Orkanskader: Modelling the impact of Hurricane Damage to Residential Property

V D Adithya Gadhamsetty	Beatrix Wen	Matthew Brownrigg	Richard Zhang
vgadham1@jh.edu	*ywen15@jh.edu*	*mbrownr1@jh.edu*	*rzhang89@jh.edu*

1 Introduction

With the damage caused by Hurricane Milton in the earlier sections of the year, it raises concerns on how we strategize to resolve challenges for coastal regions that deal with the aftermath of the large-scale hurricane. Furthermore, due to the variation of climate change, rising sea surface temperatures and fluctuating atmospheric conditions, the frequency and intensity of hurricanes have significantly increased in the recent years. The consequences, for which, are far-reaching and impacts infrastructural and economical damage. Among these, the financial damage borne by owners of residential properties remains a major concern due to the implications on property value and household financial stability of homeowners and insurers.

In recent times, hurricane forecasting tools have improved substantially, however they largely focus on the prediction of the trajectory and intensity of the hurricanes. Furthermore, they rely on meteorological and topological parameters coupled with high performance simulation technology. Yet, these tools often lack the ability to predict the spatial reach of a storm's destruction, particularly the radius of impact - which is a key factor in understanding the counties/areas that are under high risk and may experience severe damage. This lack of actionable insights into the affected areas exposes a significant gap in hurricane responsiveness for stakeholders, property owners and insurance agencies.

In order to bridge the divide, there is a demand and need for tools that can predict the path and radius of the storm along with its area-of-effect. Accurate modelling of this radius is important for the estimation of damage that may be incurred by homeowners and insurance companies. Especially so, for cases such as Florida, where residential properties are densely concentrated along hurricane-prone coastlines.

2 Motivation and Importance

The aftermath of a hurricane often leaves homeowners, insurers, and governments struggling to assess and respond to the damage effectively. Traditional forecasting systems are limited in their ability to quantify the spatial dimensions of destruction, leading to imprecise assessments of the regions at risk. This shortcoming creates significant challenges for insurance companies, which play a critical role in mitigating financial losses and aiding recovery. Without accurate information on the hurricane's radius of impact, insurers face difficulties in estimating the potential claims they might encounter, which in turn hinders their ability to set appropriate premiums, and plan financially for large-scale disasters.

Our proposed solution addresses these limitations through a two-fold approach:

1. **Radius Prediction**: We aim to develop a model that predicts the hurricane's radius of impact based on critical meteorological factors such as latitude, longitude, maximum wind speed, and radii of benchmark wind speeds (34, 50, 64 knots). This radius provides insights into the spatial extent of a hurricane's destructive force, offering precise information on the regions most likely to be affected.

2. **Damage Estimation**: Using the predicted radius, we estimate the economic damage incurred by residential properties based on their proximity to the eye of the storm. We further investigate appropriate distribution models that represent the radial impact of hurricanes, enabling a more systematic analysis of property damage severity.

The importance of this solution lies in its ability to provide actionable data that can benefit multiple stakeholders. Insurance companies, for instance, can leverage these predictions to refine their risk models, leading to more accurate estimation of losses and better financial preparedness. This enhanced risk modeling not only ensures the financial stability of insurers but also allows them to offer fair and sustainable premium structures for homeowners. Furthermore, improved damage assessments help disaster management agencies prioritize resources and plan recovery efforts effectively, minimizing the long-term economic burden on affected communities.

Through the combination of historical datasets that offer an aspect of meterological analysis and residential economical impact modeling, our research contributes towards the development of *Orkanskader*, a robust framework for understanding and providing actionable insights to mitigate hurricane-induced damage. This empowers decision makers and first responders to prepare effectively for natural disasters.

3 Data Ingestion and Usage

Orkanskader will incorporate data from various sources to establish a data pipeline for radii prediction and property damage modeling. To predict radii we will use the dataset provided by IBTrACS, yet our algorithm could be extended to other datasets such as Xu et. al [8] as our selected parameters can be commonly found across different publicly available hurricane dataset. The datasets consist of historical hurricane recordings and their `latitude`, `longitude`, `maximum wind speed` and `radius`. The paper also models wind speed (knots) at different radii labeled as R_{max}, R_{64},R_{50}, R_{34}. The data set has sparse data on these values. We hope to fill in the sparse data with our model's predictions.

Our dataset includes data that are missing some parameters. In some cases, the missing columns can be repaired by considering their relationship with other columns. For example, if the maximal wind speed is less than 64, we can infer that R_{64} is 0 even when the data point is missing this value. However, we are unable to impute the missing data and have to drop them in other cases as the relationship between known and unknown data is not always clear.

For the second phase, we attempted to find datasets that satisfied all requirements but to no avail. To combat the lack of a perfect dataset, our data is comprised of a merge between the US county boundaries [5] and the Florida county boundaries for coastal information [4]. This provides us with a geospatial viewpoint of the counties that may be affected due to the storm. Once we have the counties that have been affected, we utilize the aggregation of the median pricing of housing units [6] and the number of housing units [3] to estimate a range of the absolute damage that may have been done.

Once we have the range of absolute damage, we filter and transform it based on the attributes of the housing. The transformation process is enabled by the SEPHER 2.0 dataset [7], which allows us to classify housing in terms of ratios as apartment types or mobile homes to minimize/exclude them from the valuation estimation entirely.

4 Implementation

Orkanskader tackles two major problems during its investigation into property damage estimation:

- **Estimation of the radius from the eye of the storm**: Training a random forest model on features extracted, including latitude, longtude R_{34}, and max wind speed to predict R_{64} and R_{max}. We model the predictor to estimate the radius of the storm having damaging speed (according to NSSL [1], such speed is 50 - 60 mph).

- **Identification and classification of affected zones**: Affected zones will be identified using a radial distribution model representing windspeed impact and the rmax. We collect the boundaries for the counties of the US and project the radius of the storm at a particular timeframe onto the counties to find out which ones overlap with the storm's circle.

4.1 Random Forest Regression

To predict the value R_{64} and R_{max}, we used random forest regressor with the parameters stated above. Random forest is used because the relationship between the hurricane's location(latitude and longitude), wind speed, and area of damage is clearly non-linear due to the chaotic nature of the atmospheric system. We also did experiments using polynomial regression, but the prediction output was not as good as random forest. On the other hand, comparing with the more commonly used deep learning approaches, random forest is more suitable for our data size as deep learning models tend to converge more slowly and requires a larger input to train upon. Aside from that, random forest is more robust to overfitting and requires less computational resources, making it more suitable for the task in hand. We tackled our data with two different approaches, in both cases with 100 estimators. The data is split between training(80%) and testing(20%) sets for evaluation.

Our first approach of using random forest directly predicts our target data using the input parameters. This approach achieves a Mean Absolute Percentage Error (MAPE) of 24.2%, and such accuracy is insufficient for the precise estimate, but it provides a good baseline for further experimentation.

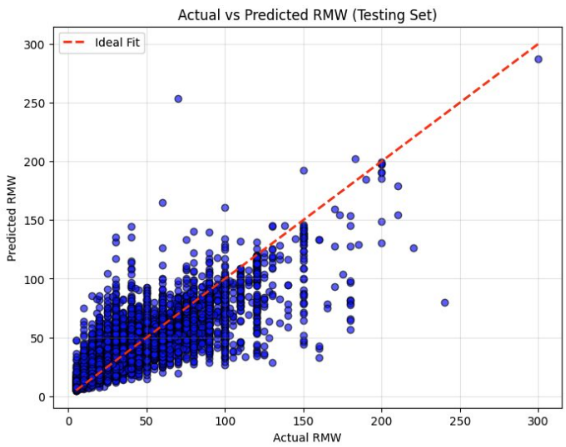

Figure 1: Direct Prediction of Radius of Max Wind Speed

The second approach utilizes the existing R_{max} estimation function retrieved from Chavas's study over estimation of cyclone radius of maximum wind from outer size [2]. While most variables used in the estimation function are available in the input parameters, $\frac{M_{max}}{M_{17.5ms}}$ can only be calculated when R_{max} is known. This allows us to make $\frac{M_{max}}{M_{17.5ms}}$ our unknown variable and benefit from the existing estimation formula. After $\frac{M_{max}}{M_{17.5ms}}$ is predicted, we then derive R_{max} using the formula below. This approach achieves a MAPE of 3.6% without any modification to the random forest regressor itself. Aside from that, we find that the equation is applicable not only in case of R_{max} calculation but also in R_{64} calculation if we replace all max wind speed related variables with their corresponding values for the wind speed of 64 miles per hour. Thus, we could estimate R_{64} with similar accuracy. This is hugely beneficial as the wind speed of 70 mph(64 knots per second in R_{64} converts to 70 mph) is enough to conduct noticeable damage to properties, and estimating the radius with this wind speed accurately allows us to determine the area where damage is significant.

1) Calculate $M_{17.5\,\mathrm{ms}}$: (Inputs: $R_{17.5\,\mathrm{ms}}$ and f)

$$M_{17.5\,\mathrm{ms}} = R_{17.5\,\mathrm{ms}} \times (17.5\,\mathrm{m\ s^{-1}}) + \frac{1}{2} f R^2_{17.5\,\mathrm{ms}}.$$

2) Calculate $M_{\max}$: (Inputs: $M_{17.5\,\mathrm{ms}}$)

$$M_{\max} = \left(\frac{M_{\max}}{M_{17.5\,\mathrm{ms}}}\right) M_{17.5\,\mathrm{ms}}.$$

3) Solve for $R_{\max}$ from $M_{\max}$: (Inputs: $V_{\max}$ and f)

$$R_{\max} = \frac{V_{\max}}{f}\left(\sqrt{1 + \frac{2 f M_{\max}}{V^2_{\max}}} - 1\right).$$

Figure 2: Formula used to calculate M and recover R_{max}

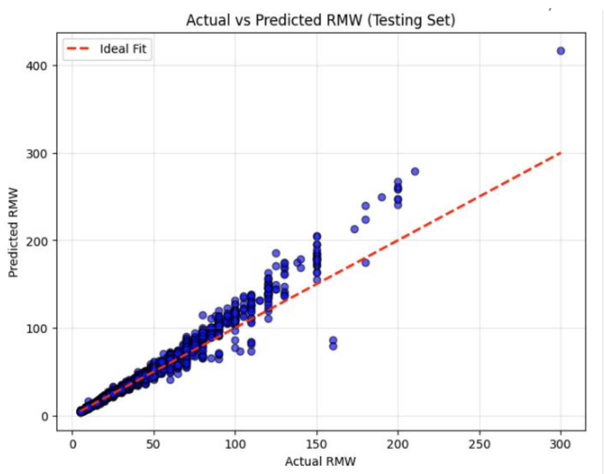

Figure 3: Indirect Prediction of Radius of Max Wind Speed

4.2 Prediction of Hurricane Location

To further improve our the practicality of our design, we attempted in predicting the movement of hurricane, thus allowing us to predict the area of maximal damage before the arrival of the hurricane. We created two separated LSTM models to learn the latitude and longitude of the hurricane in the next timestamp given the information of this particular hurricane in all previous timestamps. LSTM is applied because it remembers not only the current moment but also the previous data, which is hugely useful as the previous path of the hurricane is definitely useful in predicting its movement in the next step. The predicted output shows some positive correlations between the true latitude/longitude and the predicted values, yet the standard deviation is too large for the result to be used practically in phase two due to practical limitations in data size and computational resources, along with the emphasis on radius estimation.

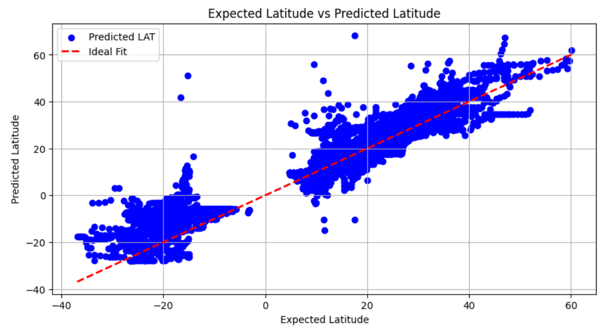

Figure 4: Latitude Prediction Result

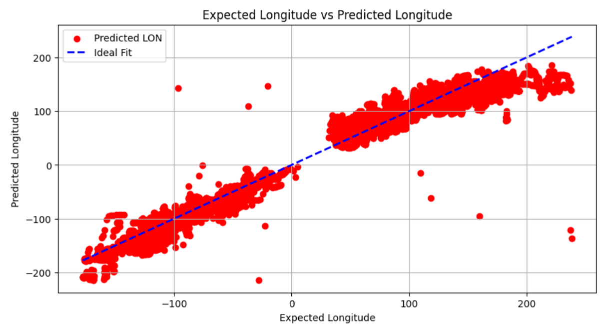

Figure 5: Longitude Prediction Result

4.3 Identification and classification of affected zones

To estimate the regions impacted by a hurricane and quantify potential economic damage, we leverage the attributes necessary for projecting an elliptical impact area on the map. These attributes include the latitude, longitude, R_{64} (radius of 64-knot winds), and R_{max} (maximum radius of hurricane-force winds).

Using these parameters, we approximate the geographical region of interest by projecting a circular area that encompasses a significant portion of the counties affected by the hurricane. The following preprocessing steps are performed to enable data aggregation and valuation:

1. **Parsing Housing Valuation Data**: We extract county-level housing valuation data by parsing publicly available PDFs that report the median price of housing units [6] grouped by county [5].

2. **State-Level Focus**: We narrow the analysis to the state of interest, in this case, Florida [4], which is frequently impacted by hurricanes.

3. **Merging Housing Unit Data**: The housing unit dataset [3], which provides the number of residential units per county, is merged with the median housing

price data. This integration allows for the estimation of total residential property valuation at a county level.

4. **Adjusting for Unaccounted Housing Types**: Certain housing types, such as apartments and mobile homes, may not be fully represented in the dataset. To account for these, we incorporate ratios derived from the SEPHER [7] dataset, which are used as adjustment factors to refine the county-level property valuations.

5. **Transformation of County Boundaries**: The geographical boundaries of Florida counties and coastal areas are transformed into a usable format, producing a comprehensive county-boundary dataset specific to the state of Florida.

Post these preprocessing steps, the parameters from the previous phase (latitude, longitude, R_{64}, and R_{max}) can be applied to project the affected areas onto a geographical map, as well as produce a reasonable valuation for the damages caused by the hurricane to residential property.

4.4 Model Evaluation on Hurricane Wilma 2005

The damage estimation model provides similar predictions to historical hurricane damage calculations. Using our model for radius estimation we can simulate damage prediction for any hurricane in the Using the IBTrACS dataset. To evaluate our damage prediction model we chose hurricane Wilma (2005), a brief hurricane that affected the lower counties in Florida. This hurricane has a short lifespan and was contained to one region of Florida. This allows for a simpler timestamp analysis of damage.

Hurricane	**latitude**	**longitude**	R_{64}	R_{max}
Wilma (2005)	27.0	-79.9	75.0	31.0

Using Hurricane Wilma data and radius prediction into our damage estimation model we are able to generate a map figure that details affected counties and their distance from the eye of the storm. The orange circle denotes the range of R_{64} while the orange counties denote which counties fall within R_{64} range. Similarly the red circle denotes the range of R_{max} and red counties are counties that fall within R_{max}.

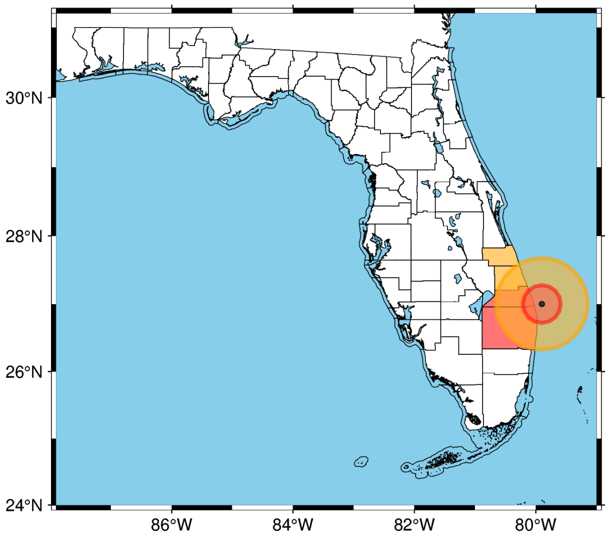

Figure 6: Modeling Hurricane Reach

The model also outputs county valuation and predicted damage given the selected counties. Damage given each county is calculated as a right skewed graph given their position.

County Region	**Damage (% of Home Value)**
Coastal R_{max} County	2.5
R_{max} County	1.5
R_{64} County	1.0

Using these Damage value assumptions we are able to calculate damage using the following Gaussian variate assumptions:

$$\mathcal{N}(d_{ij}, 0.10)$$

sampled h_{ij} times.

where d denotes the damage percentage of home and h denotes number of single family homes. i denotes the county region listed above and j denotes the county within county region.

$$\mathcal{N}(v_{ij}, \frac{v_{ij}}{2})$$

sampled h_{ij} times.

where v_i denotes the median value of a home and h_i denotes number of single family homes. i denotes the county region listed above and j denotes the county within county region.

From sampling both Gaussian variates we are given vectors for each county that we can sum together to get damage within county region. We can then sum each county region damage to get total damage.

With these calculation assumptions we are able to achieve similar damage values to historical records from hurricane Wilma (2005):

Model	**$ Valuation (billions)**
Wilma (2005)	19.0
Predictive Model	6.678

Although the valuation appears to be much lower than Wilma's reported damage it is important to note that Wilma reports the total damage over a time series that spans most of lower Florida. It is fair to assume that our model is reporting a third of Wilma's damage. This results in closer estimation of $6.33 billion which is a 5.5% difference in estimation.

5 Conclusion

With global warming and changing weather conditions, we inch closer to a more unpredictable and "stormy" future. Understanding the trends, frequency and intensity of hurricanes is key in mitigating threats to coastal regions like Florida. Despite current forecasting strategies being effective at predicting the overarching storm's trajectory, it fails to provide us with a granular and closer-to-earth perspective on the impacts of the hurricane. This limitation hinders effective disaster planning and risk management.

Orkanskader addresses this gap by chaining two approaches: predicting the radius of a hurricane's affected area and estimating the economic damage to residential properties based on their proximity to the storm's eye. By leveraging attributes of meteorological data, such as latitude, longitude, maximum wind speed, and radius, our model aims to provide actionable insights into the spatial distribution of hurricane-induced destruction. This data-driven framework can help provide missing radius data for other scientific study to do cross reference. It also enables insurance companies to refine risk assessments, allocate financial reserves, and set premiums with greater precision. Furthermore, it helps in allowing local authorities to strategize and prioritize recovery efforts more effectively.

Future work can refine this model by incorporating real-time data, more granular (zip code based) data points, and features that contribute towards the structural integrity of households near the coastal line.

References

[1] Severe weather 101 - damaging winds basics. https://www.nssl.noaa.gov/education/svrwx101/wind/#:~:text=Most%20thunderstorm%20winds%20that%20cause,those%20exceeding%2050%2D60%20mph.

[2] CHAVAS, D. R., AND KNAFF, J. A. A simple model for predicting the tropical cyclone radius of maximum wind from outer size. *Weather and Forecasting 37*, 5 (2022), 1235–1251.

[3] GOV, U. National, state, and county housing unit totals: 2020-2023. https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-housing-units.html.

[4] OF FLORIDA GEOPLAN CENTER, U. Florida county boundaries. https://fgdl.org/zips/metadata/htm/cntbnd_sep15.htm.

[5] OPENDATASOFT. Us county boundaries. https://public.opendatasoft.com/explore/dataset/us-county-boundaries/table.

[6] REALTOR, N. County median home prices and monthly mortgage payment. https://www.nar.realtor/research-and-statistics/housing-statistics/county-median-home-prices-and-monthly-mortgage-payment.

[7] TEDESCO, M., HULTQUIST, C., AND DE SHERBININ, A. A new dataset integrating public socioeconomic, physical risk, and housing data for climate justice metrics: A test-case study in miami. *Environmental Justice 15*, 3 (2022), 149–159.

[8] XU, Z., GUO, J., ZHANG, G., YE, Y., ZHAO, H., AND CHEN, H. Global tropical cyclone size and intensity reconstruction dataset for 1959–2022 based on ibtracs and era5 data. *Earth System Science Data Discussions 2024* (2024), 1–23.