms and the location that will have the most positive or negative sentiment. When the location of the most negative sentiment is known, decision-makers can use this information to provide assistance faster to the areas in need. This was explained in a study done by Enenkel et. al (2018). Public opinion and trends in social media posts can provide an abundant amount of information to decision-makers and emergency response teams to prepare prior to a storm and assist during and after a storm.

Sentiment analysis was used in this study to evaluate Twitter data from the southeast region of the United States and how the region reacted on different days before, during, and after the two major hurricanes in 2018: Hurricane Florence and Hurricane

Michael. The data was also analyzed to find the topics within the tweets that were most predominant during the period of the data and the topics that received the most attention on Twitter. The data was also examined for correlations between actual cloud cover temperature data and sentiment data. The results from these analyses could help identify when, where, and how much assistance is needed through the duration of severe weather events. Much of the current work in the literature has focused on sentiment analysis or physical analysis. This study is innovative as it provides sentiment analysis and then correlates that data with physical impacts of extreme weather events.

The contributions for this study include: (a) Sentiment analysis was performed to measure the emotions of Twitter users during the extreme weather event. The information was used to determine when sentiment was positive and when it was negative during the storms and if there was a relationship among the two storms of when tweets tended to be positive and when they tended to be negative. Authorities can then use this sentiment analysis for improving prevention before another storm and recovery following future storms. (b) Sentiment trends were analyzed over the life of the storms based on the Twitter data collected during Hurricanes Florence and Michael. (c) This study identified frequent keywords and topics of tweets used during the duration of the storms. This was done using wordclouds and list of most frequent words within text in R. (d) Analyzing correlations between sentiment and physical data through the duration of a storm allowing officials to use social media data to gauge physical impacts of storms and where the most aid is needed.

This thesis is organized as follows: Section two describes the background information of the storms used in the study and the models used for analysis. Section three states the research objective. The fourth section provides a literature review of similar studies. The fifth section explains the rationale for the research, the challenges of the study and future research recommendations. The sixth section demonstrates the methodology used in the current study. The seventh section explains the results and discusses the analysis of the results. Finally, the research conclusion is presented in the eighth section of the thesis.

2.0 Background

2.1 Background Information about Hurricane Florence and Hurricane Michael

Hurricane Florence formed on August 30, 2018 off the west coast of Africa. It started as a strong tropical wave and organized steadily to form a tropical depression and then as a tropical storm on September 1, 2018. The strength of Hurricane Florence fluctuated as it moved eastward over the Atlantic Ocean. On September 4th and 5th the hurricane started to form and rapidly intensified to a Category 4 major hurricane. The maximum sustained winds of hurricane Florence were 130 mph. By September 7th Florence was downgraded to a tropical storm due to strong wind shear tearing the storm apart. Later, on September 7th, the currents forced the storm to turn westward where it regained strength. Hurricane Florence then became a major threat to the United States coastline. States of emergency were declared in Maryland, Washington, D.C., Virginia, North Carolina, and South Carolina. Mandatory evacuation was issued for select coastal communities in Virginia, North Carolina, and South Carolina on September 10th and 11th. Hurricane Florence regained Category 4 major hurricane status by late afternoon on September 10th with winds peaking at 140 mph. Florence lost some strength when it went through an eyewall replacement cycle, but quickly regained strength on September 11th. Over the next few days wind shear increased and caused tapering of the storm. Hurricane Florence was downgraded to a Category 1 by September 13th. The storm was headed for the Carolina coast and began to stall as it neared land. Landfall was made by Florence on September 14th near Wrightsville Beach, NC. As the hurricane slowly moved inland it lost strength and was named a post-tropical cyclone on September 17th when it was over

West Virginia and merged with a frontal storm on September 19th. Figure 1 shows the path of Hurricane Florence and the area that was impacted by the storm.

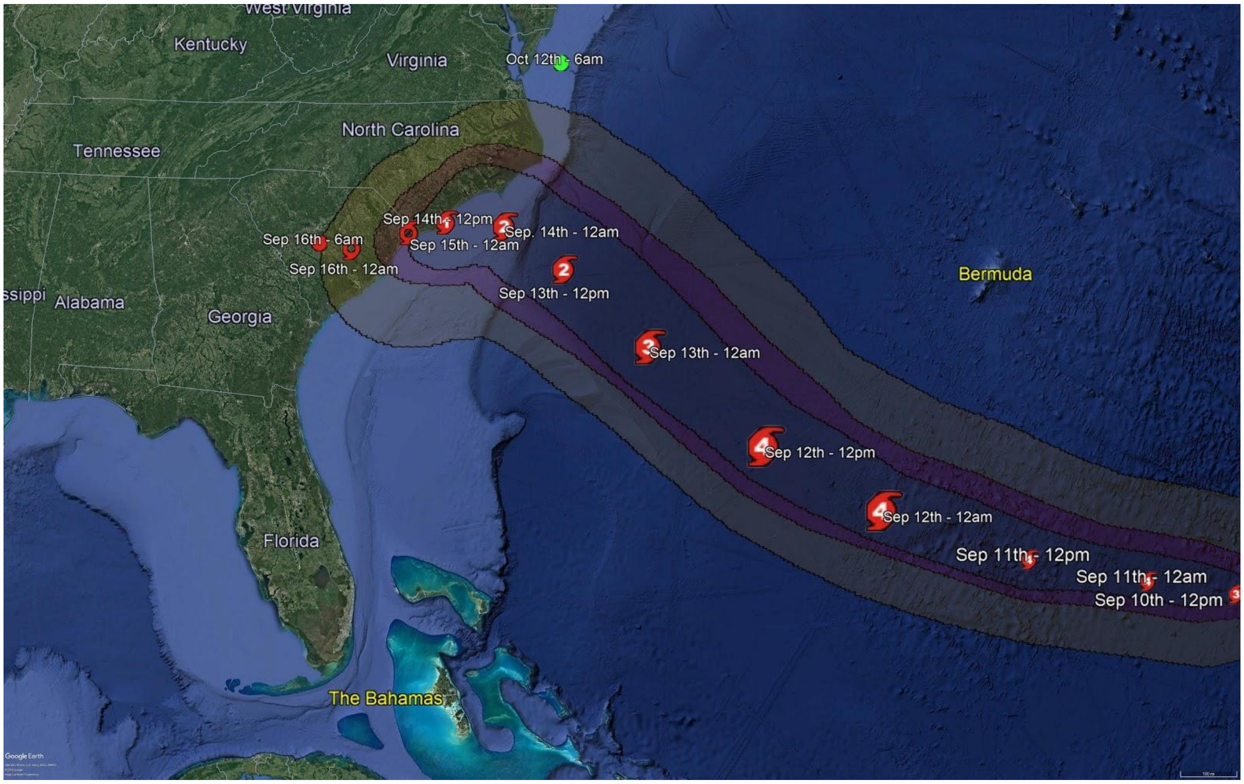


Figure 1. Hurricane Florence Track with Impact Area (National Hurricane Center, 2019).

The Carolinas experienced damaging wind speeds as the Category 1 storm ripped through the states. Florence was stalled over the land for several days due to a high pressure over the eastern United States and it slowly moved west to southwest. Heavy rains and storm surge were experienced over the coastal areas of North and South Carolina beginning on September 13th and lasting through September 15th. Inland regions experienced heavy rain from September 15th through September 17th as the storm slowly turned northeast while losing strength. Widespread flooding across Virginia, North Carolina and South Carolina, on the coast and inland, resulted from the heavy rain. Rainfall was record-breaking in many locations in North and South Carolina with more than 30 inches of rain in some places. In North and South Carolina, more than 500,000 people lost power and at least 51 people died (Assessing the U.S. Climate in 2018, 2018).

Hurricane Michael was also a devastating hurricane for the United States that came on the heels of Hurricane Florence that hit the United States just a month prior. It was the strongest storm to ever hit the panhandle of Florida, and the fourth strongest hurricane, with landfall pressure of 919 mbar, to make landfall in the United States (HURDAT, 2019). Hurricane Michael caused devastation across the Florida Panhandle, Georgia, South Carolina, North Carolina, and Virginia. Hurricane Michael formed from a low-pressure system on October 2, 2018 in the southwestern Caribbean Sea. On October 7th, after slow development, it was deemed a tropical depression. Near Cuba, Michael intensified to a hurricane on October 8th and continued to move north. The Gulf of Mexico provided the optimum conditions for Michael to strengthen rapidly to major hurricane status by October 9th. It was just shy of a Category 5 storm as it approached the Florida Panhandle. Maximum sustained winds of the Category 4 Hurricane Michael were 155 mph as it approached land near Mexico Beach, Florida on October 10th. Tyndall Air Force Base was in the direct line of landfall of Hurricane Michael and measured maximum wind gust of 139 mph as it went over the base. Sustained winds of 86 mph were recorded just prior to the inner eyewall going over the base, when the station failed. Michael began to weaken as it moved inland with a northeastward trajectory. The storm entered Georgia as a Category 3 hurricane with peak winds at 115 mph in southern Georgia. When the storm was over Georgia it was downgraded to a tropical storm. The storm passed through South Carolina, North Carolina, and Virginia and Maryland as it made its way to the Atlantic

Ocean. Michael was then downgraded to an extratropical cyclone on October 12th when it went off the Mid-Atlantic coast. It gained power again when it returned to the Atlantic Ocean but eventually dissipated by October 16th. At least 45 people in the United States died as a result of the storm (Assessing the U.S. Climate in 2018, 2018). Catastrophic damage was experienced along the Florida Panhandle due to storm surge and extreme winds.

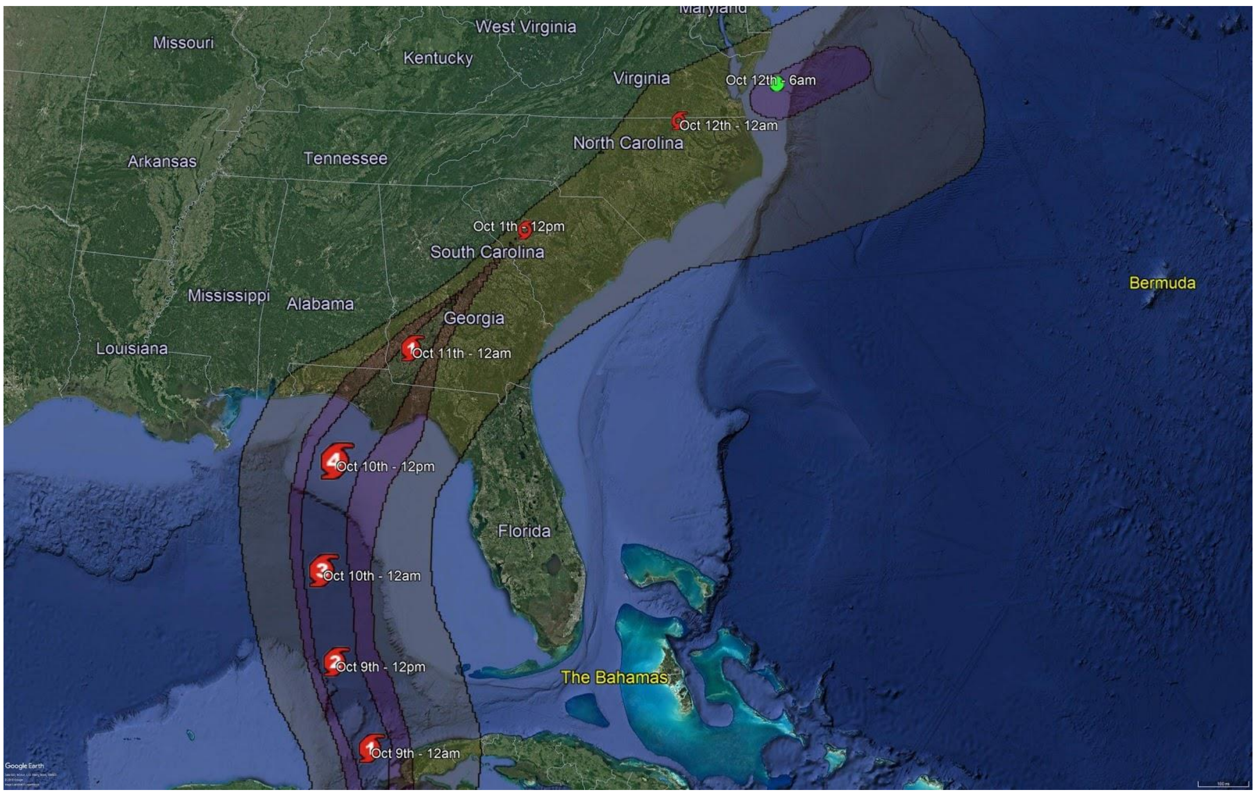


Figure 2. Hurricane Michael Track with Impact Area (National Hurricane Center, 2019).

Hurricane Florence and Hurricane Michael were chosen for this study because of the currency of the storms and the impact they had on the southeastern United States. Both of these storms devastated the Southeast in a little over a month. The attention these two storms received on the Internet was comparable. Hurricane Michael, according to Google Trends, demonstrated slightly more activity than Hurricane Florence from September 2018 through October 2018, but Hurricane Florence had a longer time span of interest in Google searches (2018). Hurricane Michael hit the United States as a stronger storm than Hurricane Florence, but was a faster moving storm. Hurricane Florence hovered over the southeast for longer than Hurricane Michael.

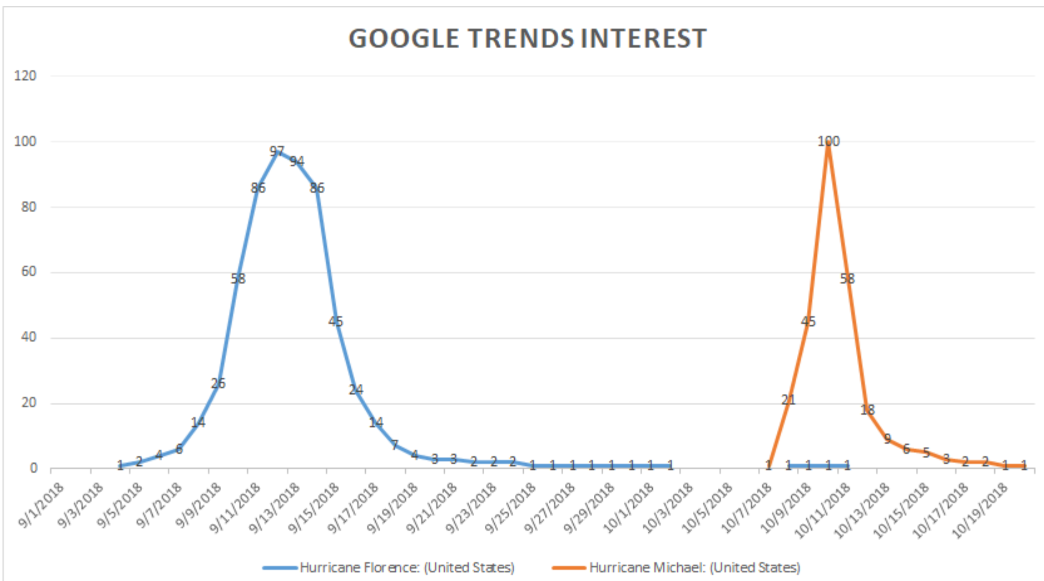


Figure 3. Google Trends Search Interest results for “Hurricane Florence” and “Hurricane Michael” relative to the highest point on the graph (Google Trends, 2018).

Interest in Hurricane Florence on Google concentrated in the southeastern United States with the highest interest being in Virginia, North Carolina, South Carolina, Maryland, West Virginia, Delaware, and the District of Columbia. The following figure shows the concentration of interest on Google across the United States. The darker the shade of blue of the state, the higher the interest was on Google between September 1, 2018 and October 20, 2018.

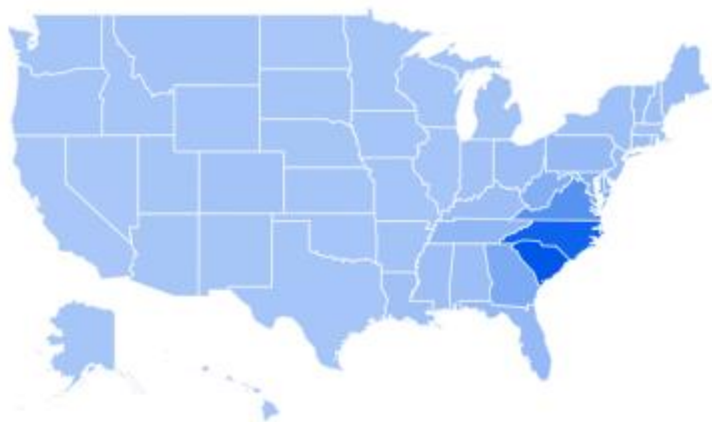


Figure 4. Google Trends Search Interest by State - Hurricane Florence relative to the highest point on the graph (Google Trends, 2018).

Interest in Hurricane Michael from September 1, 2018 to October 20, 2018 on Google concentrated in the southeastern United States as well, but the states with the highest concentration of interest differed from Hurricane Florence. The states with the highest interest included Alabama, Florida, Mississippi, and Louisiana. The following map shows the concentration on interest on Google across the United States. The darker the shade of red of the state, the higher the concentration on interest on Google.

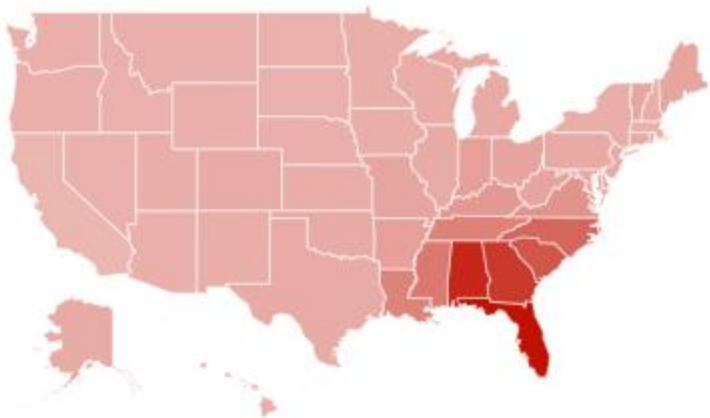


Figure 5. Google Trends Search Interest by State - Hurricane Michael relative to the highest point on the graph (Google Trends, 2018).

3.0 Research Objectives

The objective of this research is to develop models for analyzing social and physical data related to extreme weather events. The model can be used to make predictions of future impacts due to future storms. Sentiment analysis of each hurricane will help to establish a pattern of sentiment over the course of an extreme weather event. Comparison analysis will help to identify correlations between sentiment and physical impacts. The research outcomes from this project will provide the stakeholders with a model that can be operated using past and future social media and physical sensor datasets. The outcomes from the model can be used to formulate solutions for improving preparation and responses of future extreme weather events.

4.0 Literature Review

4.1 Text and Sentiment Analysis of Social Media

Much of the literature reviewed for this study involved text mining and sentiment analysis with natural language processing (NLP). In a study by Soni and Mathai (2016), tweets were clustered by k-means and classification trees were used to analyze the clusters. The data became domain-specific following the clustering and the classification was shown to be more accurate than without the clustering. The k-means Classification and Regression Tree (CART) accuracy was 74.85%, but SVM, CART, and Random Forest accuracies were only lower than the clustered analysis by a couple percentages.

Diakopoulos and Shamma (2010) examined the 2008 United States presidential debate Twitter message sentiment. They analyzed the sentiment reaction of Twitter messages to the debate video. The sentiments of the tweets were analyzed according to topics. The study presented that events that are interesting can be detected using anomalies in the pulse of the sentiment signal. The results of this study depend on the event being polarized in structure. The use of polarized events like the debate can give ideas about sentiment of different aspects of the debate, but there are times when tweet sentiment annotations are not distinctive, or the entities within the tweets are not distinctive. This can lead to misclassification of tweets. Saif, et. al (2013) annotated tweets and entities individually to better classify sentiment. They evaluated the sentiment of the tweet itself, as well as the sentiment of the entities within the tweets.

Classification of tweets was conducted by Tewari, et.al (2017), who compared SVM, NLP, naive Bayes (NB), and k nearest neighbor (KNN) classification techniques to twitter data. They analyzed complexity, amount of memory required, whether independent or dependent feature work better, decision boundary, speed of prediction, and speed of training of spam versus non-spam tweets. NB was shown to be the simplest that required the least amount of memory and performs better with independent features. NB's boundaries are linear/parabolic/elliptic, and both its prediction and training speeds were shown to be fast. SVM was complex and memory intensive, and performed better with dependent features, worked with any boundaries, and both speeds were moderate. KNN was shown to be moderately complex, memory intensive, performed well with both independent and dependent features, worked for any decision boundaries, but its prediction and training speeds were slow.

Social media has been used to explore correlations between weather and human mood. Li, et. al (2014) evaluated mood on Twitter as it related to meteorological data from NOAA. They looked at relationships between four different mood dimensions and average temperature, temperature change, types of precipitation, snow depth, wind speed, solar energy, and weekday effect. The mood dimensions they used were hostility-anger, depression-dejection, fatigue-inertia, and sleepiness-freshness. They found that mood was not sensitive to average temperature, but it was to the temperature change. There was a negative correlation between precipitation and mood, as well as between snow depth and mood. They did not find correlation between wind speed and mood.

4.2 Extreme Weather Events Social Media Analysis

4.2.1 Tweeting Concentrations

Shelton et. al (2014) used Twitter data to examine sociospatial networks during Hurricane Sandy. The largest concentration of tweets was shown to be in the areas that were hit hardest by the hurricane. Researchers used a small subset of big data from Twitter for the social and spatial analysis. The mixed approach to the big data was important in the study. Tweet density was quantitatively mapped and the actual tweets with the intended context were qualitatively measured. Jessop et al. (2008)'s territories, places, scales, and networks conceptual framework was used to identify complexities of the content within the tweets and their sociospatial relations.

The Australian floods of 2010-2011 were used to analyze social media and its relation to the extreme weather events by Cheong and Cheong (2011). Tweets were collected after the impact of the Queensland floods, whereas data was collected for the New South Wales and Victorian floods before and during impact. Social network analysis was used to identify interactions among Twitter users. General information was found to be tweeted after the Queensland floods and Queensland was found to be the most active community. During the New South Wales floods, Twitter activity was minimal. Activity during the New South Wales and Victorian floods was shown to be by volunteers who had been active during the Queensland floods.

Supertyphoon Haiyan and Twitter activity was examined by David et. al (2016). They found that tweets mainly focused on damage and disaster relief. There was a high level of activity of retweets in the early days of the event. Original tweets from ordinary users were more likely to be emotional, showing support, and politically charged. The findings during the event included a majority of the posts being a retweet of information content with approximately 80% of Twitter traffic being retweets of news messages. Data on the day when the storm made landfall and five days after showed that retweets were the largest percentage of Twitter activity. Twitter activity fell rapidly after the fifth day, which is typical of Twitter news issue cycles. David et. al (2016) found similar findings to that of Cheong and Cheong in that tweets from ordinary people were more likely to be emotional, about relief efforts, and more personal information. Over time of the dates of analysis, Twitter activities began as mainly information about the typhoon, and then moved to more disaster relief messages, reactions, emotions, and stories in the aftermath.

Twitter activity during Hurricane Irma was analyzed by Gadidov and Le (2018). Researchers analyzed the reaction of people in affected areas of the storm before, during, and after the storm through topic modeling. Trends were suggested to be helpful in relief efforts of future extreme weather events. Activity of tweets pertaining to Hurricane Irma was shown to peak as the hurricane made landfall. Four topics were identified; two included general discussion, one mentioned power outages and the fourth contained hopes and prayers. The results of this study were suggested to be used in the future to create a baseline trend of reactions that are to be expected before, during, and after a storm.

4.2.2 Evacuation

Martin, et. al (2017) used big data to analyze near real-time measurements of evacuation order compliances. Spatiotemporal variability in social media response was examined using Twitter. Tweets were used to assess resident evacuation responses. The study showed that prior to Hurricane Matthew there was a peak in Twitter responses. Once the storm passed, responses dropped quickly. Geotagged tweets showed that residents evacuated the coast, with timing of the evacuation dependent upon the state from which they were evacuating. When the state of South Carolina was analyzed, there was overall compliance with the evacuation orders. The study also analyzed residents evacuation times and destinations. Stowe et. al. (2018) collected data from Twitter API and used Density Based Spatial Clustering (DBScan) for clustering the tweets according to the users' coordinates. Tweets were also examined using temporal clustering up to the time of the storm with weekday evenings showing the most activity on Twitter. Researchers used annotation to determine whether they complied with evacuation orders, sheltered in place, or it was undetermined. Classification was used to predict what the Twitter users' actions might be during extreme weather events. Tweet semantics were represented by word embedding and then combined with the temporal and spatial features. Adding up to 20 classifiers improved the performance of the model, but more than 20 decreased performance. The results of the study found that linguistic and geospatial features can be used to predict evacuation behaviors using Twitter.

4.2.3. Damage

Storm damage has been examined using Twitter data before, during and after Hurricane Sandy in 2012. Kryvasheyeu, et. al used Twitter to identify a correlation between Twitter activity and actual damage caused by the storm (2012). Using specific keywords within tweets, they found that activity on Twitter increased with proximity to the storm. FEMA assistance grants and insurance claims that were associated with Hurricane Sandy were used to determine if the activity on Twitter was an actual predictor of damage. They found a strong correlation between economic damage and Twitter activity. They also found a correlation between sentiment on Twitter and damage from the hurricane. They proposed that big data from social media can be used by officials to rapidly assess damage caused by extreme weather events.

Enenkel et. al (2018) analyzed damage during Hurricane Harvey and Hurricane Irma. They used the spatial distribution of tweets to approximate damage from both hurricanes, and suggested the map could interpret the preliminary estimation of the distribution of damage. High Twitter activity was shown to correlate with areas of high damage from Harvey and Irma. This correlation was strongest following the disaster. Researchers suggest this could be used to approximate damage and help with disaster preparedness.

Rice University's Kinder Institute for Urban Research developed a platform to augment the Federal Emergency Management Agency's (FEMA) model for identifying damage estimates. These FEMA models missed many areas, in the wake of Hurricane Harvey, that were heavily impacted. The study by Rice University suggests that their model, when used in conjunction with the FEMA model, can help to improve disaster response and recovery. Immediate damage estimates by FEMA can miss approximately 46% of damage estimates. Social media and emergency crowdsourced sites were shown, in this study, to enhance the FEMA model and provide more accurate information about damage estimates.

4.2.4. Emotion

Gruebner, et. al (2018) used Twitter data to analyze emotions before, during, and after Hurricane Sandy of 2012. The tweets were taken from the New York City area only. Negative emotions were shown to be more prevalent after the storm than during the storm. The concentration of the negative emotions varied among neighborhoods across New York City, with the highest concentration being in Staten Island. Other factors were suggested to contribute to the differences in concentration among the boroughs, including socio-ecological factors. Three of the boroughs showed significant association of negative emotions when comparing emotions before and after the hurricane.

Twitter posts during Hurricane Sandy were used for sentiment classification and then plotted on a geographical map (Caragea, et. al, 2014). Almost 13 million tweets were collected between Oct 26, 2012 and November 12, 2012. Naive Bayes and SVM classifiers were used for the data. Using various feature types, performance of the classifiers was between 67% and 76%. Tweets during this time period were plotted on maps to visually examine the arrangement of tweets. Clustering tendency of tweets was statistically measured based on the proximity to Hurricane Sandy's landfall. Researchers found that proximity to landfall correlated with increased tweeting. Maximum tweets were shown to occur during maximum impact of the storm and then quickly spread. Sentiments of Twitter users was found to correlate with the location of the user and their proximity to the storm. Negative sentiments were shown to cluster closer to the proximity of Hurricane Sandy. Sentiment expression was significant with regard to social and spatial environment of the storm. Researchers suggest that these real-time maps of physical disaster combined with anomalies in emotional activity with proximity to any storm could assist is response and recovery.

Disaster situation awareness was examined for developing a credibility framework using Twitter data. The approaches intent is to be used to identify trustworthy events from big data of social media during extreme weather events. "...crowdsourcing, which states that errors propagated in volunteered information decreases as the number of contributors increases" was used for this framework (Yang et. al 2019). Twitter data from Hurricane Harvey was collected. The data was limited to tweets related to situation awareness using specific keywords. Tweets were aggregated by topic and spatiotemporal characteristics. Each tweet was given a credibility score and each event was given an accumulated credibility score. Credibility of the tweets was analyzed against scales of location, time, and social impact. The model provided reliable identification of events with the highest credibility scores. Spatiotemporal characteristics and social impacts were analyzed. Evaluating credibility of information generated by Twitter users was improved by identifying flexible and dynamic clusters of tweets as events.

Retweetability of tweets during Hurricane Harvey was examined by Neppalli et. al (2016). They suggest that their model, when paired with models that identify the trustworthiness of Twitter information, can help promote accurate, reliable information via social media. The researchers analyzed tweets that were retweeted to identify aspects that affect the retweetability of a tweet. The model automatically predict the retweetability of the tweets. Specific features were taken from tweets and information of Twitter users was collected to develop a model that was used to predict retweetability. This classification had better performance than the "bag or words" approach to classification.

Alam et. al used a multidimensional approach of text and images from tweets during three extreme weather events, Hurricane Harvey, Hurricane Irma, and Hurricane Maria (2018). Sentiment analysis was performed on the collected data. Through all of the days that the data was collected, sentiment was predominantly negative. Random Forest was used for classification of humanitarian topics and LDA was used for topic modeling. When analyzing the image data, the total number of image tweets per day was examined for each hurricane. Hurricane Harvey was demonstrated to have the highest daily volume of tweets, on average using image classification models. The model is intended to help with crisis management and emergency responses.

Twitter trends during Hurricane Sandy were examined chronologically and thematically, then compared to psychological theories of trauma recovery (Lam, 2013). The trends on Twitter followed the psychological stages of trauma recovery. Trends of the spread of information and anxiety/precaution increased as the storm approached. As the potential of the destruction of Hurricane Sandy increased, the trends increased. They showed an influx in tweets with relation to Hurricane Sandy. After the event, support of relief efforts on Twitter correlated with recovery and coping mechanisms of the psychological theories. The three stages of anticipation, experience, and recovery overlapped in the study with anticipation and experience being short term and recovery lasting more long term.

Baylis et. al analyzed correlations between sentiment and weather conditions of 511 million tweets (2016). The different weather factors that researchers examined were all associated with more negative sentiment. There was a statistically significant change in expressed sentiment of Twitter posts associated with the weather. Sentiment was shown to be worse when the weather conditions were less than ideal. When weather terms were eliminated from the data, the associations remained the same, but with less significance.

5.0 Rationale for the proposed research

This topic was chosen because of the experience of living in an area that has been in direct impact of several extreme weather events and previous research of data analytics. Four recent severe hurricanes; Michael, Florence, Irma, and Matthew, have impacted my area of residence. Social media responses before, during and after extreme weather can be used to help government officials, emergency management teams, and decision-makers plan for and respond to extreme weather events. Combining this social media data with physical data using machine learning will be useful in making predictions for future extreme weather events by emergency management officials.

5.1 Limitations of the existing work

Using machine learning for social media analysis is relatively new. Previous research, including that conducted by Barnes, et. al, (2008) included data from newspaper articles. The data they collected was limited in the amount of information that came directly from those involved in Hurricane Katrina. Using social media for data collection and sentiment analysis provides more data directly from the population. The study by Diakopoulos and Shamma (2010) included only those tweets that were weather related, a limitation that this study aims to improve upon by including tweets that are unrelated to the weather event. Yang, et. al (2019) identified the credibility of Twitter data during Hurricane Harvey, which was extended by this current research to include the identification of the event on Twitter and then using that data for other applications. Tewari, et. al (2017) classified tweets from Twitter and this research took the analysis a step further by classifying and then using the tweet classification for sentiment analysis. Some of the previous research contained a small number of tweets. We collected a dataset of over

100,000 tweets to address for this limitation in other studies. When smaller datasets were required for analysis due to memory constraints of the computer, a random sampling of the original dataset was used. Other research, including those by Lam (2013) and Nazer, et. al (2017), was also limited in that social media data was not compared with physical data. This research seeks to explore possible relationship between two factors of extreme weather events, although correlations were not found between the social media and this particular physical datasets of this research. Future research could be conducted to identify correlations using different physical datasets, further cleaning of social media datasets, or a combination of these modifications. Identification of possible correlations between social media data and physical data can help emergency management officials in the future.

5.2 Motivation and Research Challenges

Deriving the best model for this problem is challenging because of the limited previous research in creating similar models combining the social and physical data. Challenges also include choosing the model(s) that will be best for identifying features of a dataset that are most indicative of social sentiments. Choosing the most appropriate classifier is useful in making forecasting models of factors affecting the "soft impacts" from extreme weather events. The datasets collected were from Twitter of Hurricane Florence (2018) and Hurricane Michael (2018). The physical data was collected from Coastal Carolina University School of Coastal Environment.

A challenge of using social media data for analysis is that it may not reflect an accurate representation of the population that is under examination. Social media is a popular method of communication, but not everyone uses social media. Some of those most affected by the severe weather event may not have access to social media. There is also selection bias of the available social media data. Retweets could also cause bias for positive or negative sentiment if they are not eliminated during the data cleaning process. Future research could use other social media venues for datasets.

The data cleaning process can present challenges when analyzing social media data. The tweet data from Twitter that is usually of interest for analysis the text data, which requires preprocessing. This preprocessing step is time consuming. The tweets may contain misinformation and rumors that can skew the data analysis. The text often contains numbers and characters due to the limit of the amount of characters and words for each post. These special characters and numbers need to be eliminated from the text, which has the potential to change the meaning of the text thus changing its sentiment value. Misspelling of words, sarcasm, and slang can lead to misrepresentation of the sentiment as well. Unimportant words also need to be eliminated from the word cloud analysis. Future work could implement this data cleaning step.

6.0 Methodology

This section provides an overview of how the data mining, processing of the text, machine learning techniques, and classification techniques were implemented for this research. Figure 8 below is a diagram of the architecture used for sentiment analysis. Figure 9 below is a diagram of the architecture used for correlation analysis of physical and social data.

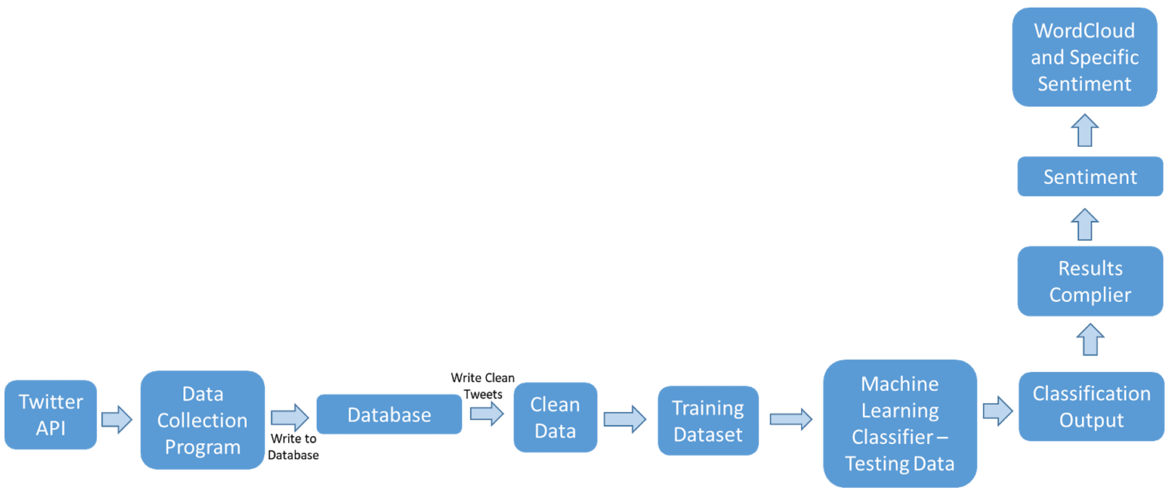


Figure 6. System Architecture for Sentiment Analysis

Figure 7. System Architecture for Correlation Analysis

6.1 Machine Learning Algorithms

Machine learning is a type of artificial intelligence in which models are built to learn from data. There are generally two types of algorithms in machine learning; supervised and unsupervised. There are also variations that incorporate both types of learning. Supervised learning involves a training dataset with expected outcomes, or target. This dataset is used to train the model to output the expected target. New data exposed to the algorithm should be mapped correctly if the training is implemented correctly. Unsupervised machine learning does not require training data and therefore is conducted to discover patterns in data that are unknown. Supervised learning was the focus of this research. After reviewing the literature of machine learning being used for sentiment analysis of hurricane data, the most used and highest performing algorithms are naive Bayes, Random Forests, and Support Vector Machine. Due to the frequency of use of these algorithms, it was decided to further investigate their suitability with these datasets and to test two other algorithms as well. Supervised machine learning is ideal for the analysis of hurricane data due to the prevalence of data for past hurricanes. This data can potentially be used to predict future hurricanes.

6.1.1. Boosting

Boosting is an ensemble learning method for classification that converts weak rules or learners into strong rules or learners. Weak rules can be combined to form a strong rule. Boosting is used to improve the prediction of a model. Each learner that is trained sequentially and corrects its predecessor. Decision trees are usually used at the base learner, with shallow trees representing weak learners. Because improvements are made in small increments, overfitting is avoided by stopping the process as soon as overfitting is detected. If x is to represent features and y is to represent the response, the following formulas can be used for gradient boosting machines (Boehmke, 2018):

1. Decision tree fit to data:

   $F_1(x) = y$

2. Decision tree fit to residuals of previous step:

$h_1(x) = y - F_1(x)$

3. New tree is added to the algorithm:

   $F_2(x) = F_1(x) + h_1(x)$

4. Decision tree fit to residuals of previous step:

   $h_2(x) = y - F_2(x)$

5. New tree is added to the algorithm:

   $F_3(x) = F_2(x) + h_1(x)$

6. Process is continued until overfitting is detected

The following is a general additive model where b is representative of the individual decision trees (Boehmke, 2018):

$$f(x) = \sum_{b=1}^{B} f^b(x)$$

6.1.2 Maximum Entropy

Maximum Entropy classifier is an exponential model that is used for solving text classification problems. Assumptions made by this classifier are minimal and it is used when there is little known about prior distributions of the data. This algorithm uses the theory that the best model of a given dataset is the model that provides the highest entropy of all the datasets that satisfy the known constraints. Neto describes how the maximum entropy theory is applied to machine learning as follows (2015): When the random variable is represented as n and the probability distribution is represented as *p(n)*, the entropy for the data is:

$$H(p) = -\sum_{i} p(i) \log p(i)$$

6.1.3 Support Vector Machine (SVM)

Support vector machine (SVM) is a non-probabilistic binary linear classifier. It is a discriminative classifier that searches from the optimal separation boundary in data that has different classes within a dataset. It can be used to estimate density and to show regression or classification. The support vectors are those points in the data that are closest to the hyperplane. These support vectors are the most difficult to classify and have a direct impact on the best location of the decision surface. SVM is used to find the optimal solution for the dataset. Training data is plotted in a multidimensional space. A hyperplane is then used to separate the classes. If linear separability is not possible, a new dimension is added to further separate the classes. In the following diagram from Anon (2011)., the original map of the objects is shown as input space using kernels (2011). The SVM map of the objects in the "Feature space" image shows linear separation of the objects.

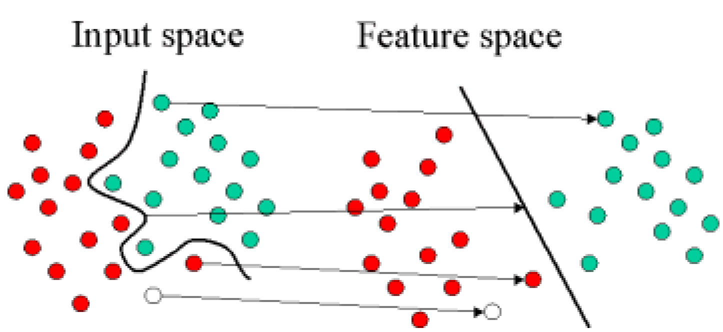


Figure 8. SVM operation (D & Rajkumar, 2016)

Two hyperplanes are plotted that will not have any points between them. The points that fall on either of the hyperplanes are called the supports. An example of finding the optimal hyperplane was demonstrated by Yu, et.al (2013) as shown below:

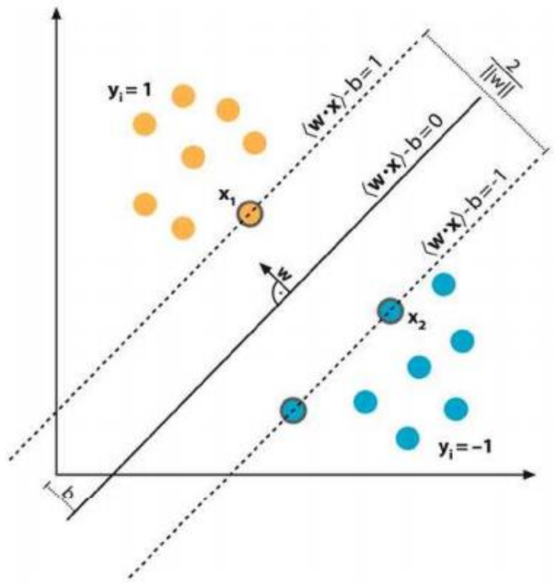


Figure 9. SVM optimal hyperplane (Yu, et. al, 2013)

6.1.4 Naïve Bayes

Naive Bayes is a probabilistic classifier that can be used for text classification. The Maximum A Posteriori decision rule is used by this classifier in a Bayesian setting. This classifier assumes that all variables in the dataset are independent of each other and come from a similar distribution. It also assumes that the features exhibit conditional independence. Naive Bayes is based on the Bayes theorem for conditional probability of events *A* and *B*. The two conditional probabilities are related according to the following formula (Khan, 2017):

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}$$

According to Khan, the independent variables are the predictors and the dependent variable is the class, or outcome. The predictors and the classes that are associated with the predictor are used to train the model and predict class based on feature values. X represents the predictor and n represents the number of predictors. The outcome variable is represented by y and k represents the number of classes. To obtain the probability of the observation coming from any class, the following equation is used (Khan, 2017):

$$P(y = C_k | X_1 = x_1, X_2 = x_2, X_3 = x_3, X_n = x_n)$$

When $B = C_k$ and $A = (x_1, x_2, x_3, \ldots, x_n)$ this can be replaced in the conditional probability formula as follows with all of A assumed to be independent conditioned on B (Khan, 2017):

$$P\ (C_k | x_1, x_2, x_3, \ldots, x_n) = \frac{P(C_k) P(x_1, x_2, x_3, \ldots, x_n | C_k)}{P(x_1, x_2, x_3, \ldots, x_n)}$$

The Bayes formula is repeatedly applied with the numerator of the equation being the joint probability of A and B leading to the following equation involving numerous conditional probabilities (Khan, 2017):

$$P(C_k, x_1, x_2, x_3, \ldots, x_n) = P(C_k) P(x_n | C_k) P(x_{n-1} | x_n, C_k) \ldots P(x_1, x_2, x_3, \ldots, x_n, C_k)$$

To simplify this expression, for a class, the predictors are independent of each other with no correlation between features. When A, B, and C are all independent events and A and B are independently conditioned on event C, the following formula can be used (Khan, 2017):

$$P(A|B,C) = P(A|C)$$

This formula is applied to conclude with the following formula (Khan, 2017):

$$P(C_k|x_1, x_2, x_3, \dots, x_n) = \frac{P(C_k)\, \pi_{j=1}^{n}\, P(x_j|C_k)}{P(x_1, x_2, x_3, \dots, x_n)}$$

The denominator in this expression is a constant for the features. When comparing the probabilities of the different classes, the numerator can be used. All possible values of *k* can be used to evaluate the numerator. The highest value is then chosen.

6.1.5 Random Forest

Random forest is another ensemble learning method for classification. It works by constructing decision trees from a randomly selected subset of the data and corrects overfitting of the training set by the trees. A final class of the test object is decided from the decision trees. Weak estimators can be combined in the random forest classifier to form strong estimators. Averaging these multiple regression trees reduces the variance of the model and improves the performance of the trees on the test dataset and avoids overfitting. Building multiple trees allows for smaller correlation between trees. If p represents the predictors of a dataset and m represents a random selection of predictors from the dataset that are chosen as the split predictors, $m = \sqrt{p}$. The regression trees are constructed and represented as $T_1, \dots, T_B$, where B represents the number of trees and x represents the variable from the tree. The random forest predictor can then be calculated using the following formula (Guillot, 2017):

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x)$$

6.2 Sentiment Analysis

Sentiment analysis can automatically extract emotions and opinions from text data. The real-world applications of sentiment analysis are numerous. Sentiment analysis can incorporate Natural Language Processing, linguistics, and machine learning. One type of sentiment analysis is natural language processing (NLP) technique that analyzes subjective information from text. NLP can take all of the unstructured data from the internet and process it, extracting meaningful content for computer processing. Knowledge-based techniques perform based on a set of rules that are manually implemented, statistical models that rely on machine learning to learn from the dataset or a blend of the two can be used for analyzing sentiment. Polarity of the text can be classified as positive, negative, or neutral. Neutral text can sometimes be ignored due to its proximity to the boundary of positive or negative. Some classifiers, including SVM, work better and produce higher accuracy when neutral classifiers are included (Koppel, 2006). Sentiment analysis of text can go beyond polarity and classify according to specific sentiments. A scaling system can be used to determine sentiment. A number range is assigned to words that are associated with negative, neutral, or positive sentiment. Sentiment can then be adjusted relative to the environment of the word. Natural language processing gives a score to each piece of unstructured text based on its relation to the concept (Augustyniak, 2015). When natural language processing is used, the sentiment values can be adjusted. These adjustments can be made relative to any modifications that are made to the sentiment value. The score can be modified if words change the sentiment. Sentiment structure can be complex and the accuracy of a sentiment analysis system requires an element of human judgement. Diakopoulos and Shamma (2010) found that when people are judging sentiment of Twitter text, the agreement among people was 65.5%. Because of its subjectivity, sentiment analysis is influenced by personal thoughts, beliefs, and experiences. Machine learning can help to reduce errors and improve consistency in the data using a sentiment analysis system. Matthew Jockers's version of Syuzhet in R can be used for sentiment detection (2017). Three sentiment dictionaries are used for the detection.

6.2.1 Natural Language Processing Analysis

Twitter data was analyzed using Natural Language Processing Analysis. Tweets were analyzed for Hurricane Florence and Hurricane Michael to determine the most frequent terms used in tweets during the hurricanes. Wordclouds, lists of the most frequent terms in tweets and Twitter activity over time were all used during analysis. Analyzing the data with wordclouds is more visually appealing than lists, but quantification of the words may be necessary for additional analysis using the wordclouds. Sentiment was analyzed as positive and negative. Sentiment type was then broken down into ten categories according to the syuzhet package in R.

6.3 Correlation Analysis

Correlation analysis can involve a correlation test, a correlation matrix, a correlation visualization, a correlation table, or a combination of these. A correlation test is used to analyze any associations that may be present between two or more variables. The correlation matrix analyzes multiple variables simultaneously. A correlation visualization is a graph that highlights the variables that are most correlated. A correlation table displays the values of correlation between the variables. Pearson is the most commonly used correlation coefficient measuring linear association between variables. The formula for Pearson correlation is the following when using 2 variables, *X* and *Y* (Dalinina, 2017):

$$P_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

The closer the p value is to 1, there is a positive correlation in that as one variable increases the other will also increase. The closer the p value is to -1, there is a negative correlation in that as one variable increases the other will decrease. If the variables are independent of one another, the p value will be close to 0. When a regression line of the two variables is plotted, the slope of the line is equivalent to the correlation between the two variables.

6.4 Method for setting up data

This study focused on social media data and cloud cover temperature relating to Hurricane Florence and Hurricane Michael. Social media data was collected from Twitter, and cloud cover data was collected from Coastal Carolina University's School of Coastal Environment for Hurricane Florence and Hurricane Michael. The research was based on a variety of datasets, areas of study, and time periods. 144,149 tweets were collected from Twitter using the Twitter Stream Application Programming Interface (API) from September 11, 2019 through September 20, 2019 for Hurricane Florence. 108,778 tweets were collected from Twitter using the Twitter Stream API from October 1, 2019 through October 18, 2019 for Hurricane Michael. Both datasets were geo-tagged for comparison purposes to determine the existence of relationships that can be used for prediction in the future. The tweets collected for both hurricanes were cleaned to eliminate non-text characters and unused features were eliminated from the dataset. Emotion classification and score were added to the dataset after being determined. Physical cloud cover data and location data of each hurricane did not need to be cleaned prior to use in this study. The cloud cover temperature included temperature at various latitude and longitude point around the hurricanes.

6.5 Platform

Microsoft Excel was used to partially clean the data. Non-text characters and unused features were eliminated from the dataset. R was used to further clean the data, analyze the data, and plot analyses. Specific features were extracted using R. Each of the tweets for both datasets was subjected to text processing and analysis, sentiment analysis, and prediction classification analysis. Correlation of sentiment and physical data was then analyzed using R.

6.6 Sentiment Preparation and Text Analysis

Social sentiments were mined using natural language processing. The initial step involved text cleaning in Excel whereby special characters were removed and all letters were converted into lowercase letters. Initial sentiment analysis was performed next. The models in this study analyzed the words and phrases for text from Twitter to identify positive, neutral, and negative sentiment. Each tweet was also assigned a numeric sentiment score of 0 to 1 by calculating the polarity of each tweet as sentiment. A value of greater than 0.599999 was deemed “positive”, value of 0.500000 to 0.599999 was given a “neutral” value, and a value less than 0.500000 was considered “negative.”

WordClouds were formed for each cleaned dataset to find the most common words within the dataset. A *wordcloud* provides a visual of the frequency of words. The number of tweets were plotted against the days of the month using a bar graph in R. Emotions for each tweet were then evaluated using the National Research Council Sentiment and Emotion Lexicons (NRC) dictionary. The R library *syuzhet* was then utilized for sentiment analysis. The package evaluates the text from the tweets and returns positive values for eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and two sentiments (positive and negative). The R library *plotly* was used to display a visual of emotions from the NRC sentiments.

Classification models, including naive Bayes, random forest, boosting, maximum entropy, and SVM, were trained and tested with the data from each of the two hurricanes. The datasets were randomized and Bag of Words tokenization was used. The data was cleaned to remove punctuation, numbers, stopwords, and white space. The document term matrix was built using the five most frequent sentiment terms. Only a portion of the data was used for the document term matrix due to memory constraints of the computer being used in this study. Word frequencies were converted to yes (presence of label) and no (absence of label). The final training and testing document term matrices were developed. Each model was then trained and tested for predictions. A table was created to compare predicted values to actual values. A confusion matrix created for each model to identify overall accuracy of predictions using the model.

6.7 Physical Model Preparation and Analysis

Cloud cover temperature maps were created for both hurricanes using R with the maps library. Latitude and longitude remained the same for each segment of time, with this being different for each hurricane. The maps display the progression of each of the storms, according to the cloud cover temperature and latitude and longitude of the datasets. The location of the center of each hurricane for each day during the hurricanes was used as physical data as well. This data was used to determine the distance between the center of each hurricane and proximity to the location of each tweet for each day. Both physical data features were used to identify possible correlations between the physical data and social media data.

6.8 Correlation Analysis of Hurricane Data

Sentiment data and physical data were analyzed together to identify any possible correlations between sentiment scores and cloud cover temperatures as well as between sentiment score and proximity of each tweet to the hurricane center. Covariation was tested and plotted for linearity. Pearson’s correlation was used, along with Kendall and Spearman correlation. Shapiro-Wilk test was performed on the data to identify if the data was normally distributed for each variable. Q-Q plots were then created to visually inspect data normality.

7.0 Results and Discussion

The social and physical effects of Hurricane Florence and Hurricane Michael were analyzed using sentiment analysis, physical analysis, and correlation analysis. Sentiment analysis for each hurricane is discussed first. Word frequency, changing sentiment with time, and prediction models for sentiment analysis are presented. Cloud cover temperature and distance between tweet location and hurricane center is used as physical data to depict the storm over time.

7.1 Sentiment Analysis

7.1.1 Natural Language Processing Analysis

7.1.1.1. Hurricane Florence

Figure 10 shows the number of tweets that were collected for each day before, during, and after Hurricane Florence. The majority of the tweets posted during the time period occurred before the storm and during landfall. The number of tweets declined as the storm passed and dissipated.

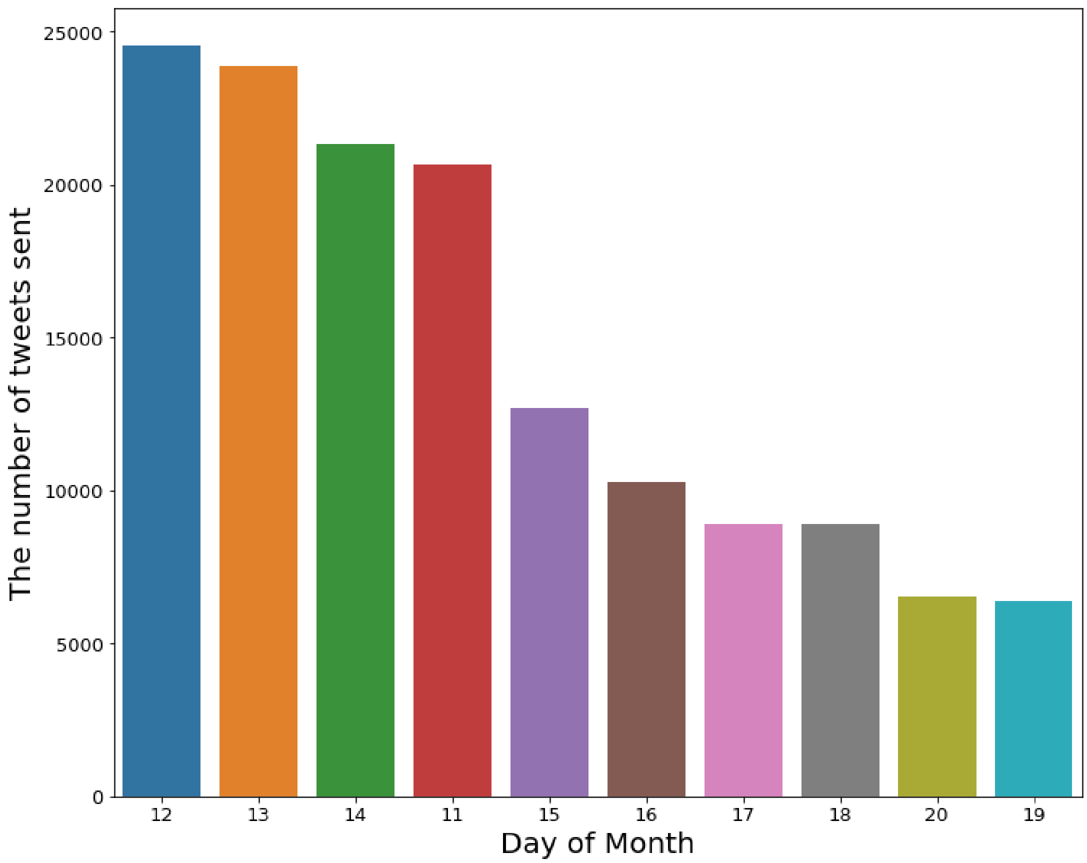


Figure 10. Number of Tweets for Each Day Data was Collected - Hurricane Florence

The wordcloud for Hurricane Florence is depicted in Figure 11. The larger the word in the word cloud, the more frequent the word appears in tweets. The color and orientation of the words are irrelevant to the data and are randomized. The majority of the most frequent terms are weather related terms or the location.

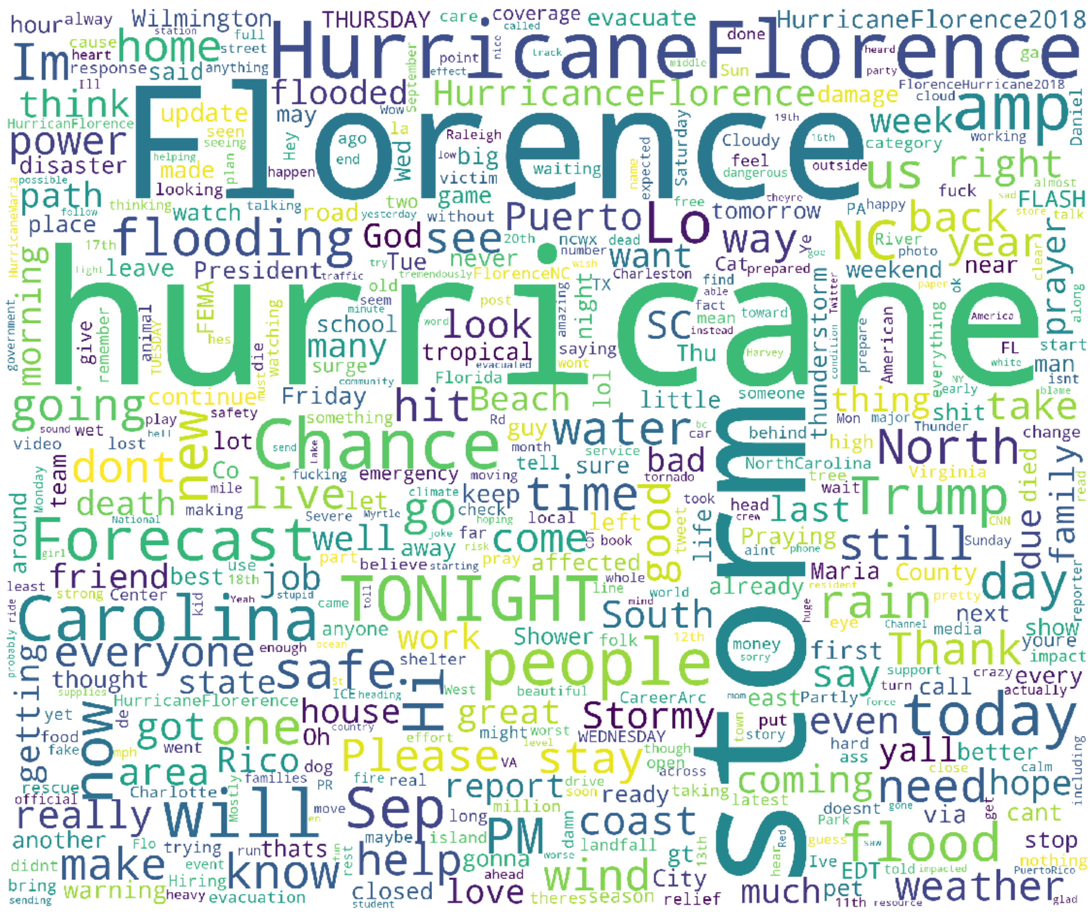


Figure 11. Word Cloud of Most Common Words in Tweets - Hurricane Florence

The list of the top 50 most frequently used words in tweets showed similar results with the majority of words being weather related. Figure 12 displays the list of terms with their frequency percentage.

```
('hurricane', 1.0),
 ('storm', 0.7559342046333618),
 ('Florence', 0.6914620740197874),
 ('HurricaneFlorence', 0.2933919628679614),
 ('Chance', 0.21572818696307153),
 ('amp', 0.20109116078335573),
 ('will', 0.19923863034892716),
 ('people', 0.1874109360368063),
 ('Carolina', 0.17833150116037622),
 ('TONIGHT', 0.17348642156263996),
 ('Forecast', 0.15805545376816904),
 ('today', 0.15524612190057407),
 ('Lo', 0.15031961239363217),
 ('Hi', 0.14736777818492733),
 ('Sep', 0.14343878506575466),
 ('now', 0.14176947192703881),
 ('NC', 0.1333211188469525),
 ('Trump', 0.12755995277065266),
 ('flood', 0.1262163592687594),
 ('new', 0.12137127967102317),
 ('safe', 0.12000732869182851),
 ('Im', 0.11860266275803102),
 ('flooding', 0.11752371646105615),
 ('one', 0.1160172631407516),
 ('PM', 0.11239363218110011),
 ('us', 0.1118846952485648),
 ('day', 0.11139611579333089),
 ('know', 0.10115630471072025),
 ('time', 0.10099344489230895),
 ('North', 0.10097308741500753),
 ('water', 0.09956842148121005),
 ('rain', 0.09861162004804365),
 ('going', 0.09209722731159155),
 ('still', 0.09134400065143927),
 ('need', 0.09067220390049265),
 ('go', 0.0902854118317658),
 ('Thank', 0.09004112210414886),
 ('stay', 0.08798501689670617),
 ('see', 0.08466674809657587),
 ('dont', 0.08417816864134196),
 ('Please', 0.083587801799601),
 ('wind', 0.08301779243516144),
 ('help', 0.08277350270754448),
 ('make', 0.08271243027564024),
 ('home', 0.08206099100199503),
 ('hit', 0.08104311713692439),
 ('good', 0.07855950490615203),
 ('live', 0.07558731322014577),
 ('everyone', 0.07493587394650054),
 ('way', 0.0738976426041285)
```

Figure 12. List of the 50 Most Common Words in Tweets - Hurricane Florence

The total sentiment score of each emotion for Hurricane Florence tweets is shown in Figure 13. Negative sentiment was the most prevalent sentiment and fear was the most common type of sentiment in tweets.

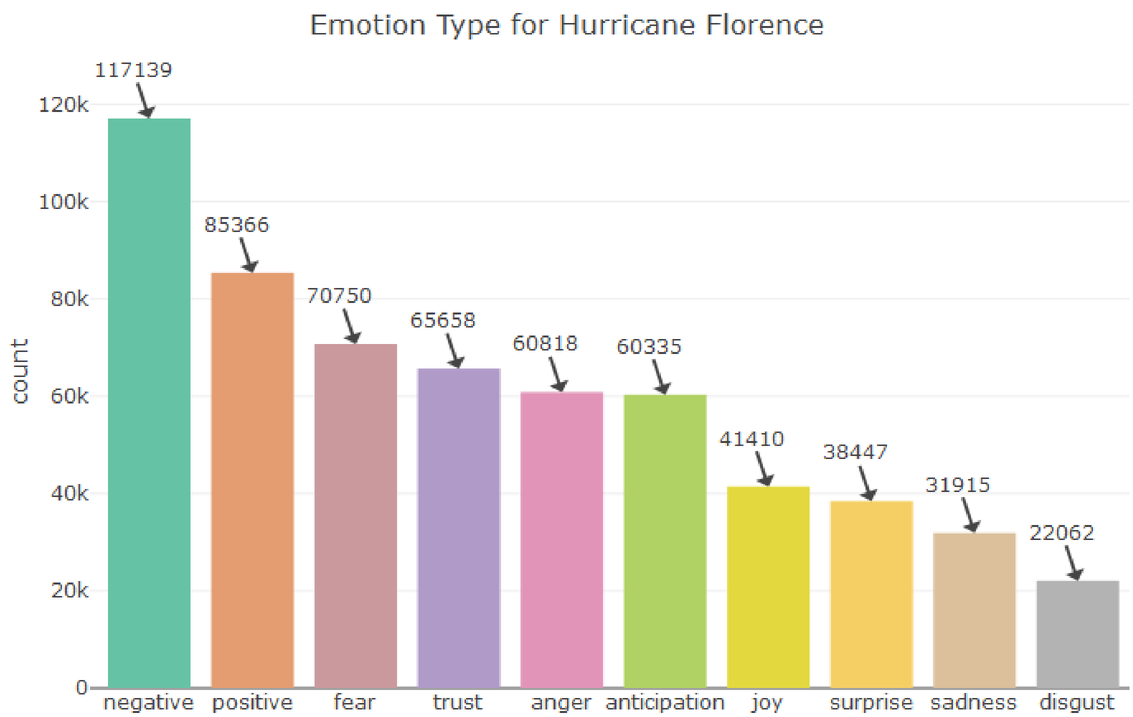


Figure 13. Number of Tweets for Each Emotion Type - Hurricane Florence

Each emotion type was analyzed for New Hanover County, NC, which was the location of the direct impact of Hurricane Florence. The counts for the different emotion types was similar to that of overall emotions regarding Hurricane Florence. Tweets from New Hanover County are displayed in Figure 14. There were more negative emotions than positive emotions, with fear being the highest emotion type and disgust having the lowest number of tweets.

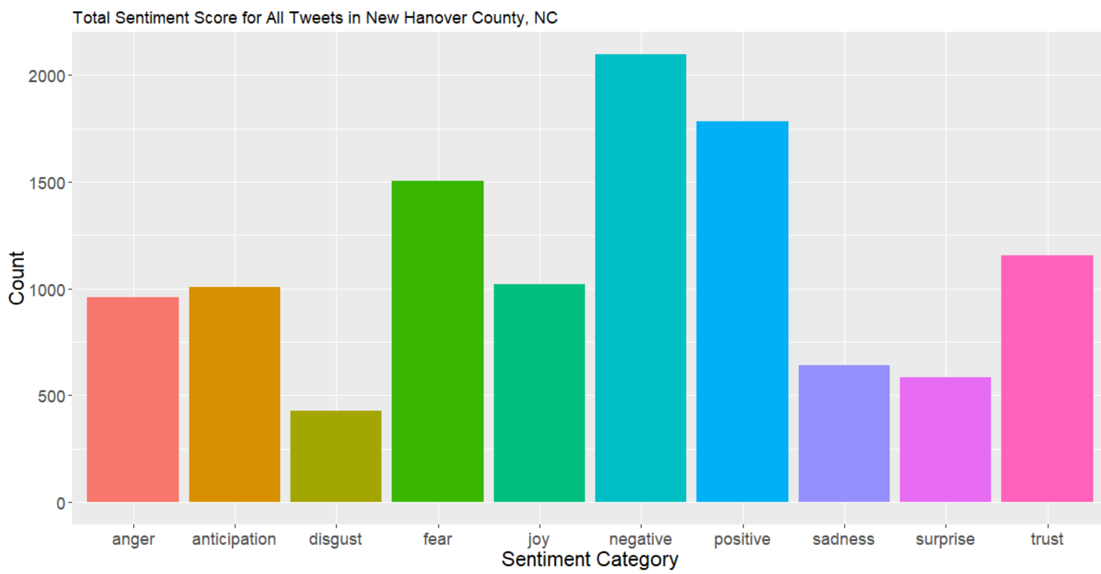


Figure 14. Number of Tweets for Each Emotion Type - Hurricane Florence - New Hanover County, NC

7.1.1.2. Hurricane Michael

The number of tweets were collected and graphed to show change over time before, during, and after Hurricane Michael. Figure 15 depicts the bar graph of number of tweets per day. The number of tweets for Hurricane Michael showed a similar pattern to that of Hurricane Florence, in that the majority of tweets were posted before and at landfall of the storm. As the storm traveled up the east coast of the United States and lost strength, volume of tweets declined.

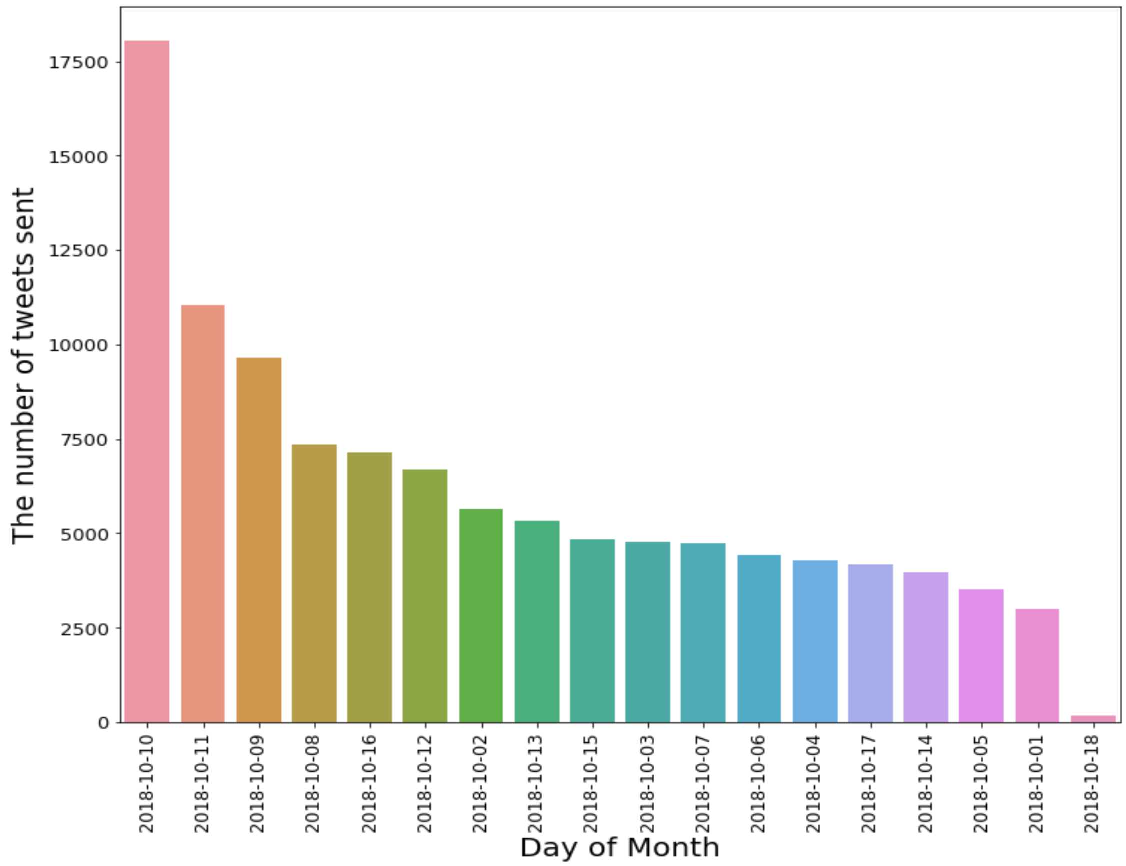


Figure 15. Number of Tweets for Each Day in Data Collected - Hurricane Michael

Figure 16 shows the wordcloud for Hurricane Michael. The largest words are, again, the most frequently used words in the tweets. The data was randomized, yielding orientation and color irrelevant to the data. Weather related terms occurred the most frequently, as they did in analysis of Hurricane Florence in both the wordcloud analysis and the list of frequent terms.

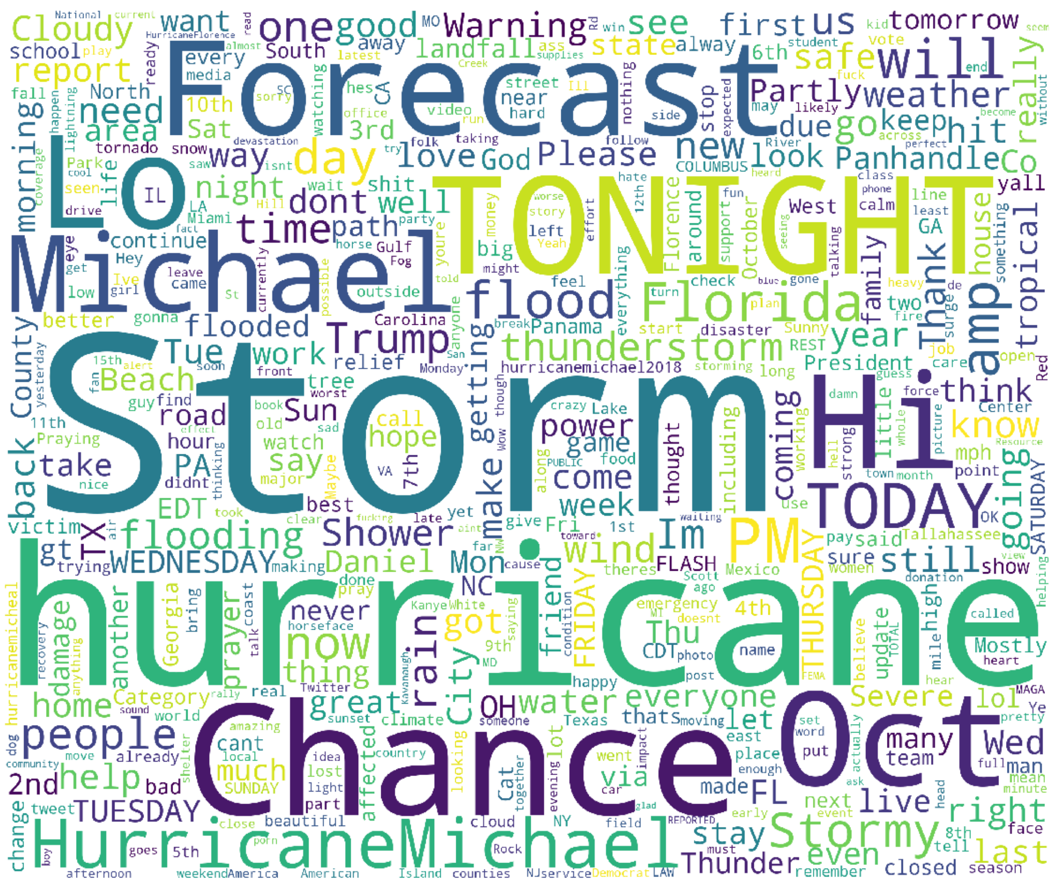


Figure 16. Word Cloud of Most Common Words in Tweets - Hurricane Michael

Figure 17 is the list of the top 50 most frequently used words in tweets with the word frequency.

```
('Storm', 1.0),
 ('hurricane', 0.7433006790027471),
 ('Chance', 0.4447986316280516),
 ('TONIGHT', 0.3400196962628933),
 ('Forecast', 0.3187166329756906 4),
 ('Oct', 0.3139221479293008),
 ('Hi', 0.3115637796091847),
 ('Lo', 0.30866117244596486),
 ('Michael', 0.2632820193852693),
 ('HurricaneMichael', 0.2616233867205722),
 ('PM', 0.22513346809723733),
 ('TODAY', 0.22500388742030789),
 ('amp', 0.16677033120821022),
 ('will', 0.15964339397709013),
 ('Florida', 0.1594619810293889),
 ('flood', 0.14308298346550563),
 ('day', 0.12499352096615353),
 ('Stormy', 0.12431970144612035),
 ('now', 0.12154667495982999),
 ('people', 0.10267972839890116),
 ('one', 0.10117659254651946),
 ('Trump', 0.09819623697714197),
 ('Im', 0.09594153319856943),
 ('time', 0.09350541647229565),
 ('rain', 0.08943658321671072),
 ('thunderstorm', 0.086274814699632),
 ('new', 0.08262063961022133),
 ('us', 0.08233556212097652),
 ('flooding', 0.08176540714248691),
 ('wind', 0.07868138703156585),
 ('Shower', 0.07738558026227128),
 ('know', 0.07601202508681905),
 ('go', 0.07375732130824651),
 ('Thank', 0.07222826932047893),
 ('need', 0.07069921733271134),
 ('weather', 0.06865184263722594),
 ('see', 0.06761519722179028),
 ('going', 0.06681179702482766),
 ('good', 0.06652671953558285),
 ('safe', 0.0660861452340227),
 ('back', 0.06598248069247914),
 ('hit', 0.06567148706784844),
 ('help', 0.06512724822474472),
 ('make', 0.06476442232934225),
 ('Cloudy', 0.0634167832892759),
 ('Warning', 0.06310578966464521),
 ('still', 0.06279479604001452),
 ('FL', 0.06162856994764940 4),
 ('come', 0.06121391178147514 4),
 ('Tue', 0.06118799564608925 4)
```

Figure 17. List of the 50 Most Common Words in Tweets - Hurricane Michael

The total sentiment score of each emotion for Hurricane Michael tweets is shown in Figure 18. Negative sentiment was the most prevalent sentiment and fear was the most common type of sentiment in tweets.

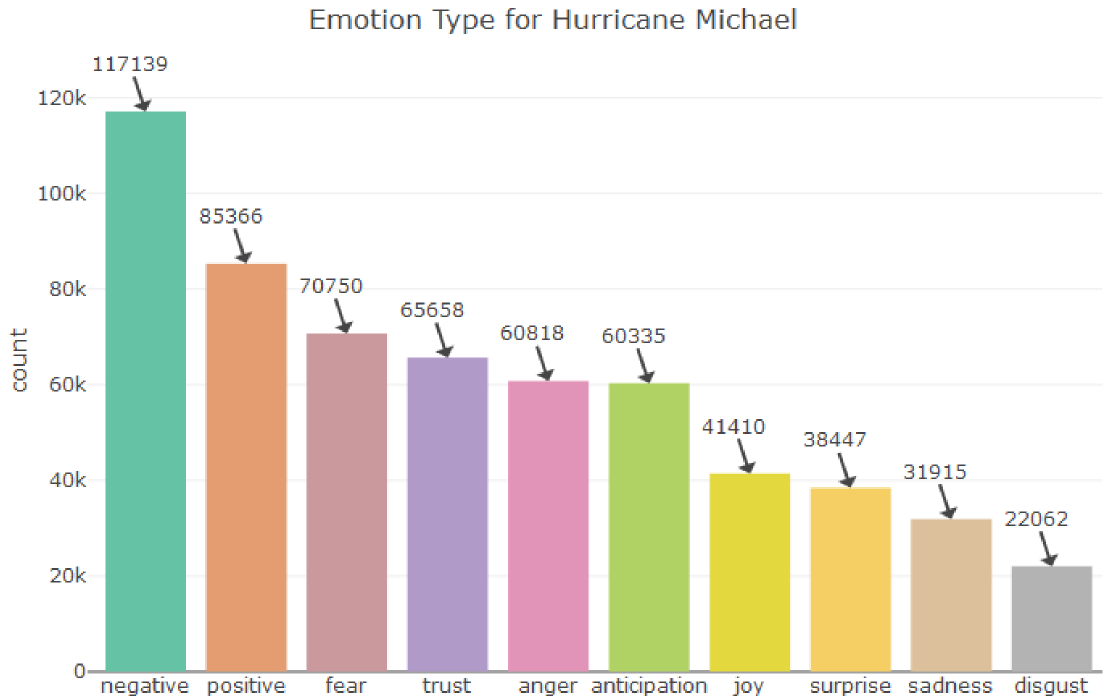


Figure 18. Number of Tweets for Each Emotion Type - Hurricane Michael

Each emotion type was analyzed for Bay County, FL, which was the location of the direct impact of Hurricane Michael. Emotion in Bay County was similar to that of the overall emotions of Hurricane Michael. Tweets from Bay County are displayed in Figure 19. There were more negative emotions than positive emotions, with fear being the highest emotion type and disgust having the lowest number of tweets.

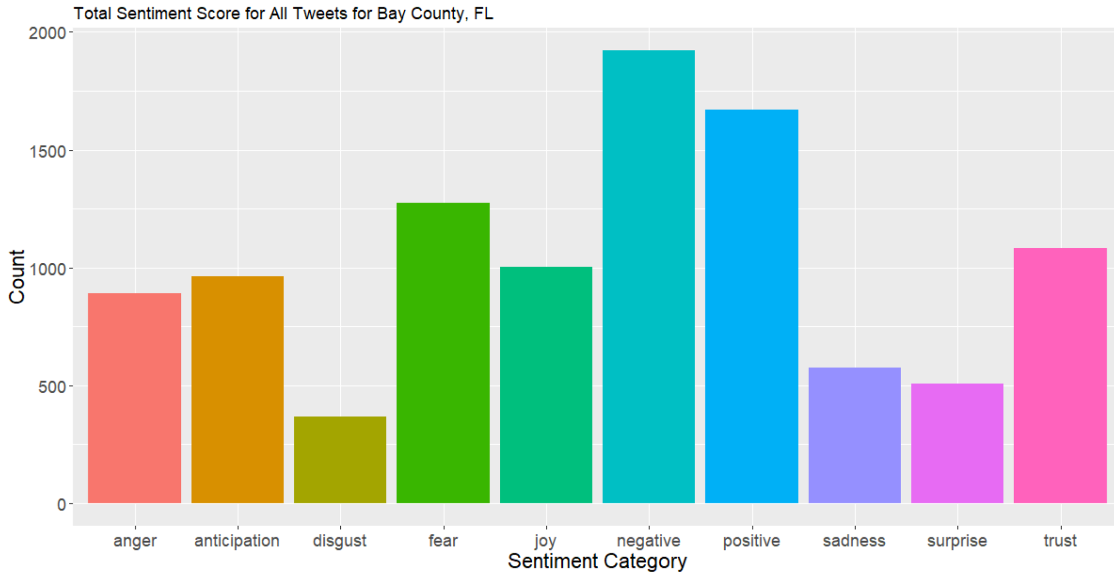


Figure 19. Number of Tweets for Each Emotion Type - Hurricane Michael - Bay County, FL

7.1.2. Temporal Patterns

Data was analyzed based on time before, during, and after Hurricanes Florence and Michael. A time series graph was created of the sentiments expressed on Twitter for each hurricane, and then for each of the counties where the hurricanes made landfall.

7.1.2.1. Hurricane Florence

The average sentiment score over time was analyzed for positive versus negative sentiment of tweets about Hurricane Florence in Figure 17. Positive sentiment appeared to remain relatively constant before, during, and after Hurricane Florence. From September 11, 2018 through September 20, 2018, negative sentiment was in contrast to positive sentiment. Negative sentiment was high at the beginning of the storm, peaked on September 13, 2018, and steadily decreased through the remainder of the storm.

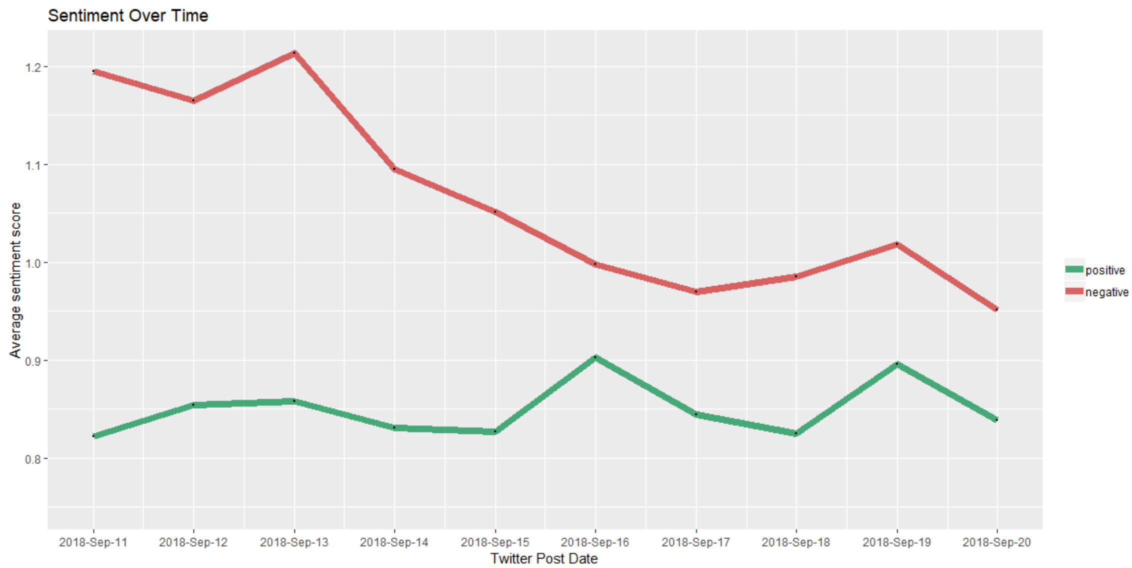


Figure 20. Average Sentiment Score Over Time - Hurricane Florence

The average sentiment score of tweets for New Hanover County, NC, as shown in Figure 21, was very different from that of all sentiment for Hurricane Florence. Positive sentiment was high on September 12, 2018 with negative sentiment being lower than positive sentiment when the wind shear increased and the storm started to taper. This was before the storm gained strength on September 13, 2018. There was a peak of negative sentiment on September 14, 2018 before landfall and dropped until September 18, 2018 when negative sentiment began to rise again. Positive sentiment had higher scores than negative sentiment from September 16, 2018, which was after Hurricane Florence made landfall. Peak positive sentiment occurred on the last day that data was collected, September 20, 2018.

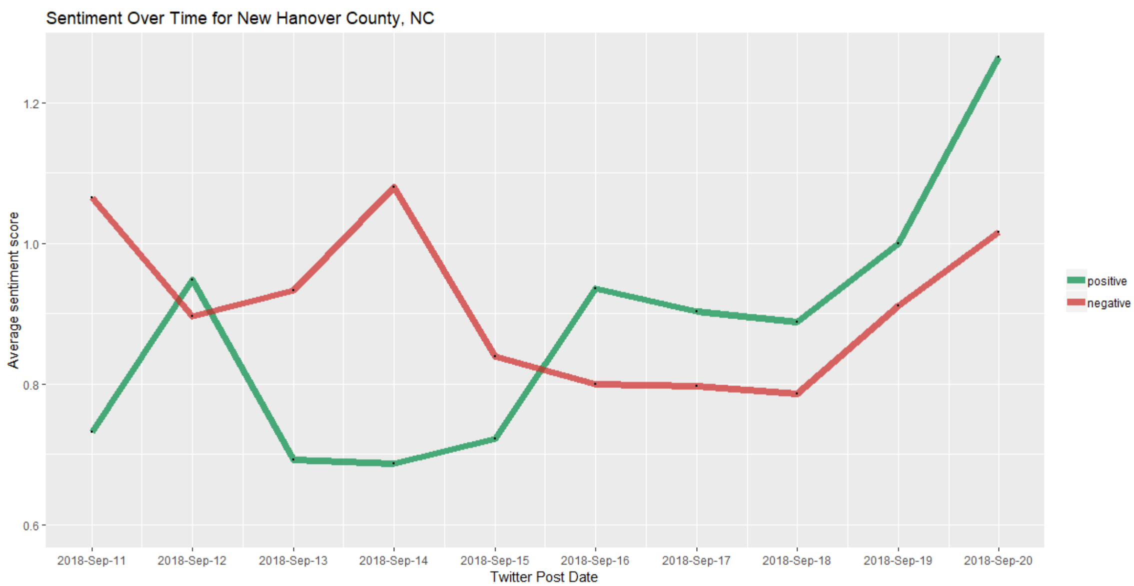


Figure 21. Average Sentiment Score Over Time - Hurricane Florence - New Hanover County, NC

When analyzing specific emotion of sentiment, the average sentiment score for each emotion was then plotted over time with fear being the most prevalent in tweets from September 11, 2018 through September 20, 2018 in Figure 22. Fear appeared to decrease as the hurricane passed through and dissipated. Surprise appeared to increase as the storm passed. Other emotions remained fairly constant through the entirety of the storm.

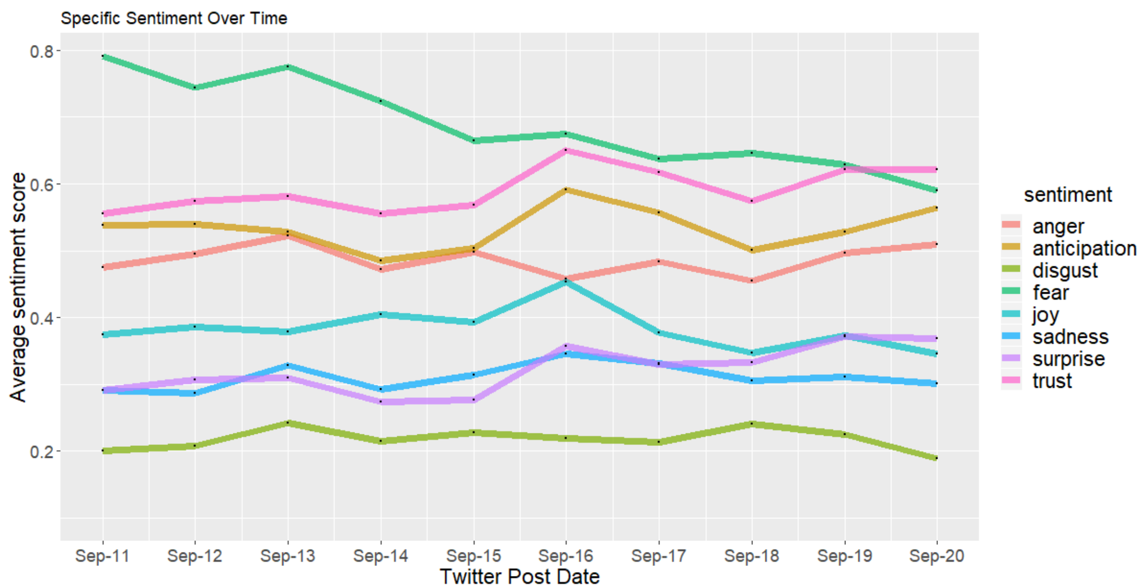


Figure 22. Average Sentiment of Specific Emotions Over Time - Hurricane Florence

Figure 23 shows the specific sentiment over time of tweets for New Hanover County, NC. Fear had the largest peak in tweets as the hurricane was making landfall. Fear decreased after, with a slight increase at the end of the data collection period. Trust appeared to stay fairly steady until post-hurricane where score increased higher than any other score. All other emotions stayed relatively stable with minor increases and decreases from day to day.

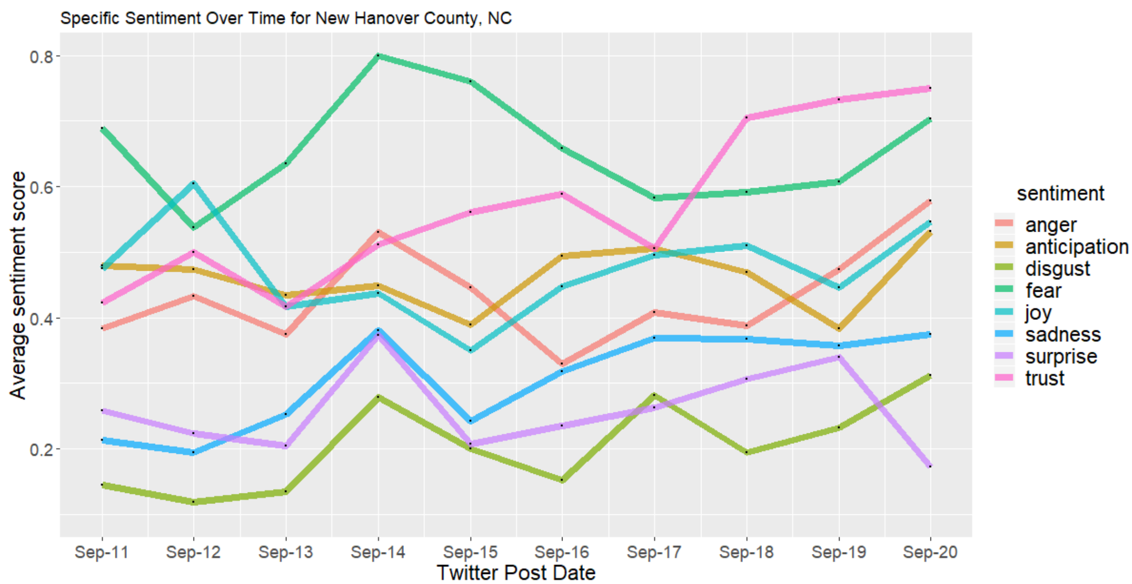


Figure 23. Average Sentiment of Specific Emotions Over Time - Hurricane Florence - New Hanover County, NC

7.1.2.2. Hurricane Michael

Positive and negative sentiments, displayed as the average sentiment score, of tweets regarding hurricane Michael over time are displayed in Figure 24. Negative sentiment peaked on October 10, 2018 when the hurricane was just about the make landfall. Positive sentiment peaked after the hurricane made landfall. There was a spike in positive sentiment on October 7, 2018 when the storm was named a tropical depression, before it was a named hurricane. Negative and positive sentiment were closer to neutral sentiment following Hurricane Michael's landfall.

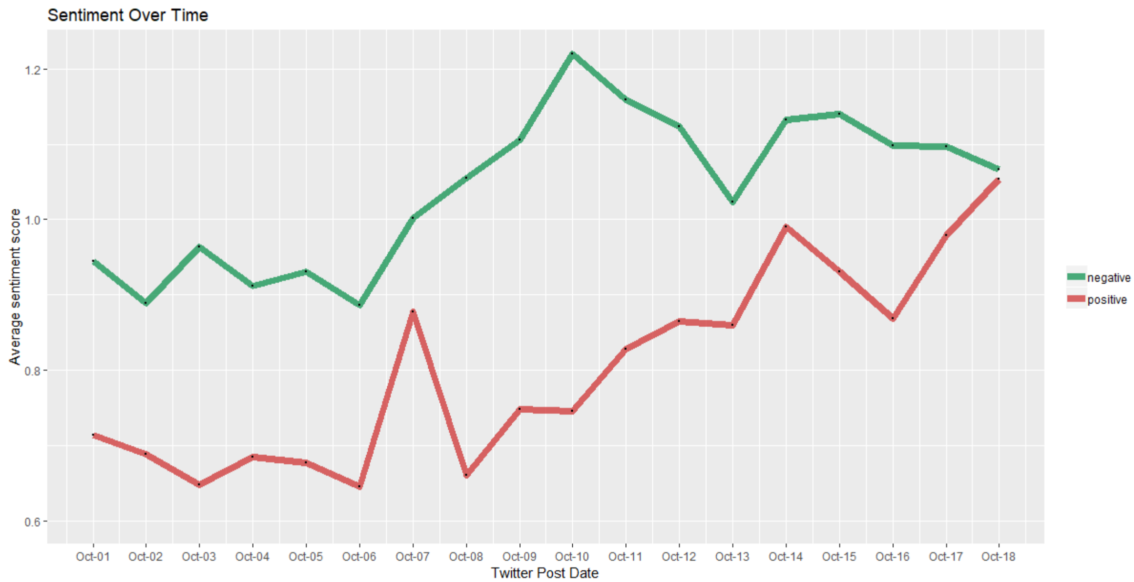


Figure 24. Average Sentiment Score Over Time - Hurricane Michael

Sentiment score over time for Bay County, FL is displayed in Figure 25. This graph displayed much different results from that of all of the tweets regarding Hurricane Michael. Positive sentiment began lower than negative sentiment on October 1, 2018. By October 2, 2018 positive sentiment increased and negative sentiment decreased to similar values. Positive and negative sentiment showed an increase and decrease with the progression of time for the hurricane event in a similar pattern. Positive sentiment showed an increase from the time the hurricane hit land until October 14, 2018. There was a decrease in positive sentiment for two days and positive sentiment began to rise again for Bay County. From October 12, 2018, after Hurricane Michael made landfall, positive and negative tweet sentiment scores rose and fell simultaneous, but positive sentiment showed higher scores than negative sentiment from October 12, 2018 through the end of the data collection period. Negative sentiment dropped on October 17, 2018 while positive sentiment continued to increase.

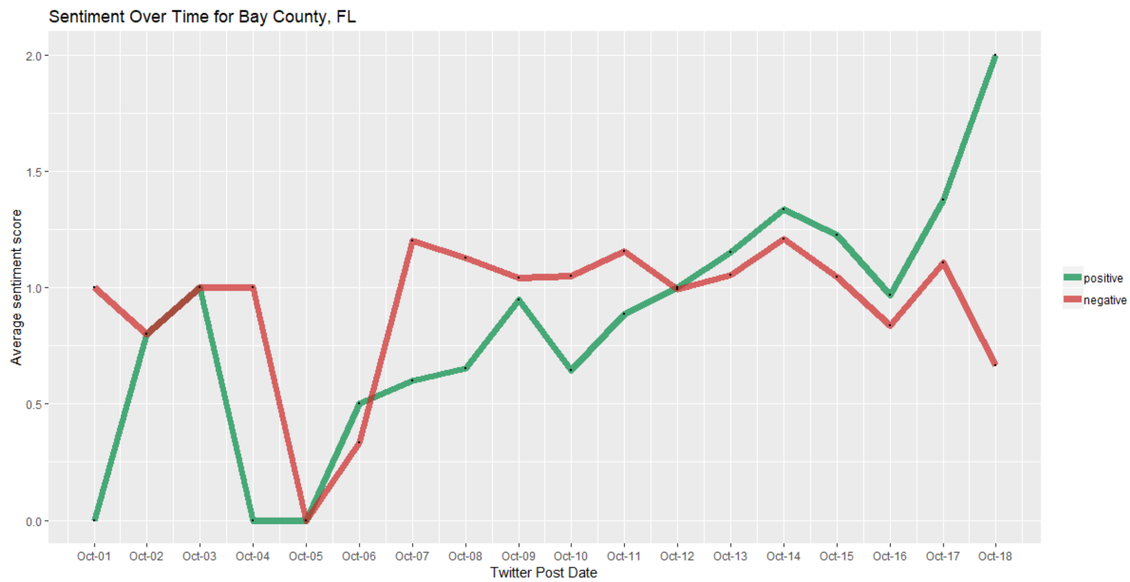


Figure 25. Average Sentiment of Specific Emotions Over Time - Hurricane Michael - Bay County, FL

The average sentiment for each emotion was plotted over time for Hurricane Michael from October 1, 2018 through October 1, 2018 in Figure 26. Trust, anticipation, surprise, and joy all peaked on October 7, 2018, prior to the storm making landfall. Fear peaked on October 10, 2018 as the hurricane made landfall. All sentiment showed a slightly increasing trend following the hurricane passed through the United States.

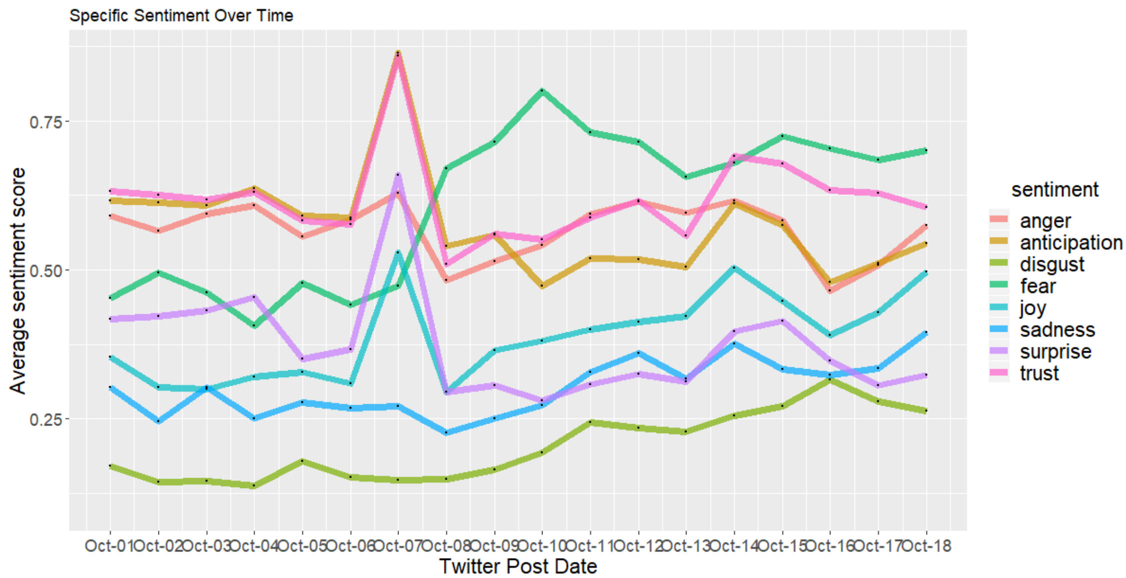


Figure 26. Average Sentiment of Specific Emotions Over Time - Hurricane Michael

Figure 27, below, shows the specific sentiment over time of tweets for Bay County, NC. Fear and joy had the largest peak in tweets. The peak in fear occurred before the hurricane made landfall, and the peak in joy occurred after the storm passed. All emotions stayed relatively stable with a minor increasing trend following the hurricane. These results were similar to those of all tweets following the hurricane.

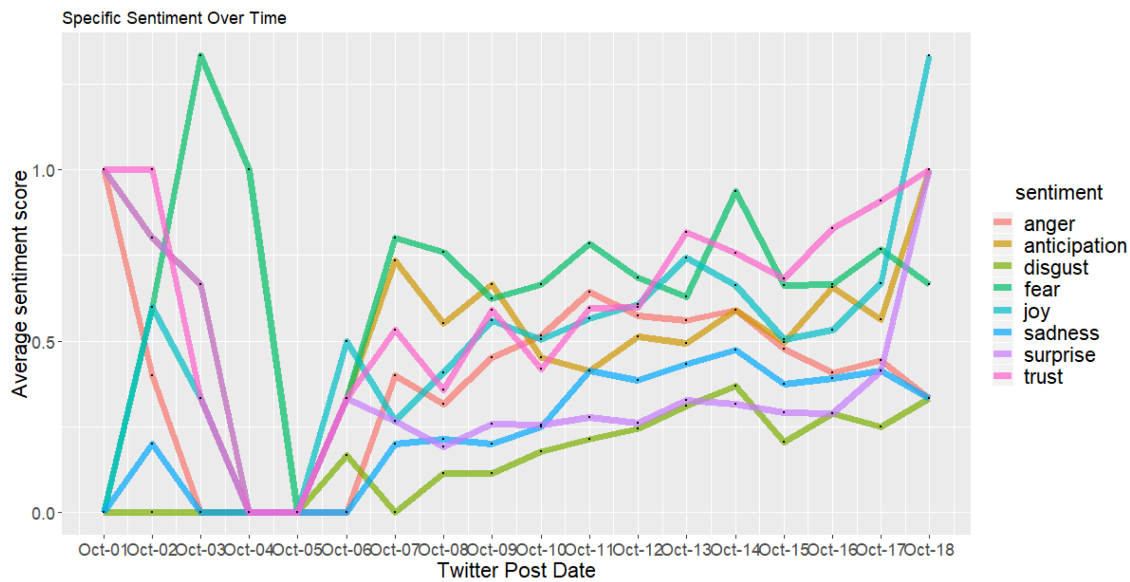


Figure 27. Average Sentiment of Specific Emotions Over Time - Hurricane Michael - Bay County, FL

7.1.3. Classification Analysis of Sentiment

This section focuses on comparing and evaluating different machine learning models. The aim is to select the best model for predicting twitter sentiment. The algorithm with the highest performance will be identified as the preferred model for prediction analysis of extreme weather event related tweets. The accuracy is the percentage of correctly classified sentiments. Naive Bayes, SVM, random forest, boosting, and maxent were all used for classification models of the sentiment datasets. Fourfold cross validation was used for the SVM, random forest, boosting, and maxent models. Each model was evaluated under four criteria; including accuracy, cross validation accuracy, precision, recall, and F-score. For both data sets, naive Bayes performed the best for predicting sentiment based on the data.

7.1.3.1. Hurricane Florence

The results of the evaluation of classification models for the Hurricane Florence Twitter dataset are shown below in Table 1. Using the method of cross validation increased the performance of the boosting classification model, but decreased the performance of the SVM, random forest, and maxent models.

Model	Accuracy	Cross Validation Accuracy	Precision	Recall	F-score
Naive Bayes	63.77%		0.55078	0.5701	0.54597
Support Vector Machine	37.64%	32.29%	0.0800	0.2500	0.1200
Random Forest	37.64%	32.28%	0.0800	0.2500	0.1200
Boosting	38.88%	86.23%	0.1475	0.2525	0.1850
Maximum Entropy	38.92%	30.75%	0.1475	0.2500	0.1775

Table 1. Model Evaluation for Hurricane Florence Tweets

The naive Bayes model that was created used a portion of the data due to memory constraints of the computer used in this research. The model produced the following table of actual versus predicted sentiments.

```
                 Actual
Predictions negative neutral positive
   negative      108      21       17
   neutral        11       5        8
   positive       64      60      206
```

The overall statistics of the model were as follows:

```
Overall Statistics

               Accuracy : 0.638
                 95% CI : (0.5942, 0.6802)
    No Information Rate : 0.462
    P-Value [Acc > NIR] : 1.920e-15

                  Kappa : 0.3758

 Mcnemar's Test P-Value : 3.942e-15
```

The accuracy of the model for predicting sentiment values was 63.8% and showed the highest accuracy of all the models tested.

7.1.3.2. Hurricane Michael

The results of the evaluation of classification models for the Hurricane Florence Twitter dataset are shown below in Table 1. Using this method of cross validation increased the performance of all of the classification models.

Model	Accuracy	Cross Validation Accuracy	Precision	Recall	F-score
Naive Bayes	66.6%		0.5891	0.6372	0.59197
Support Vector Machine	2.784%	43.20%	0.1433	0.3333	0.2000
Random Forest	2.784%	43.31%	0.1433	0.3333	0.2000
Boosting	2.784%	100%	0.1200	0.3333	0.1767
Maximum Entropy	2.784%	43.31%	0.1433	0.3333	0.2000

Table 2. Model Evaluation for Hurricane Michael Tweets

A portion of the data was used for the naive Bayes model due to the memory constraints of the computer. The following table shows the reference versus the predicted sentiment using the model.

	data_test_labels1		
sms_test_pred1	**negative**	**neutral**	**positive**
negative	**714**	**123**	**113**
neutral	**68**	**149**	**72**
positive	**309**	**317**	**1135**

The overall statistics of the model are as follows:

```
Overall Statistics

               Accuracy : 0.666
                 95% CI : (0.6488, 0.6829)
    No Information Rate : 0.587
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4503

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: negative Class: neutral Class: positive
Precision                     0.6544        0.25297          0.8598
Recall                        0.7516        0.51557          0.6445
F1                            0.6997        0.33941          0.7368
Prevalence                    0.3167        0.09633          0.5870
Detection Rate                0.2380        0.04967          0.3783
Detection Prevalence          0.3637        0.19633          0.4400
Balanced Accuracy             0.7838        0.67663          0.7476
```

The accuracy of the naive Bayes model for the Hurricane Michael data for predicting sentiment values was 66.6% and had the highest accuracy of all models tested.

7.2 Physical Impact Analysis

Cloud cover temperature was collected for specific latitude and longitude respective of each hurricane. This data was plotted on maps in R. The maps of Hurricane Florence's cloud cover temperatures per time period are displayed in Appendix A. The maps of Hurricane Michael's cloud cover temperatures per time period are displayed in Appendix B. The maps in Appendices A and B show the progression of cloud cover temperature over the time period of each of the hurricanes. The maps, when viewed as a progressive collection of images, show each of the storms moving into the coast and then offshore. The latitude and longitude of each hurricane was used to determine proximity of tweet location to the center of each hurricane. The data was then used in correlation analysis. The results of that analysis are found in Section 7.4 of this study. Specific cloud cover temperature data for each of the counties, where each hurricane made landfall, was then used to evaluate and determine any correlation that may be present with sentiment during the time period of each of the storms

7.3 Correlation Analysis

Sentiment and physical data were analyzed for correlation between features within the datasets. Average sentiment score, distance between hurricane center and tweet location, and average cloud cover temperature per day were analyzed for correlation of sentiment and physical data. Pearson's product-moment correlation test was used along with Kendall rank correlation test and Spearman rank correlation coefficient test. Q-Q plots were then created to identify possible correlations in a visually display

7.3.1. Hurricane Florence

7.3.1.1. Sentiment and Cloud Cover Data

Average sentiment scores and average cloud cover temperatures were plotted against each other in Figure 28. The relationship between the two variables does not seem to be a linear relationship.

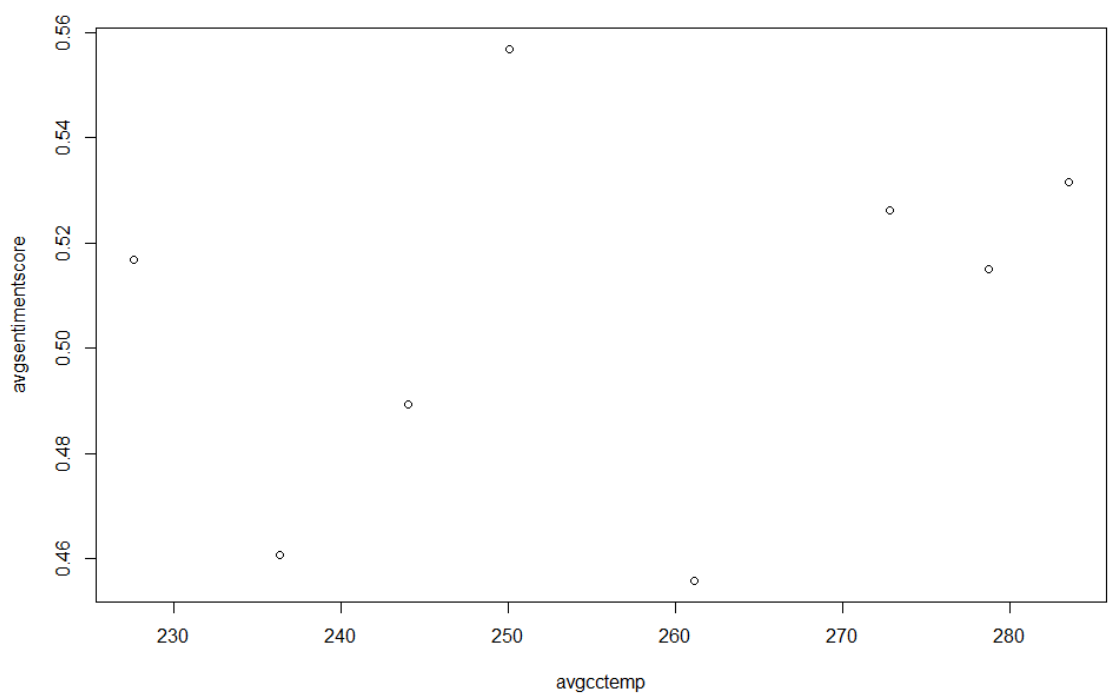


Figure 28. Covariation of Average Sentiment Score Versus Average Cloud Cover Temperature - Hurricane Florence

Correlation coefficients were calculated from the data for Hurricane Florence using the Pearson method, the Kendall method, and the Spearman method. The Pearson's method produced a correlation coefficient of 0.2904746. 0.2142857 was the correlation coefficient when using the Kendall method. 0.3095238 was the correlation coefficient using the Spearman method.

A preliminary test was conducted to identify if there is linear covariation and the results were plotted. The variables were then analyzed to identify if they follow a normal distribution. Shapiro-Wilk normality test was used for each variable. The p-value for average sentiment score was 0.5466 and the p-value for the average cloud cover temperature was 0.7011. From the output, the two p-values are greater than the significance level of 0.05. This implies that the distributions of the data for each variable are not significantly different from normal distribution. Normality can be assumed in this case. The Q-Q plot of average sentiment score versus average cloud cover temperature is shown in Figure 29. The Q-Q plot of average sentiment scores versus theoretical values is shown in Figure 30, and the Q-Q plot of average cloud cover temperatures versus theoretical values is shown in Figure 31. From visual inspection of the Q-Q plots, it is concluded that both populations of data may come from normal distributions.

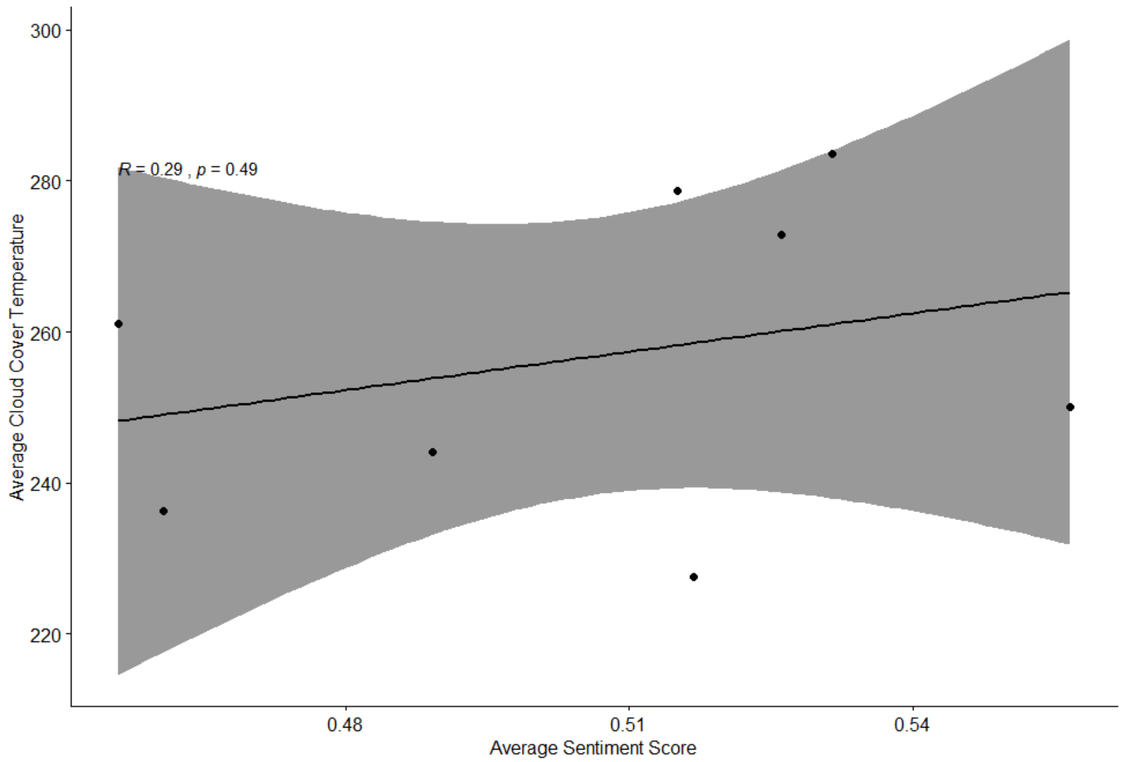


Figure 29. Q-Q Plot of Average Sentiment Score Versus Average Cloud Cover Temperature

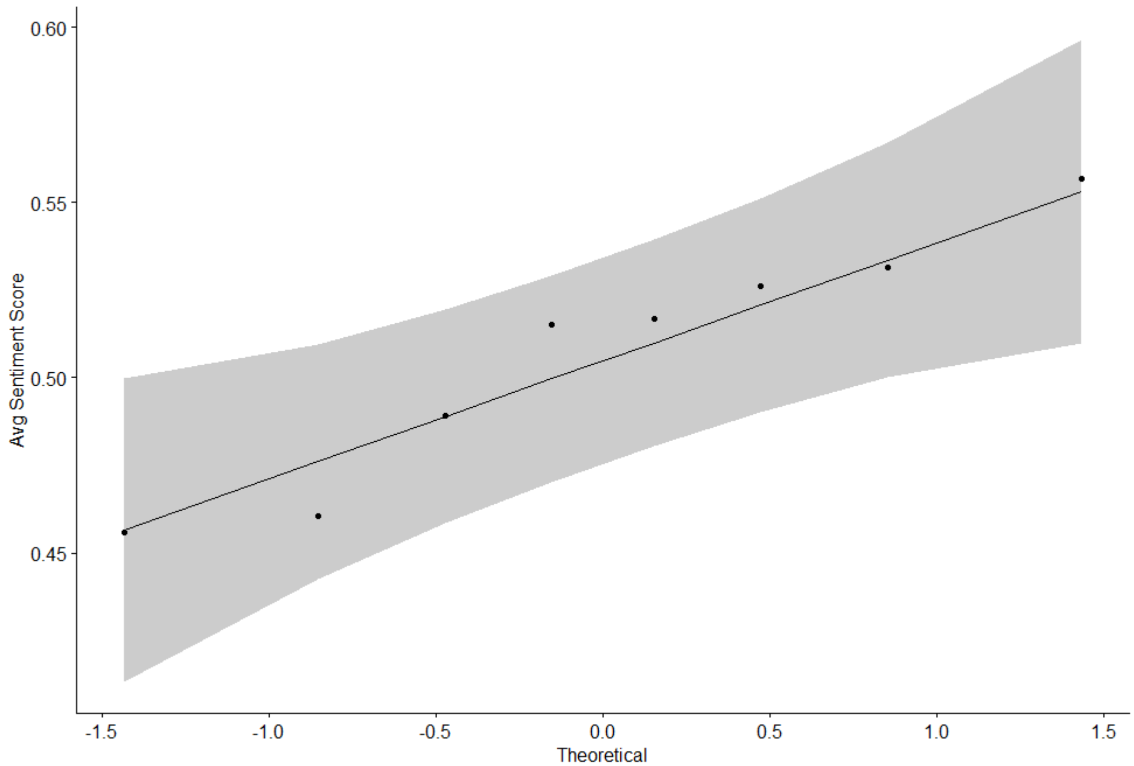


Figure 30. Q-Q Plot of Average Sentiment Scores Versus Theoretical Values

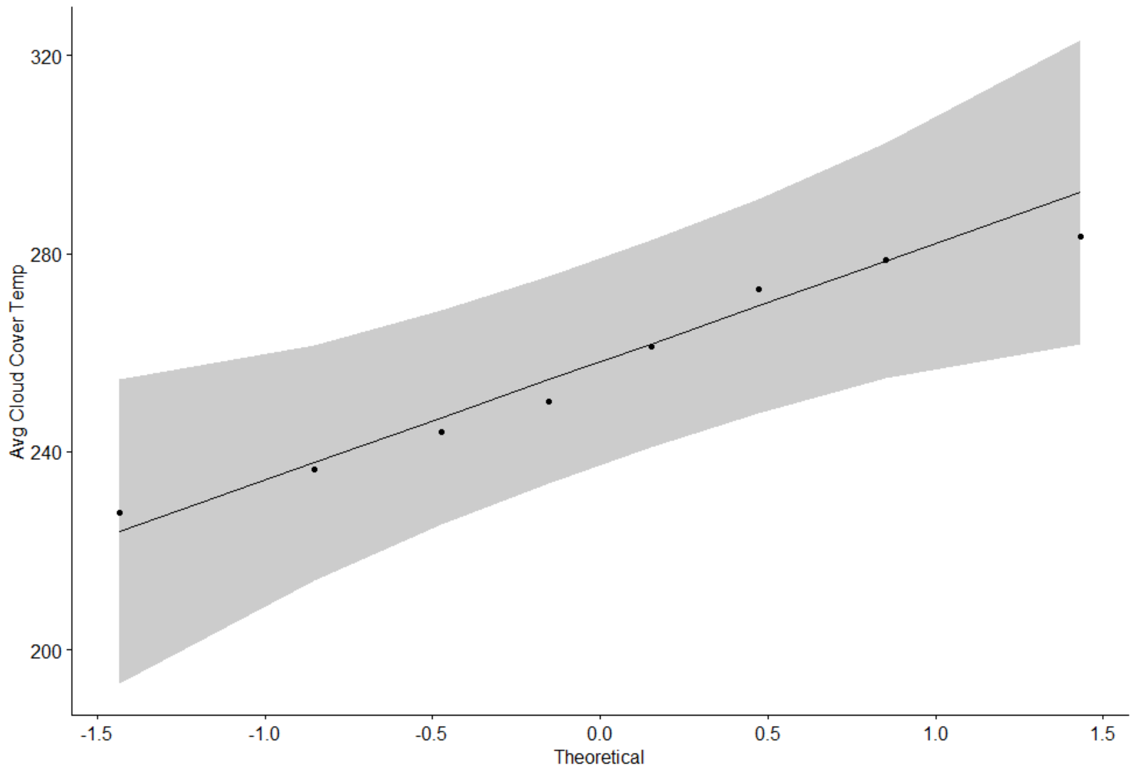


Figure 31. Q-Q Plot of Average Cloud Cover Temperatures Versus Theoretical Values

The Pearson correlation test, Kendall rank correlation test, and the Spearman rank correlation coefficient test were conducted on average sentiment score versus average cloud cover temperature variables. The Pearson's product-moment correlation coefficient was 0.2904746. The correlation coefficient is positive, indicating a positive correlation between the two variables, but because the value is close to zero, great variation in the data around the line of best fit. The p-value was 0.4852 and greater than the significance level alpha = 0.05 indicating that the correlation coefficient is not statistically significant and the variables are not significantly correlated.

Kendall rank correlation test was used to estimate a rank-based measure of association. The test was used due to the data not necessarily coming from a bivariate normal distribution. The test did not yield a higher correlation coefficient between the two variables than with the Pearson correlation test. The correlation coefficient was 0.2142857 and the p-value was 0.5484. There is great variation around the line of best fit and the variables are not significantly correlated due to the p-value being greater than the significance level of alpha = 0.05.

The Spearman rank correlation coefficient is estimated based on the two variables. This test is also used when the data does not come from a bivariate normal distribution. The test yielded the highest correlation coefficient of all three tests. The correlation coefficient was 0.3095238 and the p-value is 0.4618. There is a weak positive correlation between the variables based on the correlation coefficient. The p-value is greater than the significance level of alpha = 0.05 indicating that the variables are not statistically significantly correlated.

All of the data for sentiment scores and cloud cover temperatures for each day during Hurricane Florence were then analyzed using the same three methods; Pearson's correlation test, Kendall's correlation test, and Spearman's correlation test. These tests were conducted on sentiment score versus cloud cover temperature variables for Hurricane Florence data. When all of the data was analyzed for correlation, -0.007207557 was the Pearson's correlation coefficient. The correlation coefficient was negative, indicating that there may be a negative correlation between the variables, but because the value is so close to zero, it is indicative of great variation in the data around the line of best fit. The p-value was 0.009033, which is less than the significance level of alpha = 0.05, thus the correlation coefficient is statistically significant. The variables do not appear to be significantly correlated. The Kendall rank correlation test was used to estimate a rank-based measure of association. This test was used in addition to the Pearson's test because the data did not necessarily exhibit being from a bivariate normal distribution. The Kendall method gave a correlation coefficient of -0.006283178 and a p-value of 0.0014. The variables did not seem to be significantly correlated, and the correlation coefficient was statistically significant due to the p-value being less that the significance level of alpha = 0.05. The Spearman method correlation coefficient was -0.008893299 and the p-value was 0.001276. The relationship between sentiment score and cloud cover temperature was very slightly negative, and the correlation coefficient was statistically significant.

7.3.1.2. Sentiment and Proximity Data

The distance between the tweet location and the center of the storm was determined and added as a variable to the dataset. Correlation coefficients were calculated using the Pearson method, the Kendall method, and the Spearman method for the data of Hurricane Florence. The Pearson correlation coefficient was 0.04074419. The Kendall correlation coefficient was 0.02492845. The Spearman correlation coefficient was 0.03738075.

The p-value for all three correlation tests is < 2.2e-16, indicating that there is a correlation between the variables, but the correlation coefficient of each test indicates that the correlation is not strong. The correlation indicates a positive relationship between the variables, with sentiment becoming more negative as the hurricane center becomes closer to tweet location. The correlation plot of the two variables is shown in Figure 32.

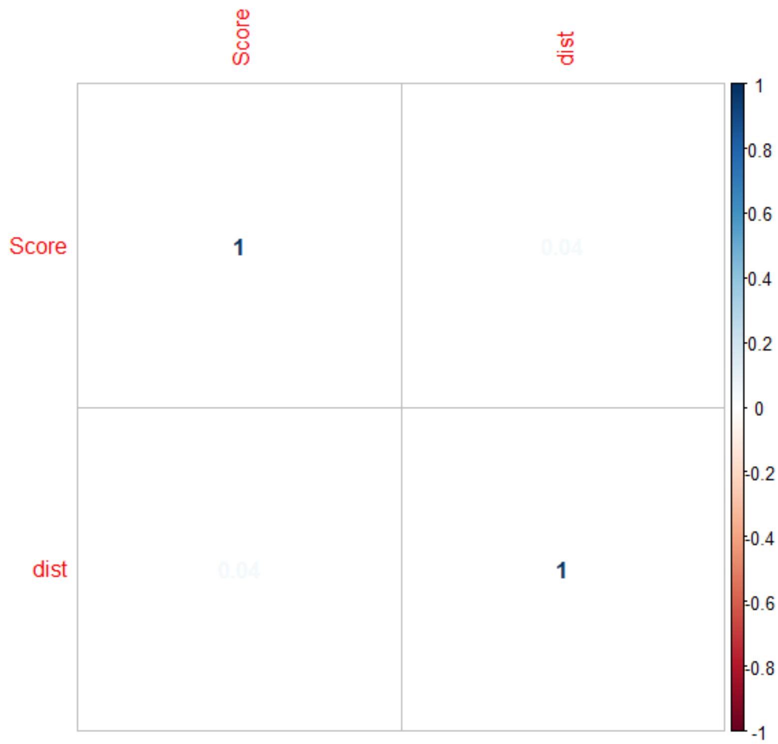


Figure 32. Correlation Plot of Sentiment Score Versus Distance of Tweet from Center of Hurricane Florence.

7.3.2. Hurricane Michael

7.3.2.1 Sentiment and Cloud Cover Data

The variables of average sentiment score and average cloud cover temperature were plotted against each other in Figure 32 below. Average sentiment score versus average cloud cover temperature per day does not appear to have a linear relationship.

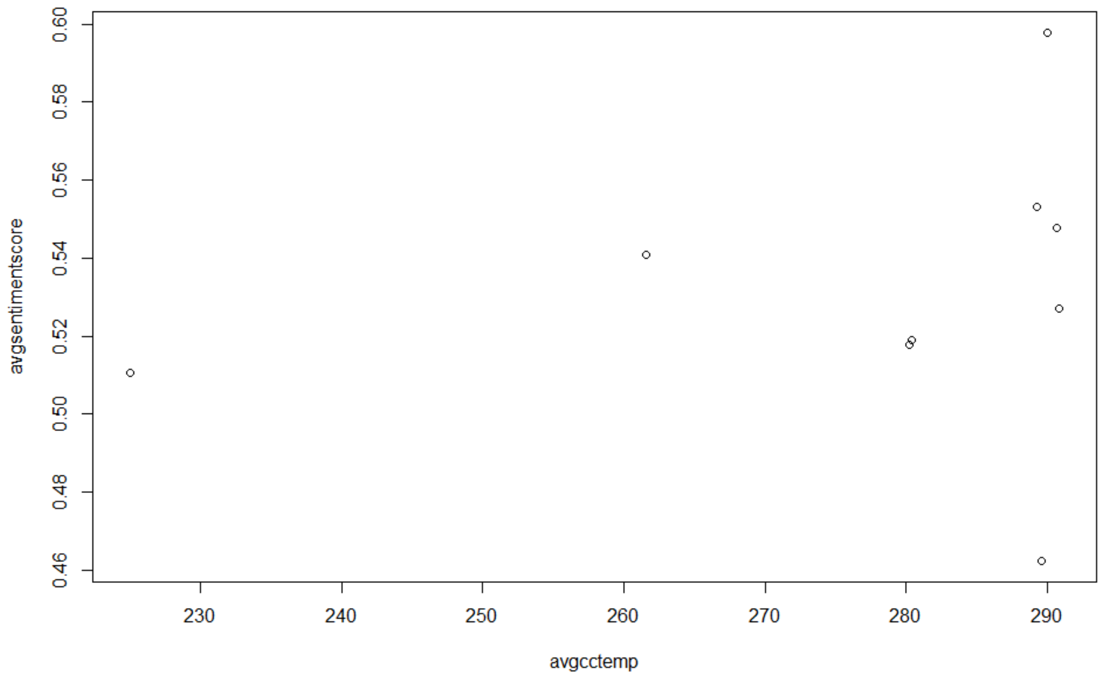


Figure 32. Covariation of Average Sentiment Score Versus Average Cloud Cover Temperature - Hurricane Michael

The correlation coefficients were then calculated using the Pearson method, the Kendall method, and the Spearman method. The Pearson's method produced a correlation coefficient of 0.1988518. The Kendall method produced a correlation coefficient of 0.2777778. The Spearman method produced a correlation coefficient of 0.4.

Linear covariation of the data was then analyzed and plotted. The variables were analyzed to identify if they follow a normal distribution. Each variable was analyzed using the Shapiro-Wilk normality test. The average sentiment score p-value was 0.7515 and the average cloud cover temperature p-value was 0.00087. The output demonstrated that the value for average sentiment score was greater than the significance level of 0.05 implying that the distribution for this data is significantly different from normal distribution. However, the p-value for the average cloud cover temperature is less than the significance level, implying that the variable data is not significantly different from the normal distribution. Normality can only be assumed for the cloud cover data from Hurricane Michael. Q-Q plots were used to visually identify correlation between a given sample and the normal distribution. The Q-Q plot of the average sentiment score versus average cloud cover temperature is show in Figure 33. The Q-Q plot of average sentiment scores versus theoretical values is shown in Figure 34, and the Q-Q plot of average cloud cover temperatures versus theoretical values is shown in Figure 35. From visual inspection of these normality plots, we conclude that the average cloud cover temperature populations may come from normal distributions, but the sentiment score populations may not.

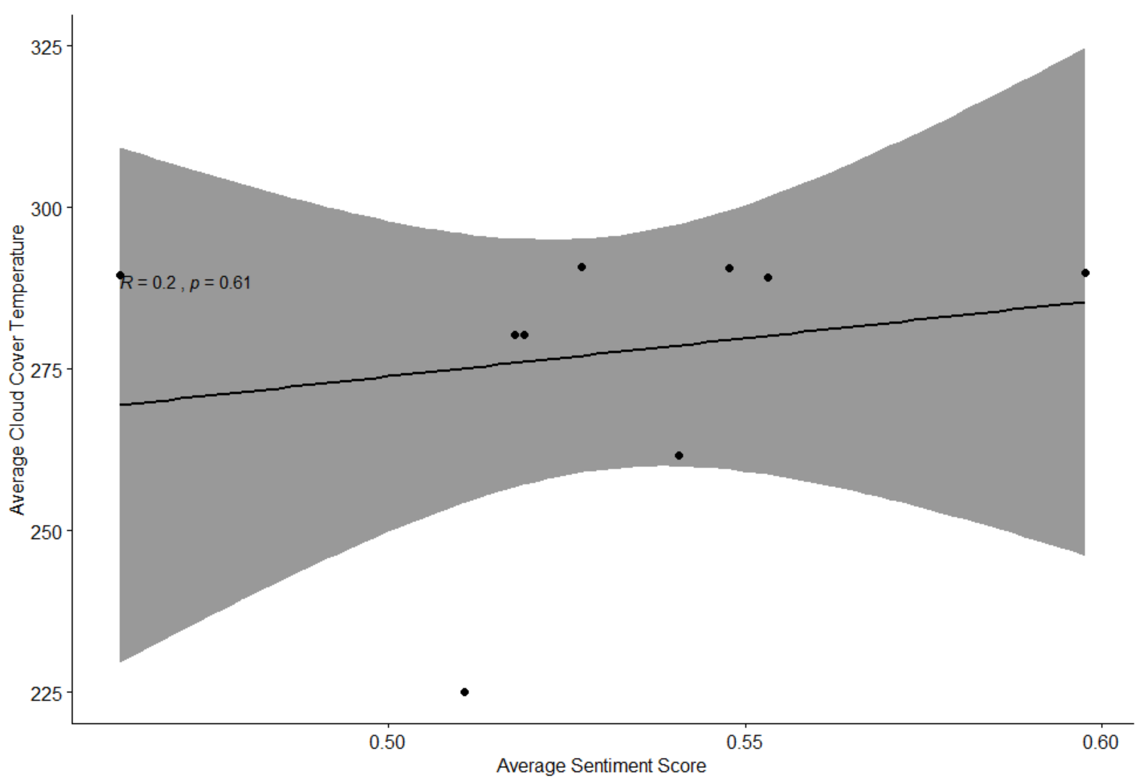


Figure 33. Q-Q Plot of Average Sentiment Score Versus Average Cloud Cover Temperature - Hurricane Michael

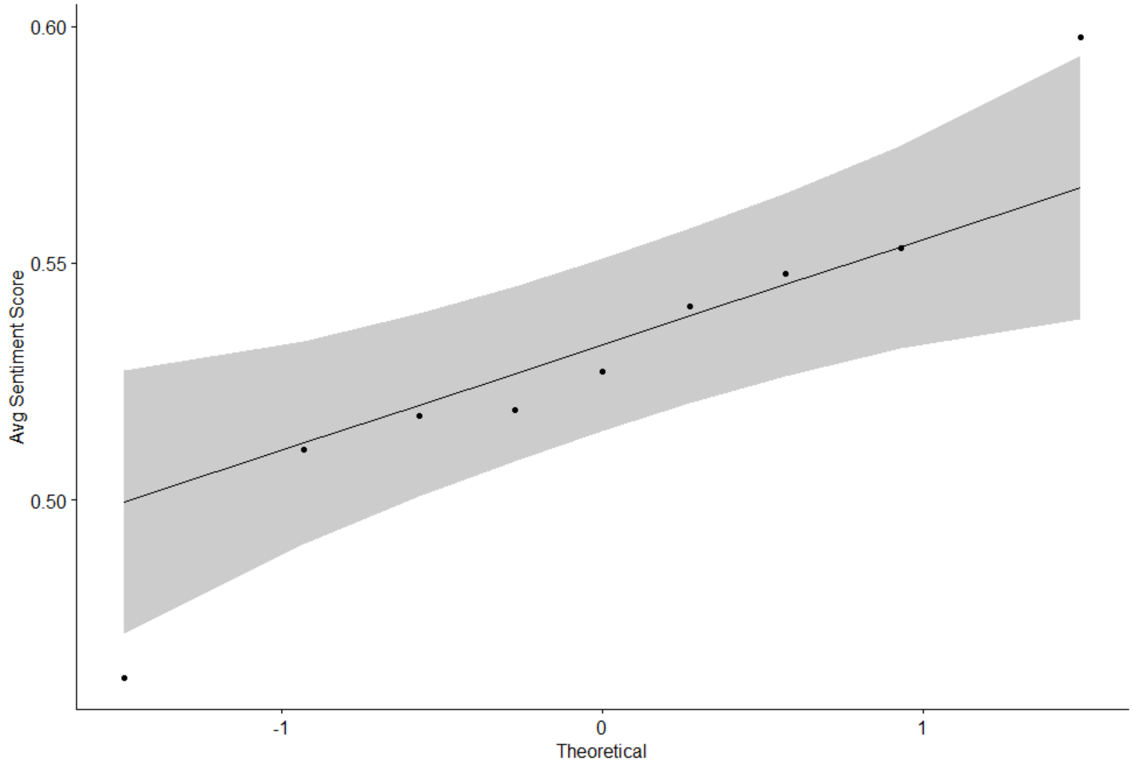


Figure 34. Q-Q Plot of Average Sentiment Scores Versus Theoretical Values - Hurricane Michael

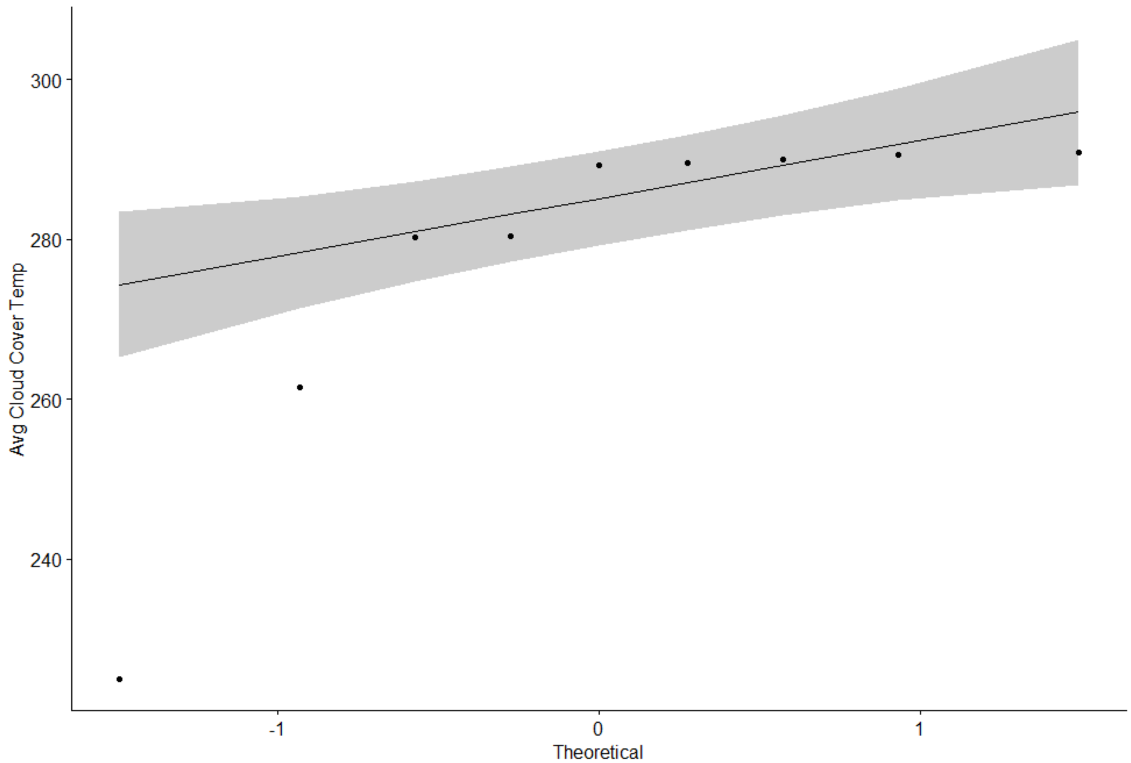


Figure 35. Q-Q Plot of Average Cloud Cover Temperatures Versus Theoretical Values - Hurricane Michael

The Pearson correlation test on the averages of sentiment scores and cloud cover temperatures revealed a p-value of 0.608, which is greater than the significance level alpha = 0.05. We conclude that the average sentiment scores and average cloud cover temperatures are not significantly correlated. The Kendall rank correlation test was used to estimate a rank-based measure of association. The correlation coefficient between the two variables was 0.2777778 and the p-value was 0.3585. This also implies that the two variables are not significantly correlated and that there may be minimal positive correlation between the variables according to the correlation coefficient. The Spearman's rank correlation coefficient test gave similar results. The p-value of 0.2912 was greater than the significance level, implying no significant correlation and the correlation coefficient of 0.4 implies that there may be a positive correlation that is not statistically significant.

All of the values for cloud cover temperature and sentiment score for each day were then analyzed using Pearson's correlation test, Kendall's correlation test, and Spearman's correlation test. The Pearson's product-moment correlation test, the Kendall rank correlation test, and the Spearman rank correlation coefficient test was conducted on sentiment score versus cloud cover temperature variables for Hurricane Michael data. When all of the data was analyzed for correlation, the Pearson's method yielded a correlation coefficient of 0.01155040. The correlation coefficient was positive, indicating that a positive correlation between the two variables may be present. The value is very close to zero, indicating that there is great variation in the data around the line of best fit. The p-value was 0.001975 and less than the significance level indicating that the correlation coefficient is statistically significant. The variables do not seem to be significantly correlated. The Kendall rank correlation test was also used on the data to estimate a rank-based measure of association. The Kendall method had a correlation coefficient of 0.008267076 and a p-value of 0.001851. There did not seem to be a significant correlation between the two variables, and the correlation coefficient was statistically significant due the value being less than that of the significance level of alpha = 0.05. The Spearman method gave similar results. The correlation coefficient was 0.01176329 and the p-value was 0.001641. The relationship between sentiment score per day and cloud cover temperature per day has an incredibly slight positive correlation, and the correlation coefficient was statistically significant.

7.3.2.1 Sentiment and Proximity Data

The distance variable was added to the dataset to identify distance between the storm and the tweet location. The Pearson, Kendall, and Spearman correlation coefficients were all calculated in R to determine is a relationship is present between the variables of distance and sentiment score. The Pearson correlation coefficient was 0.02365856. The Kendall correlation coefficient was 0.01559867. The Spearman correlation coefficient was 0.02332885.

The p-value for Pearson's correlation tests is < 6.764e-11. The p-value for Kendall's correlation test is 1.113e-10. The p-value for Spearman's correlation test is 1.236e-10. All three of these tests indicate that there is a correlation between the variables, but the correlation coefficient of each test indicates that weak correlation. The correlation results were all positive, which indicates a positive relationship between the variables. Sentiment becomes more negative as the hurricane center becomes closer to tweet location. The correlation plot of the two variables is shown in Figure 37.

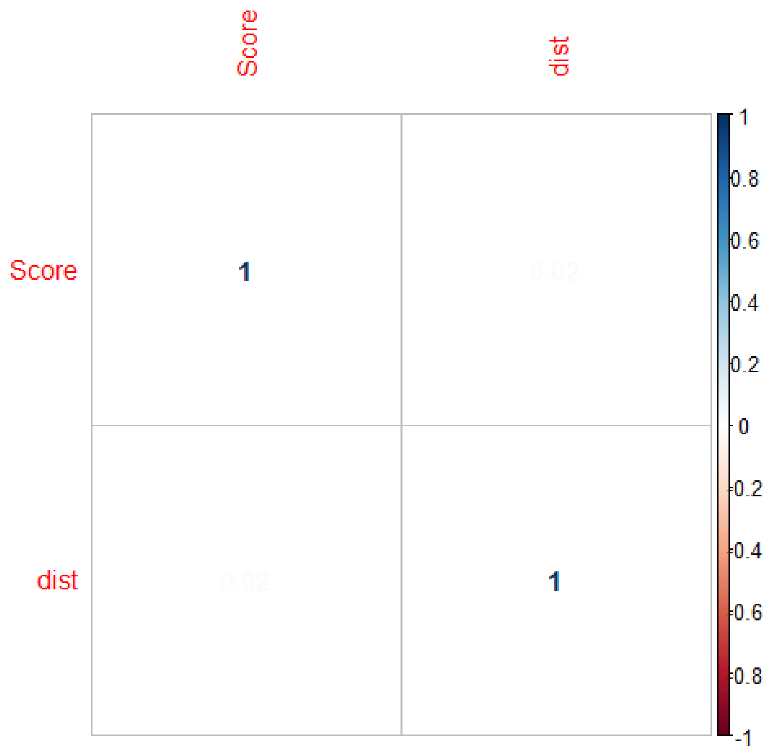


Figure 37. Correlation Plot of Sentiment Score Versus Distance of Tweet from Center of Hurricane Michael.

8.0 Conclusion

This research study has illustrated that effective sentiment analysis can be performed on a Twitter dataset. Correlation analysis did not identify correlation between the social media and cloud cover temperature physical data, resulting in a need for further research. Correlation analysis did find slight correlation between sentiment score and distance the location of the tweet was from the hurricane center. Many different data analysis tools were utilized during the course of this investigation to collect, clean and mine physical and sentiment data from the datasets. This analysis could provide valuable feedback to emergency responders and government officials to provide information before, during, and after an extreme weather event and help to make predictions for future storms using Twitter data. Discovering trends earlier will help to enhance recovery and assistance to those in need.

It is evident from this research that machine learning classifiers used in this study have an effect on the accuracy of the sentiment analysis. The algorithms used in this study are commonly used for text classification. Evaluating the different algorithms, the Naive Bayes model produced the highest accuracy of predicting sentiment for both hurricanes using these datasets. Twitter data provides a virtually unlimited source of emotions and sentiment that can be used for analysis and correlation with other data. Hurricanes Florence and Michael were two recent storms to hit the southeastern United States. Studying the sentiment and physical data from these storms can give insight into the feelings of those directly impacted by the events. It can also help to provide information to decision makers about what help is needed and where it is needed, which can then be used to predict where focus needs to be for future extreme weather events.

9.0 Works Cited

[1] Alam, F., Ofli, F., & Imran, M. (2018). A Twitter Tale of Three Hurricanes: Harvey, Irma, and Maria. arXiv preprint arXiv:1805.05144.

[2] Anon., 2011. StatSoft Electronic Statistics Textbook. [Online] Retrieved from: https://www.statsoft.com/textbook/support-vector-machines

[3] Augustyniak, Łukasz; Szymański, Piotr; Kajdanowicz, Tomasz; Tuligłowicz, Włodzimierz (2015). "Comprehensive Study on Lexicon-based Ensemble Classification Sentiment Analysis". Entropy. 18 (1): 4.

[4] Barnes, M. D., Hanson, C. L., Novilla, L. M., Meacham, A. T., McIntyre, E., & Erickson, B. C. (2008). Analysis of media agenda setting during and after Hurricane Katrina: Implications for emergency preparedness, disaster response, and disaster policy. American journal of public health, 98(4), 604-610.

[5] Baylis, P., Obradovich, N., Kryvasheyeu, Y., Chen, H., Coviello, L., Moro, E., ... & Fowler, J. H. (2018). Weather impacts expressed sentiment. PloS one, 13(4), e0195750.

[6] Biau, G., & Scornet, E. (2016). A random forest guided tour. Test, 25(2), 197-227.

[7] Bik, H. M., & Goldstein, M. C. (2013). An introduction to social media for scientists. PLoS biology, 11(4), e1001535.

[8] Blei, David M. (2012). "Introduction to Probabilistic Topic Models". Comm. ACM. 55 (4): 77–84.

[9] Blei, David M., Ng, Andrew Y., Jordan, Michael I., Lafferty, John, ed. (2003) "Latent Dirichlet Allocation". Journal of Machine Learning Research. 3 (4–5): pp. 993–1022.

[10] Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013). "New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems. 28 (2): 15–21.

[11] Caragea, C., Squicciarini, A. C., Stehle, S., Neppalli, K., & Tapia, A. H. (2014, May). Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. In ISCRAM.

[12] Cerutti, V., Fuchs, G., Andrienko, G., Andrienko, N., & Ostermann, F. (2016). Identification of disaster-affected areas using exploratory visual analysis of georeferenced Tweets: application to a flood event. Association of Geographic Information Laboratories in Europe: Helsinki, Finland, 5.

[13] Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., & Ertl, T. (2012, October). Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on(pp. 143-152). IEEE.

[14] Cheong, F., & Cheong, C. (2011). Social Media Data Mining: A Social Network Analysis Of Tweets During The 2010-2011 Australian Floods. PACIS, 11, 46-46.

[15] D, Aju & Rajkumar, Rajasekaran. (2016). A Multi-Stage Hybrid CAD Approach for MRI Brain Tumor Recognition and Classification. IIOAB Journal. 7. 7-20.

[16] Dalinina, Ruslana (2017). Introduction to Correlation. Oracle + DataScience.com. Retrieved from https://www.datascience.com/blog/introduction-to-correlation-learn-data-science-tutorials

[17] David, C. C., Ong, J. C., & Legara, E. F. T. (2016). Tweeting Supertyphoon Haiyan: Evolving functions of Twitter during and after a disaster event. PloS one, 11(3), e0150190.

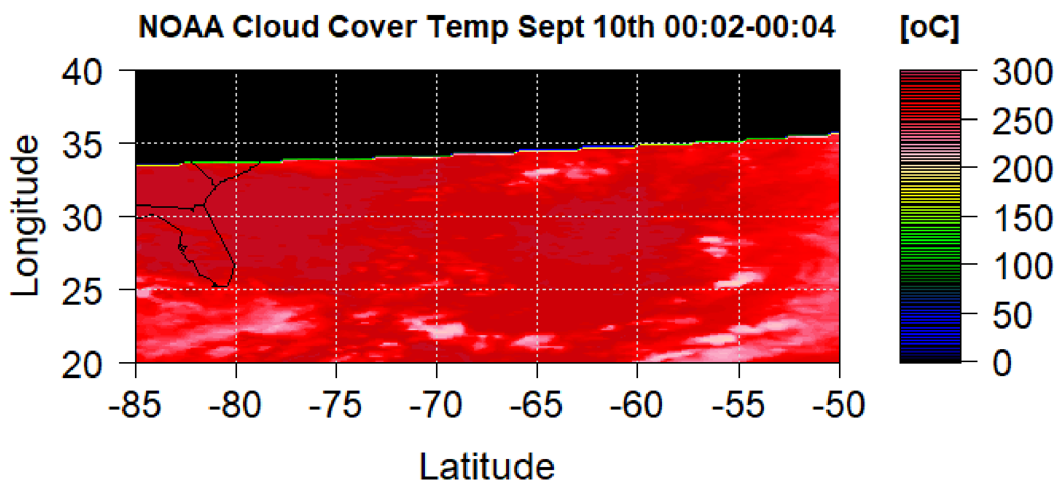
[18] Diakopoulos, N. A., & Shamma, D. A. (2010). Characterizing debate performance via aggregated twitter sentiment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1195-1198). ACM.
[19] Dodd, J. (2014). Twitter sentiment analysis (Doctoral dissertation, Dublin, National College of Ireland).
[20] Enenkel, M., Saenz, S. M., Dookie, D. S., Braman, L., Obradovich, N., & Kryvasheyeu, Y. (2018). Social Media Data Analysis and Feedback for Advanced Disaster Risk Management. arXiv preprint arXiv:1802.02631.
[21] Feaster, T. D., Weaver, J. C., Gotvald, A. J., & Kolb, K. R. (2018). Preliminary peak stage and streamflow data for selected US Geological Survey streamgaging stations in North and South Carolina for flooding following Hurricane Florence, September 2018 (No. 2018-1172). US Geological Survey.
[22] Gadidov, B., Le, L. (2018). A Case Study of Mining Social Media Data for disaster Relief: Hurricane Irma. Retrieved from https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/2695-2018.pdf
[23] Google Trends (2018) Retrieved from https://trends.google.com/trends/explore
[24] Gorakala, S. K. (2013, October 6). Topic Modeling in R. Retrieved from https://www.r-bloggers.com/topic-modeling-in-r-2/
[25] Gruebner, O., Lowe, S., Sykora, M., Shankardass, K., Subramanian, S. V., & Galea, S. (2018). Spatio-temporal distribution of negative emotions in New York City after a natural disaster as seen in social media. International journal of environmental research and public health, 15(10), 2275.
[26] Guillot, Dominique (2017). Mathematical Techniques in Data Science: Random forest. Retrieved from http://www.math.udel.edu/~dguillot/teaching/MATH567-S2017/lectures-handout/MATH567-Week12-handout-Random_forest.pdf
[27] Hassan, A., Abbasi, A., & Zeng, D. (2013). Twitter sentiment analysis: A bootstrap ensemble framework. In 2013 International Conference on Social Computing (pp. 357-364). IEEE.
[28] Hossain, A. (2018, May). Text analysis on the tweets about Bangladesh. Retrieved from https://rstudio-pubs-static.s3.amazonaws.com/400502_639fc3d0594d4d36891dcdfd739124f3.html
[29] HURDAT (2019). Atlantic hurricane best track (HURDAT version 2). Hurricane Research Division (Database). National Hurricane Center. Retrieved February 26, 2019.
[30] Insurance Information Institute (2018). Facts + Statistics: Hurricanes. Retrieved from https://www.iii.org/fact-statistic/facts-statistics-hurricanes
[31] Insurance Information Institute (2018). Facts + Statistics: Hurricanes. Retrieved from https://www.iii.org/fact-statistic/facts-statistics-hurricanes
[32] Jockers, Matthew (2017). Syuzhet. Retrieved from https://github.com/mjockers/syuzhet
[33] Khan, Riaz (2017). Naive Bayes Classifier: theory and R example. RPubs. Retrieved from https://rpubs.com/riazakhan94/naive_bayes_classifier_e1071
[34] King, L. J. (2018). Social Media Use During Natural Disasters: An Analysis of Social Media Usage During Hurricanes Harvey.

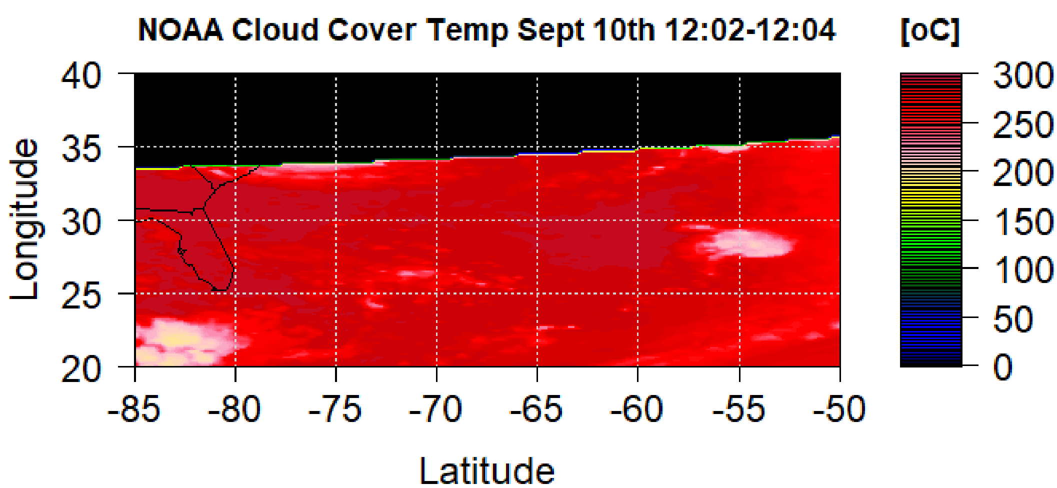
[35] Koppel, Moshe; Schler, Jonathan (2006). The Importance of Neutral Examples for Learning Sentiment. Computational Intelligence 22. pp. 100–109.
[36] Krishnan, A., Adivi, S., M.B., Mishra, Nitesh. (2016) Train machine to classify tweets. RPubs - Tweet Classification.
[37] Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. Science advances, 2(3), e1500779.
[38] Kumar, S., Barbier, G., Abbasi, M. A., & Liu, H. (2011, July). Tweettracker: An analysis tool for humanitarian and disaster relief. In Fifth international AAAI conference on weblogs and social media.
[39] Lam. (2013). Hashtagging Sandy: How are the psychological effects of natural disasters expressed in social media? Analyzing the response to Hurricane Sandy on Twitter. Retrieved from https://library.gwu.edu/sites/default/files/eckles/Hashtagging%20Sandy%20-%20How%20are%20the%20psychological%20effects%20of%20natural%20disasters%20expressed%20in%20social%20media.docx.pdf
[40] Maron, D. F. (2013). How social media is changing disaster response. Scientific American, 7.
[41] Martín, Y., Li, Z., & Cutter, S. L. (2017). Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. PLoS one, 12(7), e0181701.
[42] Martinez, M. (2018). Assessing the Impact of Hurricane Harvey Through Twitter Data. Rice University Houston Education Research Consortium.
[43] The National Hurricane Center (2019). NHC GIS Archive - Tropical Cyclone Best Track. Retrieved from https://www.nhc.noaa.gov/gis/archive_besttrack.php?year=2018
[44] Nazer, T. H., Xue, G., Ji, Y., & Liu, H. (2017). Intelligent Disaster Response via Social Media Analysis A Survey. ACM SIGKDD Explorations Newsletter, 19(1), 46-59.
[45] Neppalli, V. K., Medeiros, M. C., Caragea, C., Caragea, D., Tapia, A. H., & Halse, S. E. (2016, January). Retweetability Analysis and Prediction during Hurricane Sandy. In ISCRAM.
[46] Neto, J. P. (2015). Maximum Entropy. Retrieved from http://www.di.fc.ul.pt/~jpn/r/maxent/maxent.html
[47] NOAA (2018). Billion-Dollar Weather and Climate Disasters: Overview. Retrieved from https://www.ncdc.noaa.gov/billions/overview
[48] O'Leary, D. E. (2015). Twitter mining for discovery, prediction and causality: Applications and methodologies. Intelligent Systems in Accounting, Finance and Management, 22(3), 227-247.
[49] Sadri, A. M., Ukkusuri, S. V., & Gladwin, H. (2017). The role of social networks and information sources on hurricane evacuation decision making. Natural Hazards Review, 18(3), 04017005.
[50] Saif, H., Fernandez, M., He, Y., & Alani, H. (2013). Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold.

[51] Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data’. Geoforum, 52, 167-179.
[52] Shen, S.S.P., 2017: R Programming for Climate Data Analysis and Visualization: Computing and plotting for NOAA data applications. The first revised edition. San Diego State University, San Diego, USA, 152pp.
[53] Smith, A. B., & Katz, R. W. (2013). US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. Natural hazards, 67(2), 387-410.
[54] Soni, R., & Mathai, K. J. (2016). Effective sentiment analysis of a launched product using clustering and decision trees. International Journal of Innovative Research in Computer and Communication Engineering, 4(1).
[55] Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. Handbook of latent semantic analysis, 427(7), 424-440.
[56] Stowe, K., Anderson, J., Palmer, M., Palen, L., & Anderson, K. (2018). Improving classification of twitter behavior during hurricane events. In Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media(pp. 67-75).
[57] Tewari, A., Sawant, P., Samtani, J., Sawant, S., & Massand, G. (2017). Multilabel Classification of Tweets. International Journal of Computer Applications, 159(1).
[58] Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios; Cai, Di; Kappas, Arvid (2010). "Sentiment strength detection in short informal text". Journal of the American Society for Information Science and Technology. 61 (12): 2544–2558.
[59] Vayansky, I. R. (2018). An Evaluation of Geotagged Twitter Data during Hurricane Irma using Sentiment Analysis and Topic Modeling for Disaster Resilience.
[60] Villegas, C., Martinez, M., & Krause, M. (2018). Lessons from Harvey: Crisis Informatics for Urban Resilience.
[61] Virtual Social Media Working Group. (2013). Lessons learned: Social media and hurricane Sandy.
[62] Yang, J., Yu, M., Qin, H., Lu, M., & Yang, C. (2019). A Twitter Data Credibility Framework—Hurricane Harvey as a Use Case. ISPRS International Journal of Geo-Information, 8(3), 111.
[63] Yu, S., Li, P., Lin, H., Rohani, E., Choi, G., Shao, B., & Wang, Q. (2013). Support vector machine based detection of drowsiness using minimum EEG features. In 2013 International Conference on Social Computing (pp. 827-835). IEEE.
[64] Zhao, Y. (2016). Twitter Data Analysis with R–Text Mining and Social Network Analysis. RDM, Canberra.
[65] Boehmke, B. (2018). Gradient Boosting Machines. UC Business Analytics R Programming Guide. Retrieved from http://uc-r.github.io/gbm_regression

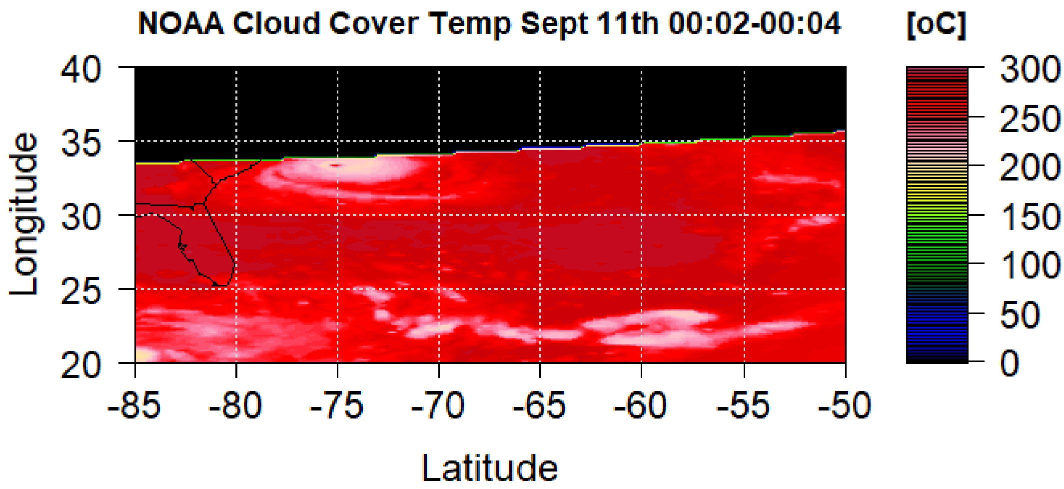
Appendix A

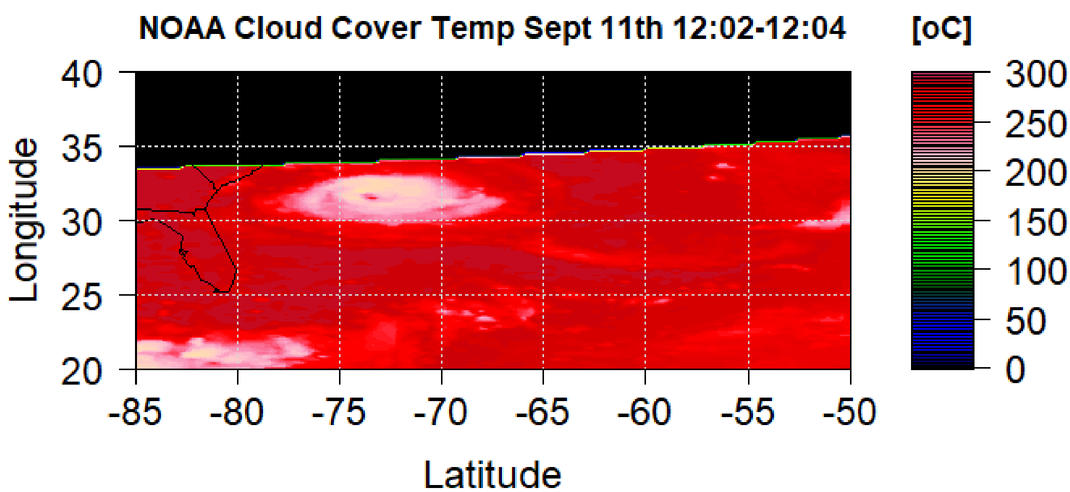
Cloud Cover Temp - Sep 10 - Sep 15

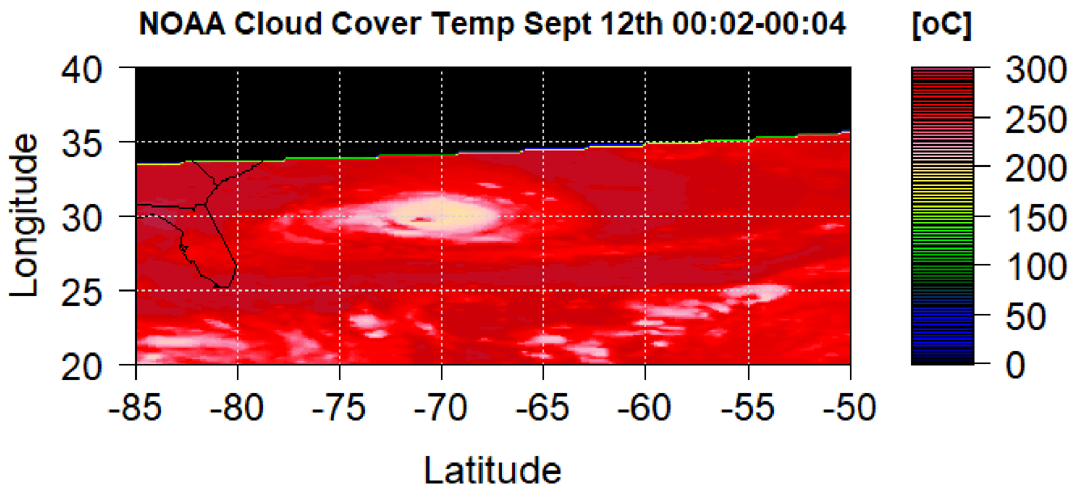


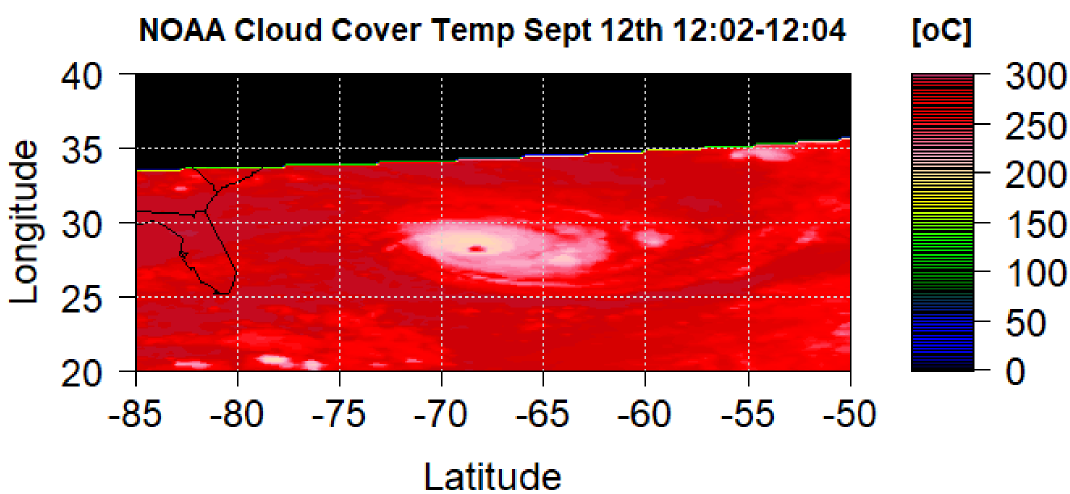


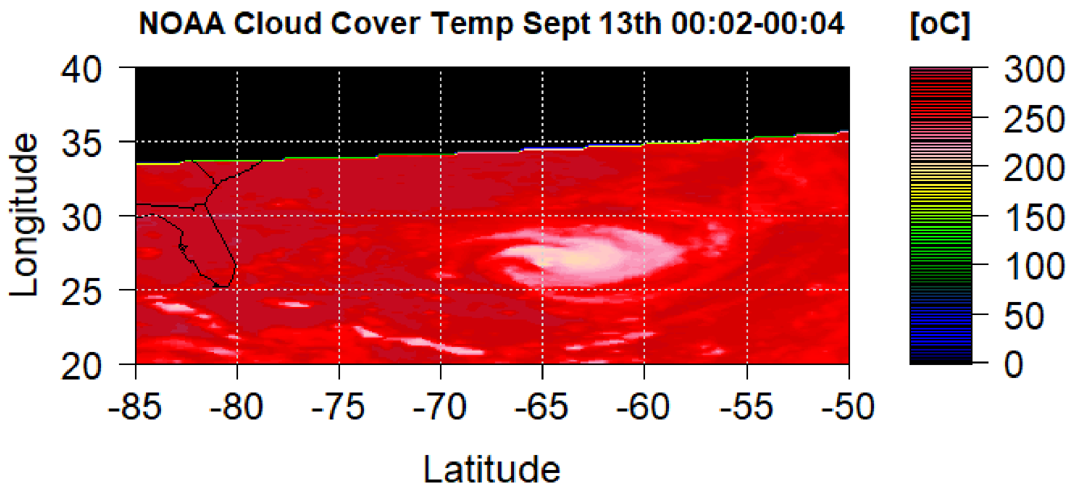
\

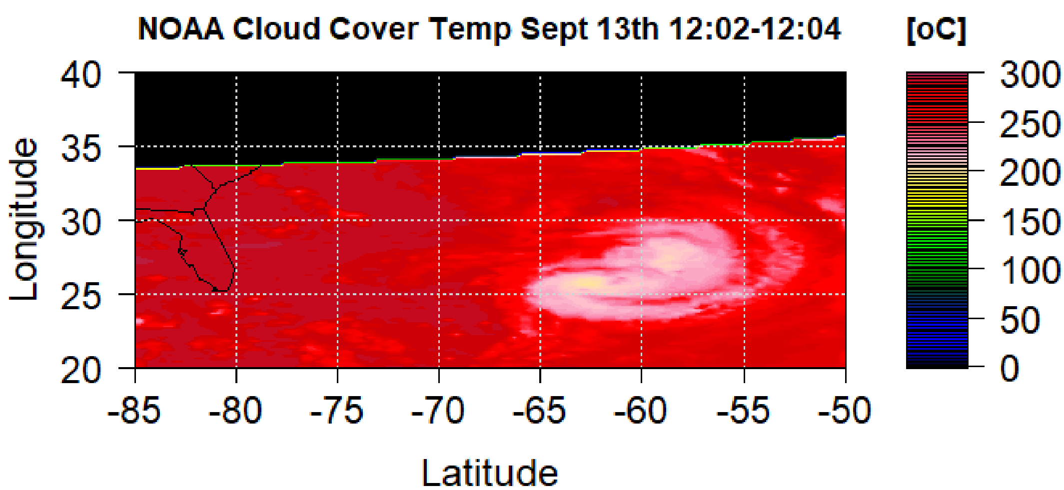

NOAA Cloud Cover Temp Sept 11th 00:02-00:04
[oC]
300
250
200
150
100
50
0
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude

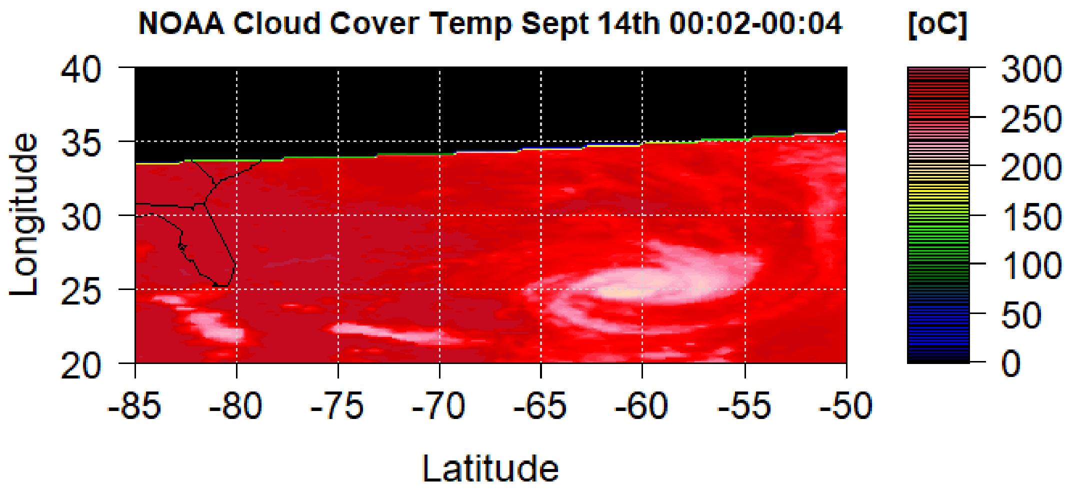

NOAA Cloud Cover Temp Sept 11th 12:02-12:04
[oC]
300
250
200
150
100
50
0
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude

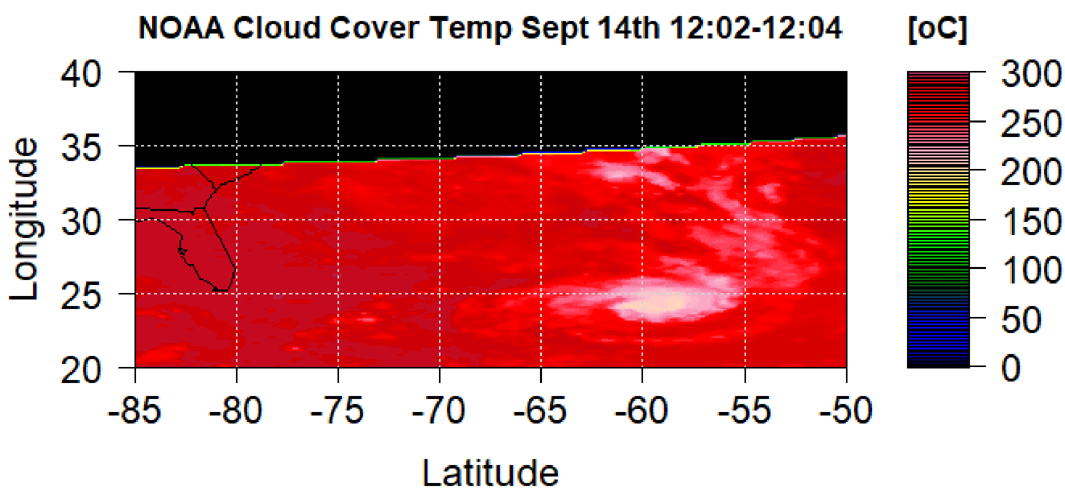

NOAA Cloud Cover Temp Sept 12th 00:02-00:04
[oC]
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude
300
250
200
150
100
50
0

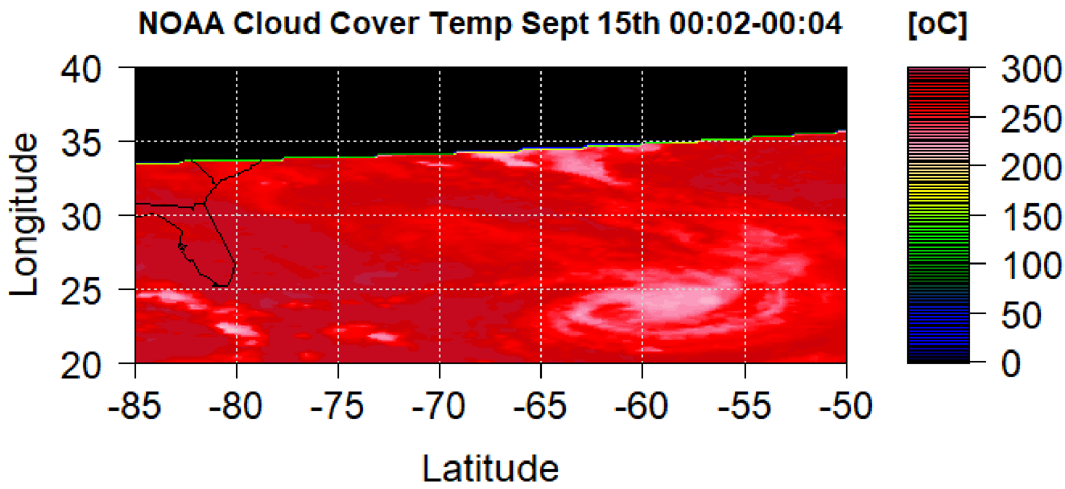

NOAA Cloud Cover Temp Sept 12th 12:02-12:04
[oC]
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude
300
250
200
150
100
50
0

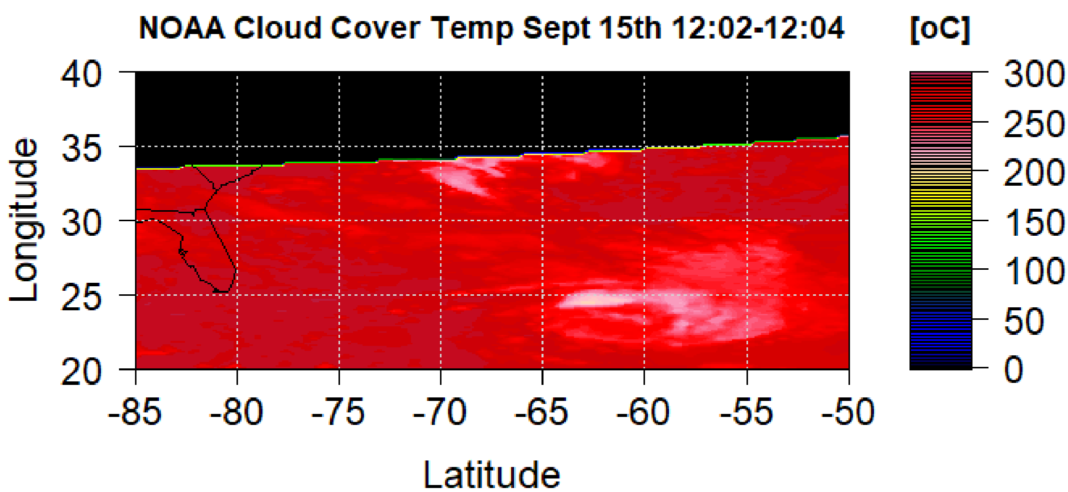

NOAA Cloud Cover Temp Sept 13th 00:02-00:04
[oC]
300
250
200
150
100
50
0
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude


NOAA Cloud Cover Temp Sept 13th 12:02-12:04
[oC]
300
250
200
150
100
50
0
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude


NOAA Cloud Cover Temp Sept 14th 00:02-00:04
[oC]
300
250
200
150
100
50
0
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude

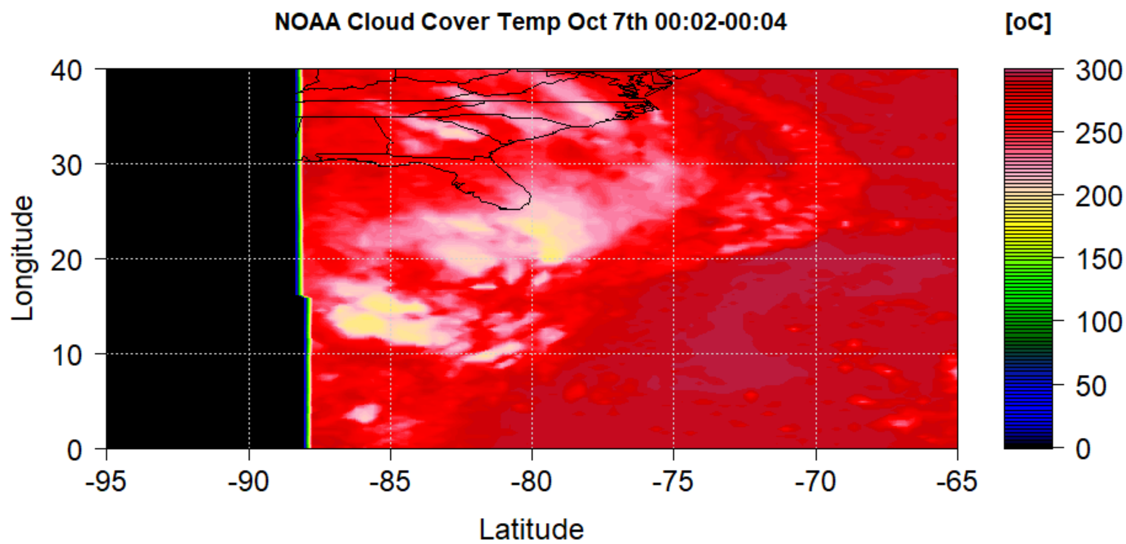

NOAA Cloud Cover Temp Sept 14th 12:02-12:04
[oC]
300
250
200
150
100
50
0
40
35
30
25
20
Longitude
-85
-80
-75
-70
-65
-60
-55
-50
Latitude

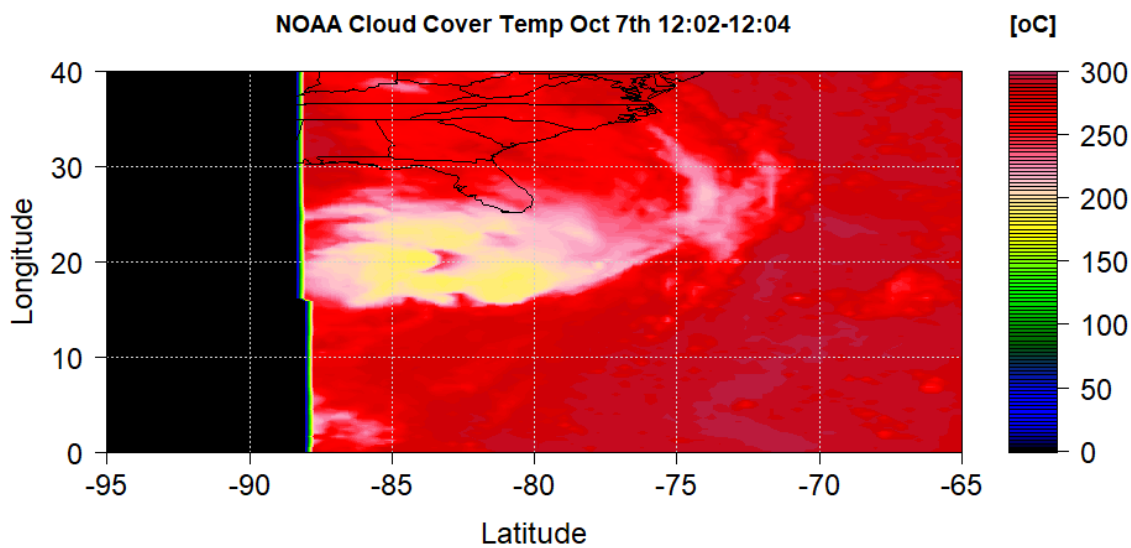

NOAA Cloud Cover Temp Sept 15th 00:02-00:04
[oC]
300
250
200
150
100
50
0
Longitude
40
35
30
25
20
-85
-80
-75
-70
-65
-60
-55
-50
Latitude

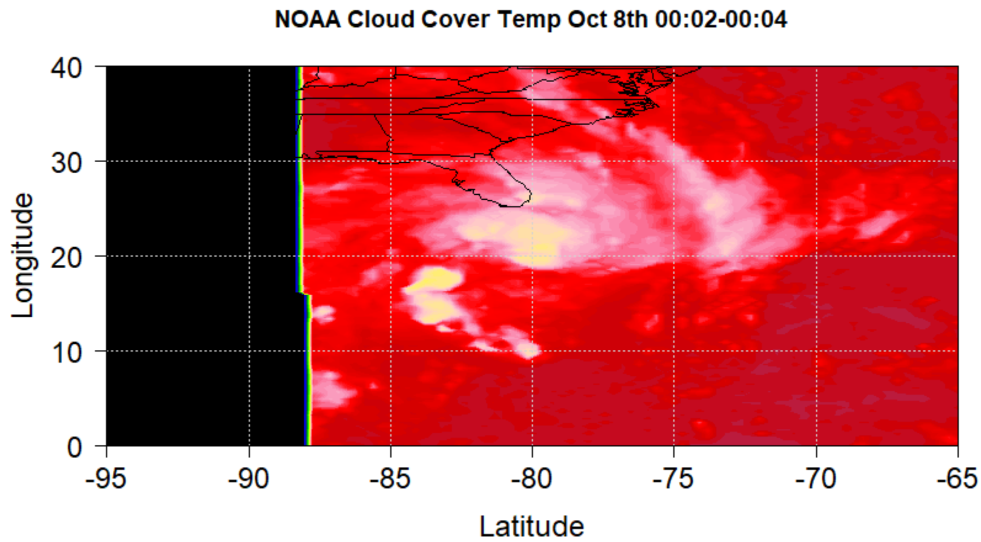

NOAA Cloud Cover Temp Sept 15th 12:02-12:04
[oC]
300
250
200
150
100
50
0
Longitude
40
35
30
25
20
-85
-80
-75
-70
-65
-60
-55
-50
Latitude

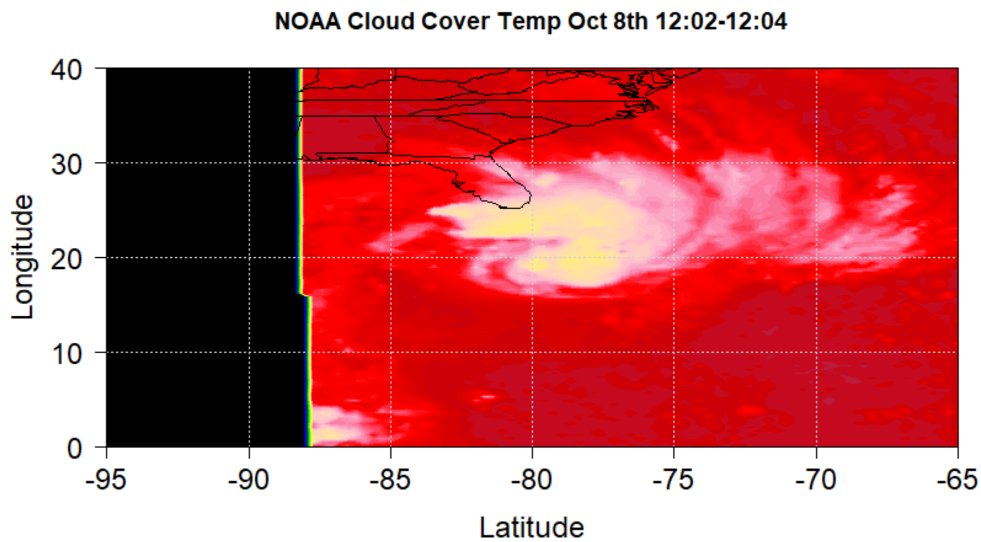
Appendix B

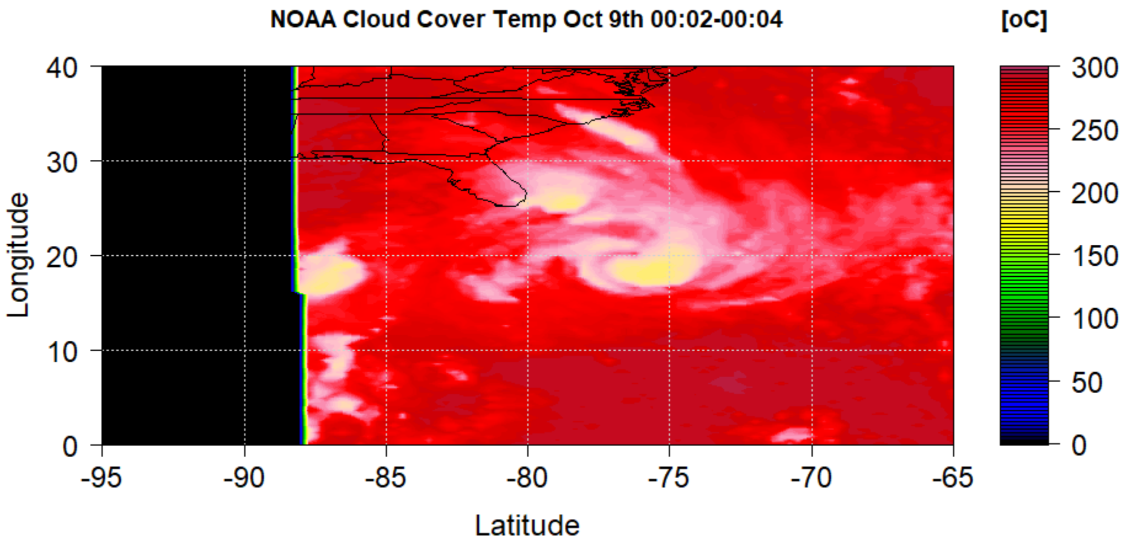
Cloud Cover Temp - Oct 8 - Oct 15

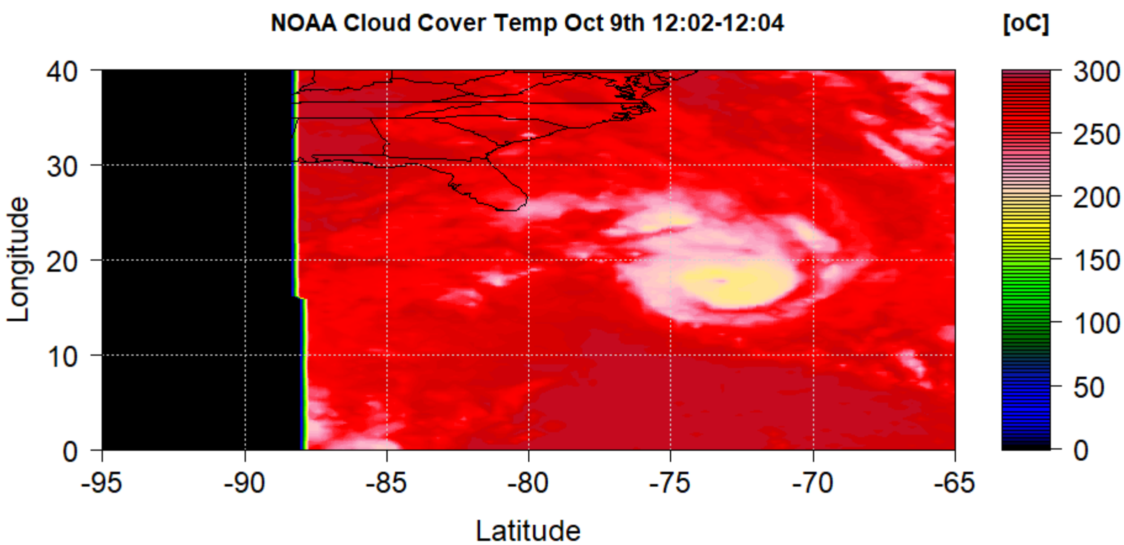


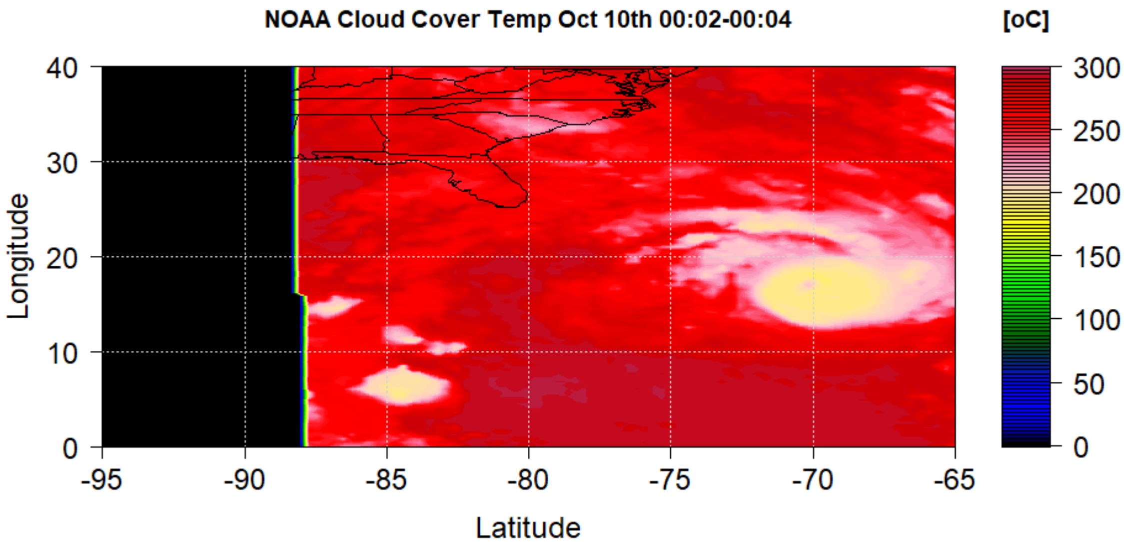


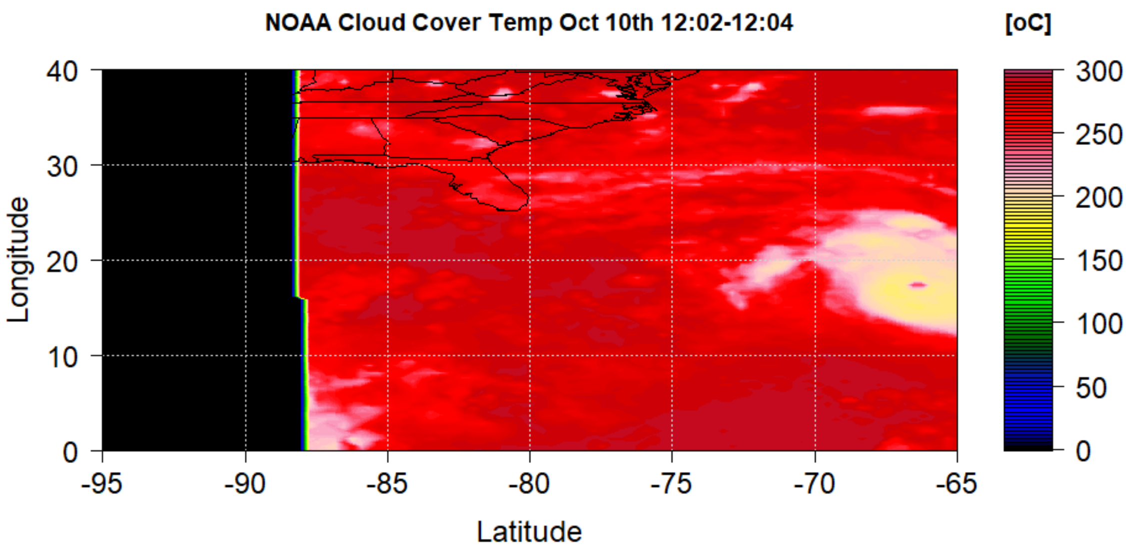

NOAA Cloud Cover Temp Oct 8th 00:02-00:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

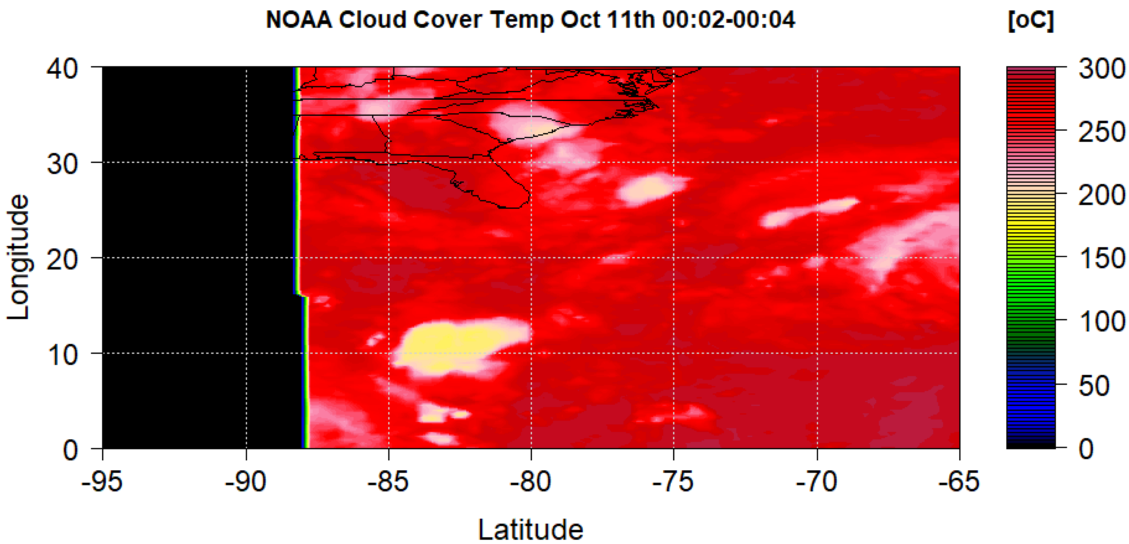

NOAA Cloud Cover Temp Oct 8th 12:02-12:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

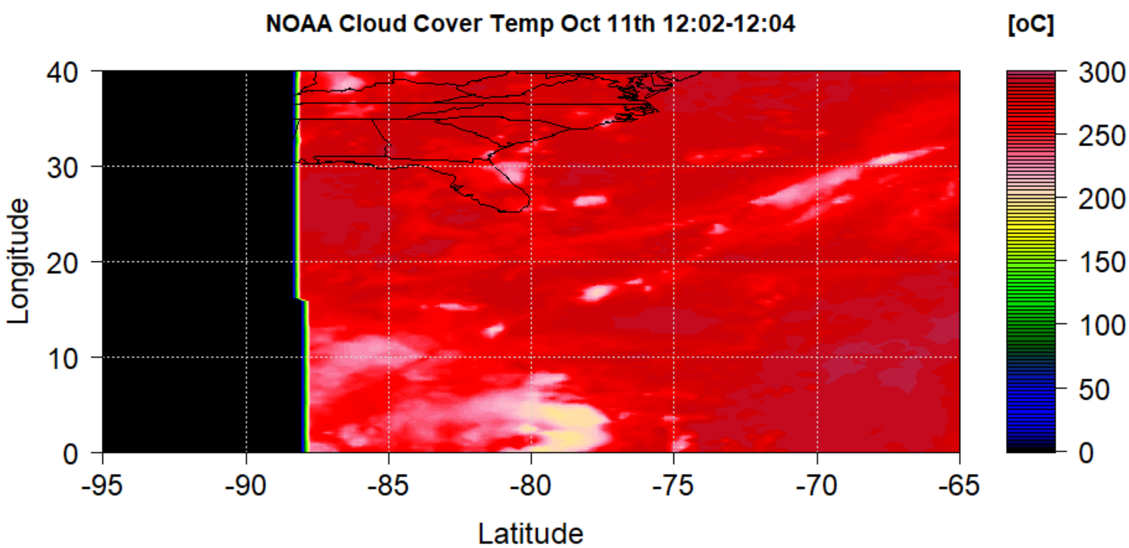

NOAA Cloud Cover Temp Oct 9th 00:02-00:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

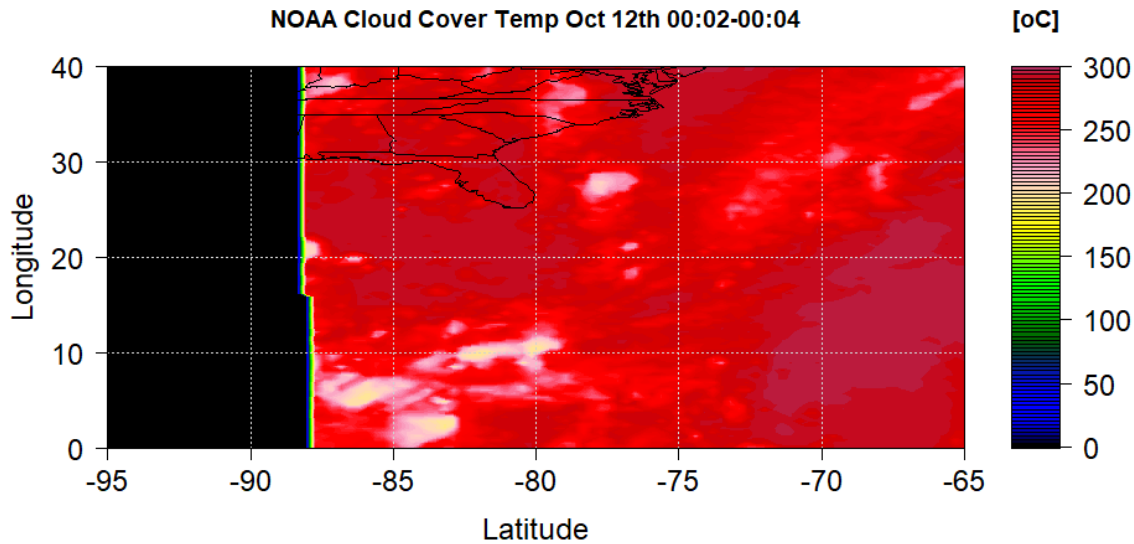

NOAA Cloud Cover Temp Oct 9th 12:02-12:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

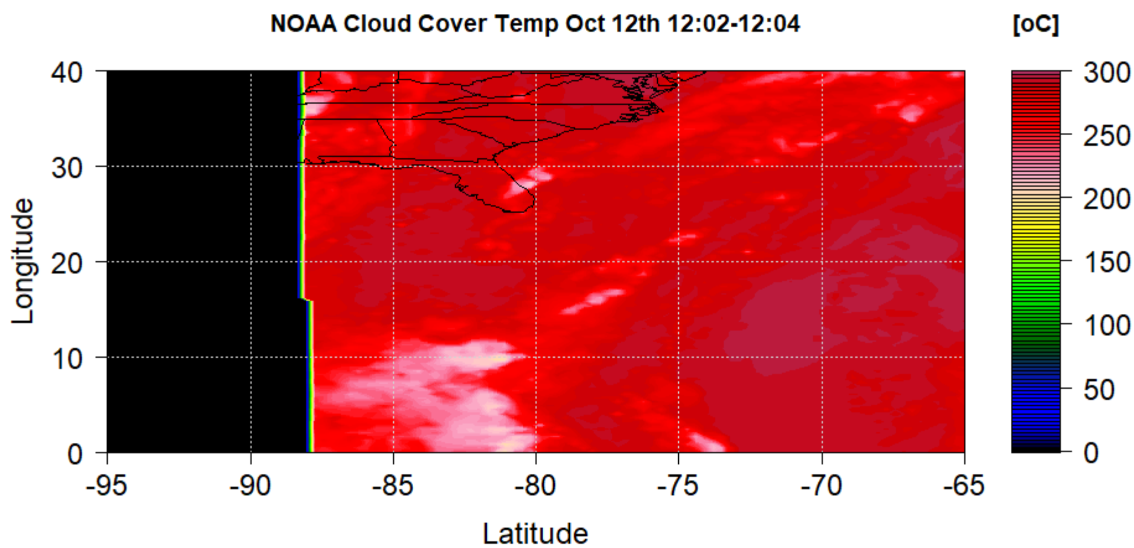

NOAA Cloud Cover Temp Oct 10th 00:02-00:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

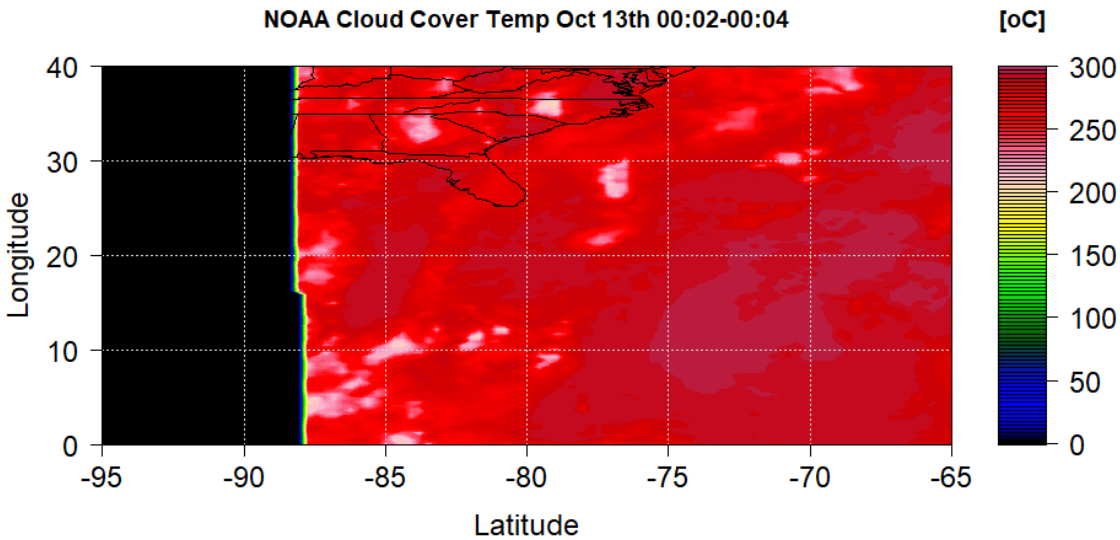

NOAA Cloud Cover Temp Oct 10th 12:02-12:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

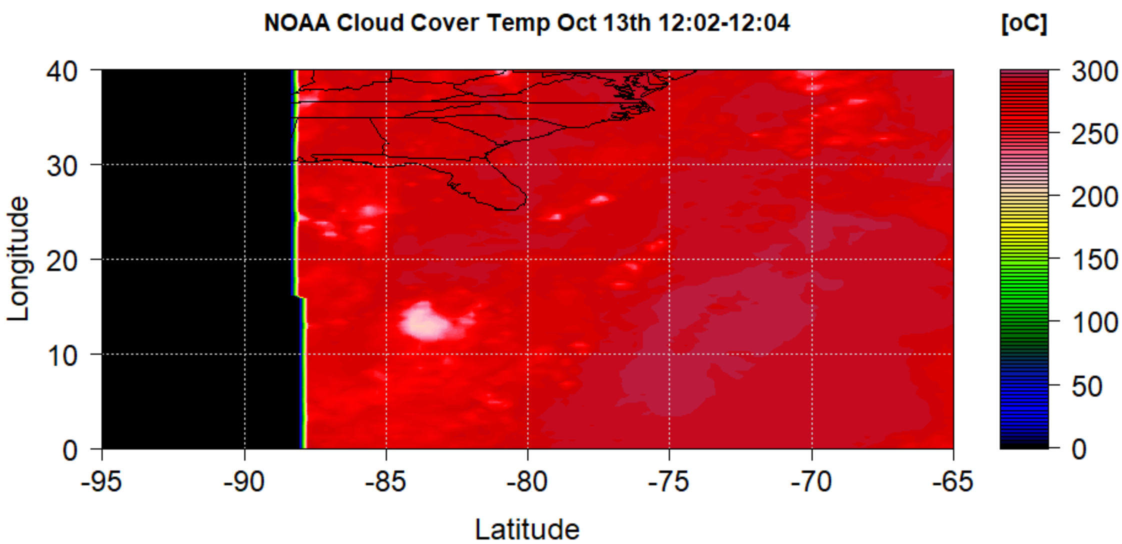

NOAA Cloud Cover Temp Oct 11th 00:02-00:04
[oC]
40
30
20
10
0
Longitude
-95
-90
-85
-80
-75
-70
-65
Latitude
300
250
200
150
100
50
0

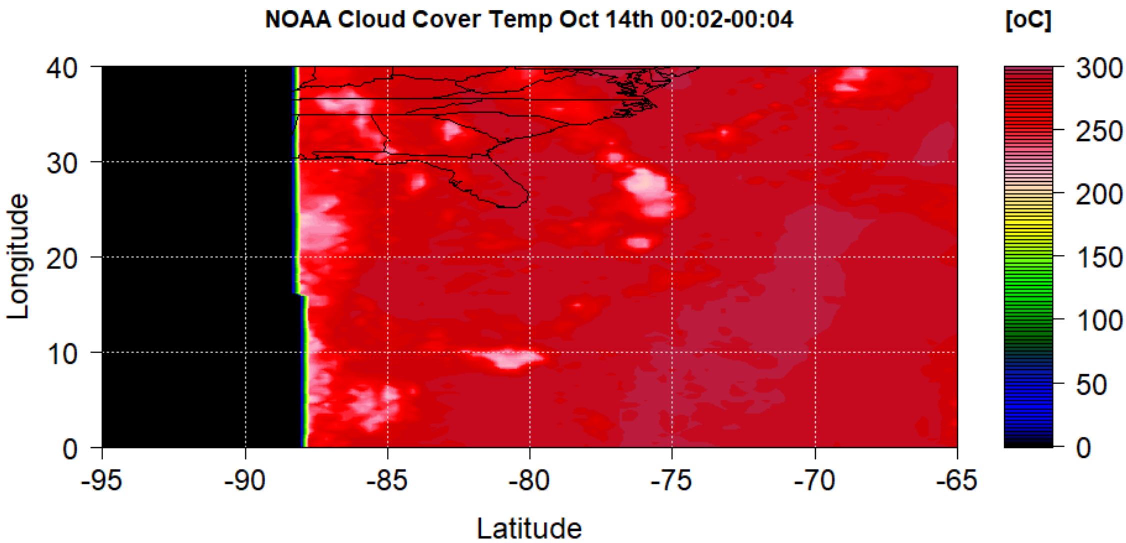

NOAA Cloud Cover Temp Oct 11th 12:02-12:04
[oC]
40
30
20
10
0
Longitude
-95
-90
-85
-80
-75
-70
-65
Latitude
300
250
200
150
100
50
0


NOAA Cloud Cover Temp Oct 12th 00:02-00:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

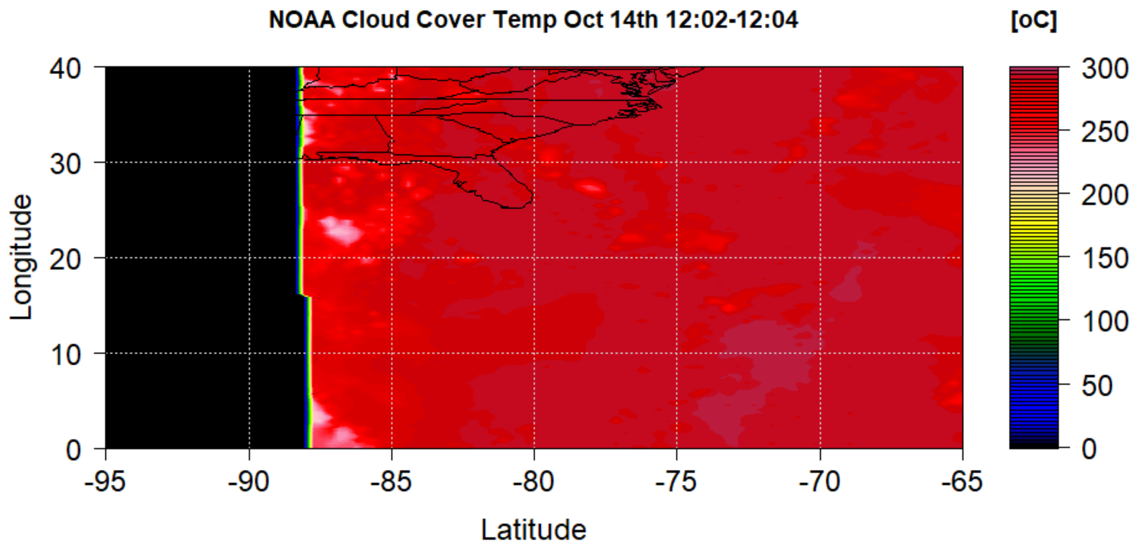

NOAA Cloud Cover Temp Oct 12th 12:02-12:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0

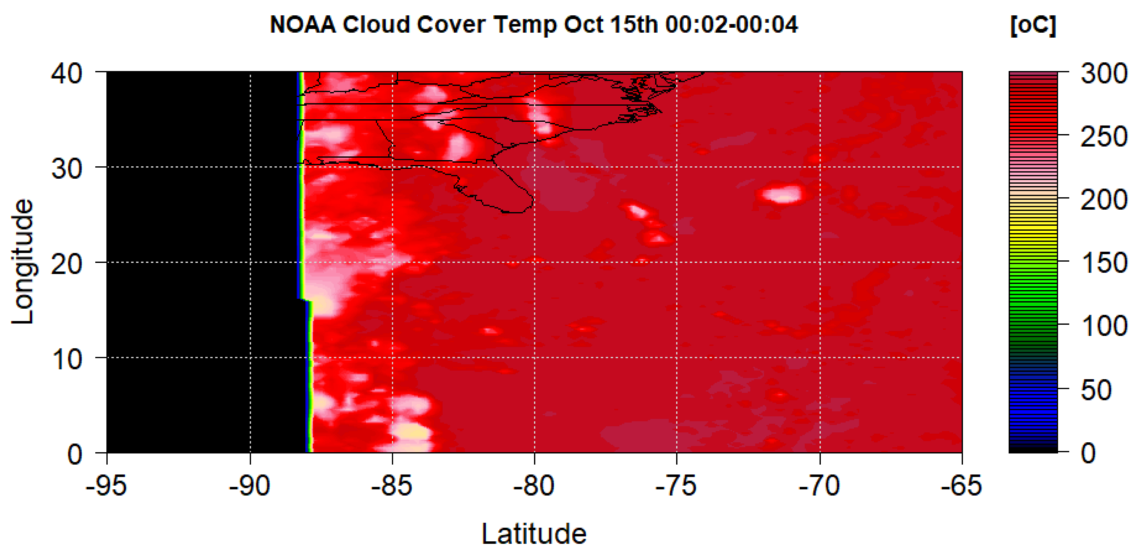

NOAA Cloud Cover Temp Oct 13th 00:02-00:04
[oC]
300
250
200
150
100
50
0
40
30
20
10
0
Longitude
-95
-90
-85
-80
-75
-70
-65
Latitude

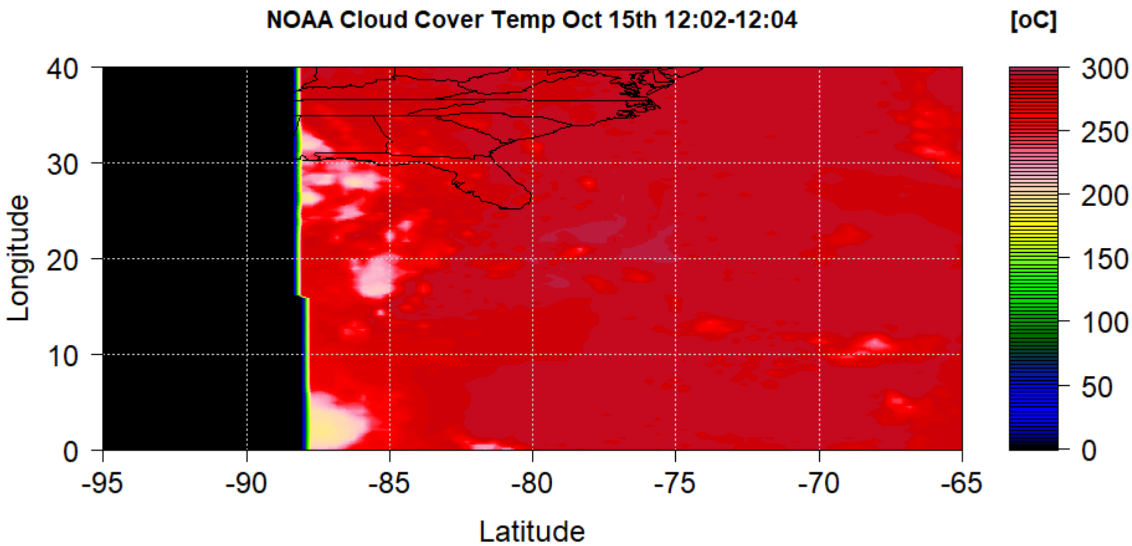

NOAA Cloud Cover Temp Oct 13th 12:02-12:04
[oC]
300
250
200
150
100
50
0
40
30
20
10
0
Longitude
-95
-90
-85
-80
-75
-70
-65
Latitude



NOAA Cloud Cover Temp Oct 14th 00:02-00:04
[oC]
300
250
200
150
100
50
0
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
Longitude
Latitude


NOAA Cloud Cover Temp Oct 14th 12:02-12:04
[oC]
300
250
200
150
100
50
0
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
Longitude
Latitude


NOAA Cloud Cover Temp Oct 15th 00:02-00:04
[oC]
300
250
200
150
100
50
0
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
Longitude
Latitude


NOAA Cloud Cover Temp Oct 15th 12:02-12:04
[oC]
Longitude
Latitude
40
30
20
10
0
-95
-90
-85
-80
-75
-70
-65
300
250
200
150
100
50
0