Juracán: Analyzing Hurricane Trajectories and Assessing Hurricane Risks for Gulf of Mexico

COSC 3337 Task 3, Fall 2024

Gabriel Becker
Computer Science Department
University of Houston

Andrew Guzman
Computer Science Department
University of Houston

Hannah Jensen
Computer Science Department
University of Houston

Shadha Khan
Computer Science Department
University of Houston

Kyle Rodriguez
Computer Science Department
University of Houston

DATA COLLECTION AND PREPARATION

Lead Contributor: Kyle Rodriguez

The data we utilized for this project was the hurricane database HURDAT, which was obtained using the file *get_data.ipynb* from the National Oceanic and Atmospheric Administration (NOAA).

STORM TRACK VISUALIZATION

Lead Contributor: Shadha Khan

For the storm track visualization in the relevant file *storm_track_visualized.py*, the Gulf of Mexico is visualized from latitude 10 degrees to 35 degrees, and longitude -100 degrees to -70 degrees. The storm data displayed in the visualization is pulled from the Python pickle file *detailed_storm_data.pkl*.

The time frame is measured over 25 years, from now, 2024 to 1999, and only the storms relevant to the time frame are presented in the visualization. The states, their borders, and the coastlines are presented in the visualization as well for clarity. Because of the large number of storm tracks presented in the final plot, the alpha (or opacity) of the lines was set to 0.5 rather than 1, in order to make sure that the coastlines and states can still be seen when all storms are visualized.

There are sliders for setting the range of years, the minimum year being 1999 and the maximum year being 2024, and only a valid range of years can be displayed (ex. The user cannot set the maximum year lower than the minimum year), and at least 1 year of storm tracks are always displayed. The user can toggle the storm names beside each displayed storm track, although for a large range of years, it becomes difficult to read the relevant storm name. The user can also toggle the city names, which displays the 23 cities that are within the given latitude and longitude range.

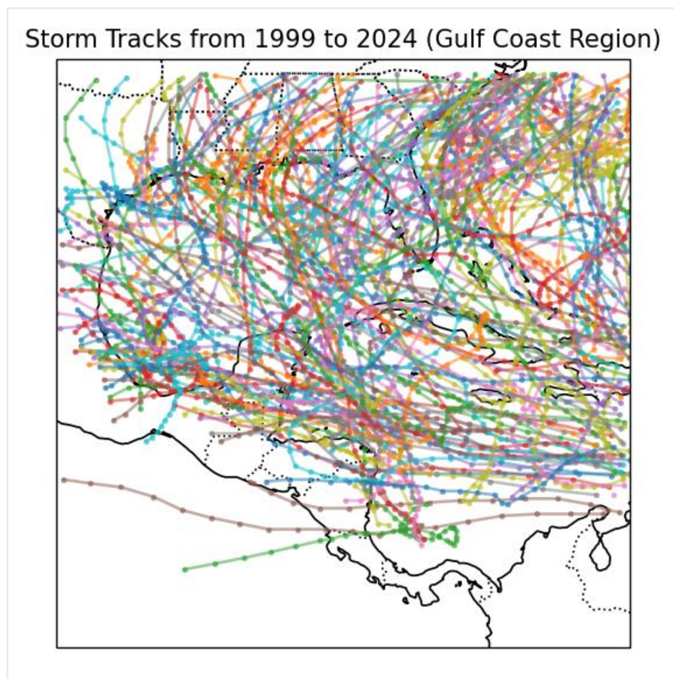

Figure 1: Visualization of all storm tracks over the full 25 years, without the storm names or cities included in the plot. The visual on the right is the visualization of the storm tracks in 2024, with the storm names and the city names included in the plot.

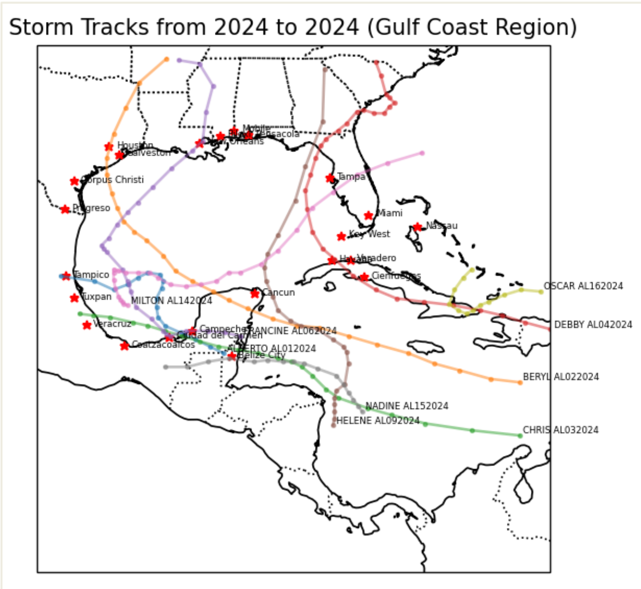

Figure 2: Visualization of the storm tracks in 2024, with the storm names and the city names included in the plot.

ANALYSIS OF HURRICANE/TROPICAL STORM TRACKS

Lead Contributor: Gabriel Becker

Preliminary Analysis

For the primary analysis of our hurricane data, our group took the data of the last twenty-five years, from 1999 to 2024 and plotted the hurricane routes over the Gulf of Mexico. We're going to be focusing on the data that was collected for our visualization, which is focused on the Gulf of Mexico and a part of the Caribbean Sea, or the area between latitude five and thirty-five and the longitude negative one hundred and negative seventy. With this data we will be looking at where they were located, their direction, their duration, and the average intensity while in our recording area. With this we will be breaking down our data by year to help us identify if there are any major trends we can notice, or even if there are any major differences throughout the years.

In 1999, there were only 10 hurricanes in the gulf, but only three noteworthy ones. The noteworthy hurricanes for that year were Bret, Irene, and Katrina, with Bret starting just north of Coatzacoalcos and continuing to move north before making an almost 80° turn to the left and moving in between Corpus Christi and Progreso before dissipating between near the Rio Grande River over 126 hours, with an average intensity of 65.91kt over that period. Irene started deep in the Caribbean Sea, moving north to hit 3 major cities directly, Havana, Key West, and Miami, relatively fast over its 114-hour lifespan with an average intensity of 52.5kt. The final noteworthy hurricane is Katrina starting in the Caribbean Sea but even lower that Irene, Katrina moved northeast and eventually hitting Belize City before dissipating at the where Mexico meets the ocean. Katrina was only around for 90 hours with a low intensity of 26.88kt but it covered a large distance relatively fast.

In the past twenty-five years we have had many hurricanes appear in the Gulf of Mexico and the Caribbean Sea, and it is important to identify how these storms affect the biggest cities in the gulf. There are three major factors that we are looking at in relation to each city and those are, the number of storms that hit a city, the average intensity of the storms at each city, and the average duration of the storms that hit a city.

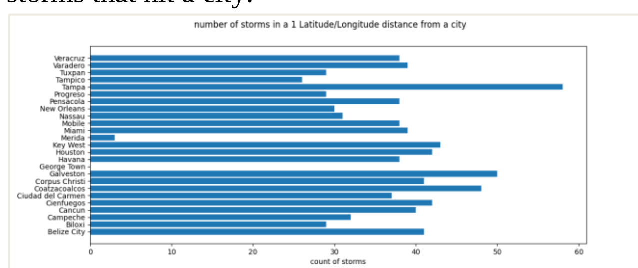

Figure 3: Storm frequency chart based on the latitudinal and longitudinal distance from a city.

For the number of storms, we counted the number of storms that entered a one latitude/longitude range of a city, or 54.6 miles east or west and 69 miles north or south of the twenty-five main cities in our study. A storm was able to hit multiple cities, and we recorded 881 total times a storm hit any of the cities, with an average of 35. What is interesting is that we have two outliers in our data, with George Town not getting a single storm, and Merida only having 3. This is most likely a cause of insufficient data collection, but it doesn't ruin our data. Ignoring the outliers we can see the city with the highest number of storms is Tampa, with 58 and the rest of the cities lying between 25 and 50 storms, excluding the outliers. This makes it so in any of these cities you can expect at least 1 hurricane a year, on average.

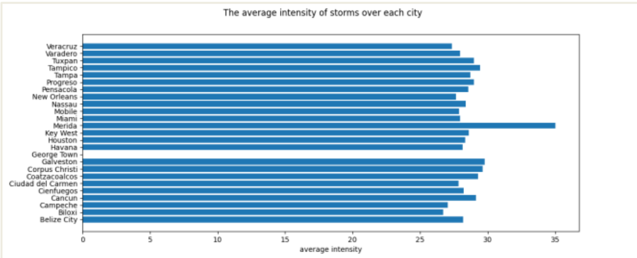

Figure 4: Average intensity of storms for each city.

With the number of storms that hit a city collected we can use this to calculate the average intensity of storms when they are over a city. With this we still have George Town and Merida as outliers, with George Town at 0kt and Merida at the highest value of 35kt. The rest of the cities don't have a lot of variances with all the values laying between the small range of 26kt and 30kt, and an overall intensity of 28kt, we can conclude from this data that as the storms hit cities, their intensity falls drastically compared to what it is when they're over the ocean.

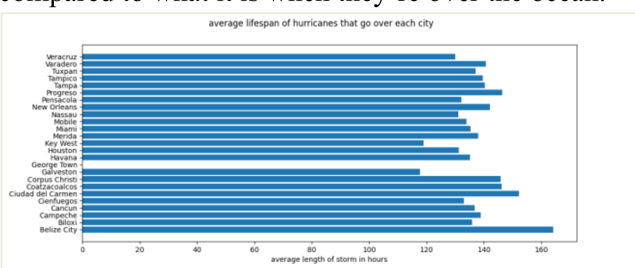

Figure 5: Average lifespan of a hurricane for each city.

The last important attribute we wanted to track is the average duration of storms that hit cities and with this attribute we only have one outlier at George Town. Besides the outlier, our data is a little more widespread but is still in a good range with a minimum of 117 hours and a maximum of 164 hours, creating a range of 47 hours. We can conclude with this that a certain number of hours is required for a storm to hit land and that seems to be at least above 100 hours.

RISK-PROFILE ANALYSIS

Lead Contributor: Hannah Jensen

There are an extreme number of factors for what causes tropical storms and hurricanes. In fact, there is a Chinese proverb that exemplifies this sentiment: "The flapping of the wings of a butterfly can be felt on the other side of the world." Very small events can lead to very large, unpredictable, and immeasurable changes. What this project desires are a way to somehow assess the risk that a certain predetermined list of cities is at, using a number of different variables, including geographical location and severities of previous storms.

For the Risk-Profile Analysis section of this project, we looked into the additional observations that can be used to assess risk aside from the historical location data. The attributes considered include sea surface temperature (SST), El Niño/La Niña patterns (ENSO), Atlantic Multidecadal Oscillation (AMO), Sahara dust levels, and upper-level wind patterns. After looking into these variables, we concluded that the ranking of importance is: SST, ENSO, and AMO ranked as the highest, followed by the upper-level wind patterns, then the Sahara dust levels last.

Sea Surface Temperature (SST)

Sea surface temperature is the measure, in temperature, of the sea's surface temperature over time. Many different organizations measure this in various ways with varying degrees of accuracy. We expect SST to not only be a good predictor for the risk of a tropical storm and/or hurricane, but also a fundamental one. This is because, as denoted by the NOAA[1], hurricanes begin when:

There is some kind of weather disturbance that pulls air in from all directions, and

The water at the ocean's surface where the hurricane begins is at least 80°F.

Heat is consequently shown to be a necessity for hurricane's formation. Additionally, the NOAA

includes that the reason hurricanes die out is because they lose touch of this hot water, either because they move over colder waters or because they reach land and subsequently die out. This fundamental role for hurricanes that ocean temperature plays in its creation is the reason that we chose to include it for our risk assessment analysis.
The dataset we found for SST was obtained from the U.S. Environmental Protection Agency (EPA). It includes an attribute that indicates how much of a deviance there is from the average ocean temperature from 1850 to the present day. Due to advancements in measurement technology, there is a stipulation in the documentation provided by the EPA that older data is less precise than newer data; to remedy this, a confidence interval was included as well.

El Niño/La Niña Patterns (ENSO)

The next attribute we looked into for its influences on tropical storms/hurricanes was El Niño and La Niña, where the latter is also known as Southern Oscillation. El Niño and La Niña describe the typical patterns seen in the ENSO (El Niño-Southern Oscillation) cycle, where El Niño is representative of the warmer phase and La Niña is representative of the cooler phase. They are categorized using trade winds, the Southern Oscillation Index, and sea surface temperature.[2] Each phase has different attributes to it. For example, El Niño is characteristic of the warming of the ocean which, like we mentioned before, is one of the factors that contributes to hurricanes, but also weaker trade winds. La Niña is the opposite, characterized by cooling water and stronger winds.

There have been numerous studies linking El Niño and La Niña to hurricane activity, including one from Florida State University that found a correlation between these patterns, where they found that, "the probability of one or more major hurricane landfall during El Niño is 23% but is 58% during neutral conditions and 63% during La Niña," (O'Brien)[3]. Thus, we expect times during La Niña in locations susceptible to hurricanes to be more at-risk during these times, and our model should reflect this.

The dataset used to capture El Niño and La Niña patterns was the intensities of those patterns given as a categorical variable measuring the ENSO value, which is based on the year (denoted as "Season"). The categories given are:

- WE: Weak El Niño
- ME: Moderate El Niño
- SE: Strong El Niño
- VSE: Very Strong El Niño
- WL: Weak La Niña
- ML: Moderate La Niña
- SL: Strong La Niña
- (blank): No particular ENSO pattern that year.

Atlantic Multidecadal Oscillation (AMO)

Atlantic Multidecadal Oscillation (AMO) is defined as the variability in the temperature of the ocean's surface. For this reason, we may expect some overlap or redundancy in this variable for our analysis with sea surface temperature, but with our SST variable measuring the anomaly and this value instead measuring the variability of the change in temperature, we included both.

According to Jeff Knight, Chris Folland, and Adam Scaife with the Advancing Earth and Space Sciences Journal, there is evidence that supports the fact that AMO is negatively correlated with tropical storms and hurricanes in the Atlantic. They support this, stating "[our] model simulation shows a similar band of significant… AMO correlations, supporting a link with the AMO [and hurricane activity in the Atlantic]." (Folland et al.)[4]. This supports the idea that this indicator will help predict risk for hurricanes.

The dataset obtained for AMO comes from the National Oceanic and Atmospheric Administration (NOAA). It contains information from 1856 to 2023, and unlike the other datasets we're using, this one is separated by months; this means that we can be more granular in our separation of this information when combining it with our original storm dataset.

Sahara Dust Levels

Sahara dust level refers to the aerosols and granules, often pollutive, that are found in the Sahara's atmosphere. There have been efforts made to reduce the dust levels, but some studies found that this may directly cause more hurricanes. It was noticeably more difficult to find studies correlating the Sahara dust levels with the creation of hurricanes, but there were negative correlations found in a few studies. For example, in a study by JGR Atmospheres measuring the effects of dust levels on tropical cyclone frequency, it was found that, "According to our results, controlling parameters for hurricane genesis do not depend crucially on dust" (Bretl et al)5. As one would expect, it was harder to model the relationship of dust and hurricanes versus the more easily quantifiable and relatable variables.

We did not use Sahara dust levels in our risk profiling analysis.

Upper-Level Wind Patterns

Our final attribute considered was upper-level wind patterns. From the ENSO phases, where one of the measurements tracked is trade winds, we expect that upper-level wind patterns will indubitably have an effect on hurricane formations.

There have been numerous studies on this variable's effect on hurricane activity as well, and models generated that help to support this. In one of these models, the air and ocean interaction was simulated under high wind conditions. There were some positive results shown, although they were highly volatile to other variables: "In agreement with previous studies, the present results indicate that the intensification of the model-simulated hurricane depends on the SST cooling due to the wind forcing associated with the hurricane" (Boa et al)[6]. It was more difficult to find appropriate datasets for this variable, and since its patterns should be supplemented by our ENSO variable, we did not include it for our risk assessment.

We did not use upper-level wind patterns in our risk profiling analysis.

SPATIAL CORRELATION ANALYSIS

Lead Contributor: Hannah Jensen

Setup

Our goal was to assess the risk that each city had of being hit by a hurricane while also factoring in how damaging the hurricane would likely be. To create a 0risk assessment in this way, we included the following variables:

• Sea Surface Temperature: "SST" (continuous variable)

• El Niño-Southern Oscillation: "ENSO" (categorical variable)

• Atlantic Multidecadal Oscillation: "AMO" (continuous variable)

We also included the distance that previous storms were from each city for this spatial correlation analysis of risk. ENSO was transformed via ordinal encoding, where Weak El Niño was ranked as the weakest value 1, up to Strong La Niña, which was ranked as the strongest at 7. This is because our research indicated that El Niño was indicative of less frequent hurricanes and conversely La Niña indicated more frequent ones.

We normalized these attributes via z-scoring, then calculated our Risk Score based on the sum of the results.

Results:

The comprehensive list of each city and their corresponding cumulative risk scores are calculated in the file *task3_spatial_analysis.ipynb*. The city that scored the highest was Mobile, AL, with a Risk Score of 11.95. The lowest score was Belize City, Belize, with a Risk Score of -21.00.

What the higher scores indicate is that the variables SST, ENSO, and AMO are all found to be indicative of a higher risk for hurricanes during the times that these cities are hit, how frequently, and the hurricanes strength. Alternatively, lower scores indicated that the three variables introduced had little influence on the formation of storms during those times. A score of zero means that there was somewhat of an influence from these variables, and these cities are moderately at risk for storms.

```
Varadero : 4.68
Key West : -3.60
Miami : -12.97
Tampico : -4.63
Houston : 6.35
Galveston : 0.31
New Orleans : 7.13
Biloxi : -12.29
Belize City : -21.00
Progreso : -7.42
Cancun : -9.29
Havana : 5.35
Corpus Christi : 3.07
Campeche : 0.24
Veracruz : 9.84
Nassau : -6.35
Pensacola : 9.27
Mobile : 11.95
Cienfuegos : 2.22
Tampa : -3.01
Tuxpan : 5.91
Coatzacoalcos : 5.03
Ciudad del Carmen : 8.33
Merida : 0.89
```

Figure 6: Results from the spatial correlation analysis based on the variables SST, ENSO, and AMO to assess the risk of each city specified.

NON-PARAMETRIC DENSITY ESTIMATION

Lead Contributor: Andrew Guzman

Introduction

This section of the Risk-Profile Analysis of the group assignment focuses on using Non-Parametric Density Estimation to analyze historical hurricane data. Based on past hurricane trajectories and their severity, the analysis targets areas with high location risks. This approach provides valuable insights into disaster preparedness and regional planning, particularly in areas prone to severe weather events.

The links provided in the Data Collection and Preparation section of the assignment were beneficial in obtaining the necessary TroPyCal Python library API information to integrate the dataset into the code. The HURDAT database from the National Hurricane Center offered a variety of data on hurricanes and cities that have been heavily impacted by them over the past 25 years. This database is a crucial resource as it contains extensive historical records, including detailed information on storm paths, intensities, and frequencies, which are very important for assessing risk patterns.

Setup

While gathering and organizing the required information was initially challenging due to the dataset's size and complexity, collaboration within the group helped resolve these issues efficiently. This teamwork also facilitated preprocessing the data and saving it as a .pkl (pickle) file for streamlined loading and analysis. Key fields such as latitude, longitude, and time were standardized, with time converted to datetime format to enable precise year-based filtering.

For the analysis, the Gulf Coast region was defined with specific latitude and longitude bounds: latitude between 10° and 35° N and longitude between -100° and -80° W. This geographic focus makes sure that the dataset captured storms that had the most direct impact on Gulf Coast areas. Having a large sample size spanning 25 years was very valuable for producing reliable and detailed predictions about location risks. The extensive dataset allowed for the identification of clear trends and patterns in storm trajectories and intensities over time. A custom filtering function, *get_gulf_storms_data*, was implemented to isolate storms that either formed within or entered the Gulf Coast region during this timeframe.

Kernel Density Estimation (KDE)

Following data filtering, Kernel Density Estimation (KDE) was applied to estimate storm occurrence density. The KDE, a non-parametric statistical technique, was particularly beneficial for this analysis as it does not assume a fixed data distribution. This flexibility is ideal for hurricane data, as storm paths and intensities are unpredictable and often influenced by numerous environmental factors. By estimating the probability density function of storm occurrences, KDE offered a clear visual representation of high-risk areas. The bandwidth parameter of the KDE was adjusted carefully to balance smoothness and clarity. This adjustment ensured that the density map effectively highlighted high-density areas without over-smoothing, which could obscure important patterns or dilute the visual.

The density map visualization was created using Cartopy, a Python library for geospatial data visualization. The map incorporated geographic context, including coastlines, landforms, and state boundaries to provide a comprehensive visual reference. The KDE output was overlaid as a color gradient, where warmer colors indicated areas of higher storm density, while lighter shades indicated regions with less frequent storm activity. This overlay made it possible to pinpoint high-risk areas with high amounts of clarity, making sure the visualization was both informative and accessible.

Results and Analysis

Along the Gulf Coast, the lower half of Mexico emerged as a significant hotspot for storm activity, particularly along its coastal areas. These regions consistently experienced high storm frequencies, highlighting their vulnerability to severe weather. The map showed dense clusters of activity that correlated with known hurricane landfall zones, underscoring the recurring nature of these events. Similarly, the coastal stretch near Honduras and the region between Cuba and the Yucatán Peninsula exhibited substantial storm activity, which is consistent with storm systems originating in the Caribbean. This area acts as a corridor for hurricanes, channeling storms into the Gulf of Mexico, where they often intensify before making landfall. In the United States, states such as Texas, Louisiana, Mississippi, Alabama, and Florida showed moderate but persistent storm densities. These states are frequently impacted by hurricanes, particularly those originating in the Gulf. While the density in these areas was not as high as in regions further south, the consistent storm activity still poses a considerable risk.

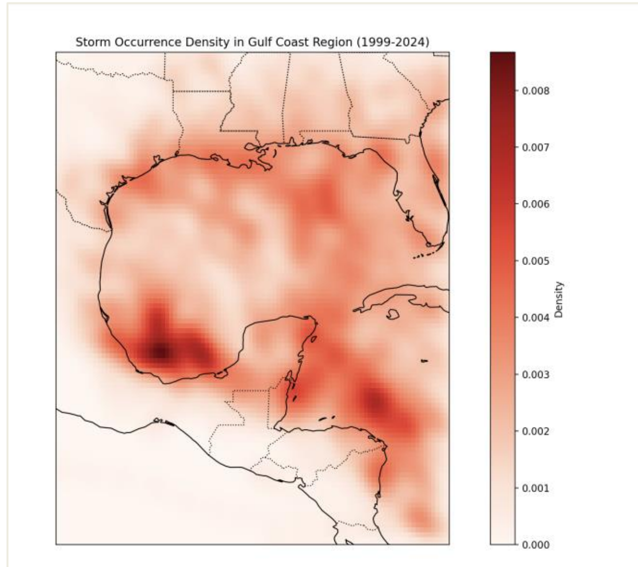

Figure 7: Geographical heatmap showing the hotspots for hurricane activity, strength, and frequency, as determined by the Non-Parametric Density Function defined in the previous section.

Despite its strengths, KDE has limitations. The smoothing process, while effective for general patterns, can obscure specific trajectory details or rare, intense storms, potentially underrepresenting critical outliers. For example, highly destructive hurricanes like Katrina or Harvey, which had significant impacts, may not be adequately highlighted in a smoothed density map. A potential improvement would involve integrating additional statistical techniques, such as clustering or outlier detection, to better identify and emphasize these extreme events. Additionally, expanding the dataset to cover a longer timeframe—such as 50 years instead of 25—could provide a more comprehensive understanding of long-term hurricane trends and their evolving impacts, particularly in the context of climate change.

COMPARISON OF RISK-PROFILING ANALYSIS RESULTS

Lead Contributors: Andrew Guzman and Hannah Jensen

The results from our spatial analysis and our non-parametric density estimation had results that were consistent with each other. For example, in the visual provided, we see the darkest portions of the graph correspond with the Mexican cities Veracruz and Ciudad del Carmen, which have the Risk Scores of 9.84 and 8.33 respectively. Additionally, those cities falling just outside of the dark red area correctly correspond with a lower Risk Score, with Tampico and Merida having values of -4.63 and 0.89 respectively.

By the nature of this dataset, we expect all of these cities to have active hurricane activity. Thus, it is important to keep in mind that this categorization is relative to hurricane-prone areas.

In our assessment, we find that the cities with the greatest risks are as follows:

• High Risk: Veracruz, Ciudad del Carmen, Mobile, New Orleans, Tuxpan, Coatzacoalcos, and Pensacola

• Moderate Risk: Havana, Varadero, Tampico, Galveston, Houston, Corpus Christy, Campeche, and Cienfuegos

• Low Risk: Key West, Miami, Biloxi, Belize City, Progreso, Cancun, Nassau, Tampa, and Merida

REFERENCES

[1] National Oceanic and Atmospheric Administration, 2007. How does the ocean affect hurricanes? *ACM* 50, 1 page (Jan, 2007), https://oceanexplorer.noaa.gov/facts/hurricanes.html#:~:text=Once%20they%20move%20over%20cold,waters%20fuel%20more%20energetic%20storms

[2] NOAA *How is El Niño measured?* (Nov, 2024), https://www.oc.nps.edu/webmodules/ENSO/how_is_ENSO_measured.html#:~:text=There%20are%20three%20main%20ways%20to%20measure%20an,using%20the%20TAO%20array%20and%20calculating%20an%20index

[3] Mark C. Bove, James B. Elsner, et al. 1998. *Effect of El Niño on U.S. Landfalling Hurricanes, Revisited,* https://journals.ametsoc.org/view/journals/bams/79/11/1520-0477_1998_079_2477_eoenoo_2_0_co_2.xml?tab_body=abstract-display

[4] Jeff R. Knight, Chris K. Folland, Adam A. Scaife, 2001. *Climate impacts of the Atlantic Multidecadal Oscillation*, AGU Advancing Earth and Space Sciences, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2006GL026242

[5] Sebastian Bretl, Philipp Reutter, et al., 2015. *The influence of absorbed solar radiation by Saharan dust on hurricane genesis*, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014JD022441

[6] J-W. Bao, J. M. Wilczak, et al. 2000. Numerical Simulations of Air–Sea Interaction under High Wind Conditions Using a Coupled Model: A Study of Hurricane Development, American Meteorologist Society AMS, https://journals.ametsoc.org/configurable/content/journals$002fmwre$002f128$002f7$002f1520-0493_2000_128_2190_nsoasi_2.0.co_2.xml?t:ac=journals%24002fmwre%24002f128%24002f7%24002f1520-0493_2000_128_2190_nsoasi_2.0.co_2.xml

[7] Hurricane Research Division, NOAA. 2024. *North Atlantic Hurricane Basin (1851-2023), Comparison of Original and Revised HURDAT*, https://www.aoml.noaa.gov/hrd/hurdat/comparison_table.html

[8] National Hurricane Center and Central Pacific Hurricane Center, NOAA. 2024. *NHC Data Archive,* https://www.nhc.noaa.gov/data/