

# Hurricane Intensity Prediction using ARIMA Model

June 29, 2025

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Background on hurricane prediction . . . . .	2
1.2	Importance of accurate intensity forecasting . . . . .	2
1.3	Overview of ARIMA modeling approach . . . . .	2
<b>2</b>	<b>Data Overview</b>	<b>2</b>
2.1	Dataset description and source . . . . .	2
2.2	Key variables and their significance . . . . .	2
2.3	Data quality assessment . . . . .	2
<b>3</b>	<b>Methodology</b>	<b>2</b>
3.1	ARIMA model explanation . . . . .	2
3.2	Data preprocessing steps . . . . .	3
3.3	Model parameter tuning . . . . .	3
<b>4</b>	<b>Exploratory Data Analysis</b>	<b>3</b>
4.1	Time series decomposition . . . . .	3
4.2	Pressure-wind correlation analysis . . . . .	3
4.3	Temporal patterns in hurricane development . . . . .	3
<b>5</b>	<b>Model Development</b>	<b>3</b>
5.1	ARIMA implementation . . . . .	3
5.2	Parameter optimization . . . . .	3
5.3	Validation using walk-forward testing . . . . .	3
<b>6</b>	<b>Results and Evaluation</b>	<b>3</b>
6.1	Model performance metrics (MAE/RMSE) . . . . .	3
6.2	Residual analysis . . . . .	3
6.3	Comparison with baseline predictions . . . . .	4
<b>7</b>	<b>Discussion</b>	<b>4</b>
7.1	Interpretation of results . . . . .	4
7.2	Impact analysis of pressure changes on intensity . . . . .	4
7.3	Early warning threshold identification . . . . .	4
<b>8</b>	<b>Conclusion and Future Work</b>	<b>4</b>
8.1	Summary of findings . . . . .	4
8.2	Limitations of the current approach . . . . .	4
8.3	Potential improvements and extensions . . . . .	4

# 1 Introduction

## 1.1 Background on hurricane prediction

Hurricanes are among the most devastating natural disasters, causing significant economic and social impacts. Accurate prediction of hurricane intensity is crucial for effective disaster preparedness and response.

## 1.2 Importance of accurate intensity forecasting

Accurate intensity forecasts enable timely evacuation orders, resource allocation, and infrastructure protection measures. Improving these forecasts can significantly reduce damage and save lives.

## 1.3 Overview of ARIMA modeling approach

ARIMA models are widely used for time series forecasting due to their ability to capture temporal dependencies in data. This project employs ARIMA to predict hurricane intensity based on historical storm data.

# 2 Data Overview

## 2.1 Dataset description and source

The dataset used in this project contains historical data of Atlantic hurricanes from 1975 to 2021, including wind speed, pressure, location, and time information. The data was obtained from [Kaggle Dataset](<https://www.kaggle.com/datasets/utkarshx27/noaa-atlantic-hurricane-database>).

## 2.2 Key variables and their significance

- **Wind:** Maximum sustained wind speed in knots (target variable).
- **Pressure:** Air pressure at the storm's center in millibars.
- **Lat/Long:** Geographic coordinates of the storm.
- **Datetime:** Combined year, month, day, and hour for time series analysis.

## 2.3 Data quality assessment

The dataset contains missing values in several columns. Columns with more than 90% missing values, such as 'category' and 'hurricane\_force\_diameter', were removed. Remaining missing values in 'tropical-storm\_force\_diameter' were filled with 0.

# 3 Methodology

## 3.1 ARIMA model explanation

ARIMA (Autoregressive Integrated Moving Average) models are used to forecast time series data by capturing its autocorrelations. The model is denoted as ARIMA(p, d, q), where:

- p: Order of autoregression (AR)
- d: Degree of differencing (I)
- q: Order of moving average (MA)

### 3.2 Data preprocessing steps

1. Remove irrelevant columns: 'Unnamed: 0', 'category', 'hurricane\_force\_diameter'.
2. Fill missing values: Replace NaN values in 'tropicalstorm\_force\_diameter' with 0.
3. Create datetime column: Combine year, month, day, and hour into a datetime object.
4. Data Scaling: Scale features to a uniform range

### 3.3 Model parameter tuning

The parameters (p, d, q) of the ARIMA model were tuned using grid search and AIC (Akaike Information Criterion) to find the best-fitting model.

## 4 Exploratory Data Analysis

### 4.1 Time series decomposition

Time series decomposition was performed to separate the data into trend, seasonality, and residual components.

### 4.2 Pressure-wind correlation analysis

Correlation analysis was conducted to understand the relationship between air pressure and wind speed, which is critical for hurricane intensity prediction.

### 4.3 Temporal patterns in hurricane development

Analysis of temporal patterns in hurricane development was performed to identify recurring trends and cycles in storm behavior.

## 5 Model Development

### 5.1 ARIMA implementation

The ARIMA model was implemented using Python's statsmodels library.

### 5.2 Parameter optimization

The ARIMA model parameters were optimized using techniques such as grid search to minimize the AIC.

### 5.3 Validation using walk-forward testing

Walk-forward testing was used to validate the model's performance on unseen data.

## 6 Results and Evaluation

### 6.1 Model performance metrics (MAE/RMSE)

The model performance was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

### 6.2 Residual analysis

Residual analysis was performed to check the model's assumptions and identify potential areas for improvement.

### 6.3 Comparison with baseline predictions

The model's performance was compared with baseline predictions to assess its effectiveness.

## 7 Discussion

### 7.1 Interpretation of results

The results were interpreted in terms of the model's ability to predict hurricane intensity based on historical data.

### 7.2 Impact analysis of pressure changes on intensity

The impact of pressure changes on hurricane intensity was analyzed to understand the key drivers of storm intensification.

### 7.3 Early warning threshold identification

Early warning thresholds were identified to provide timely alerts for potential high-intensity hurricanes.

## 8 Conclusion and Future Work

### 8.1 Summary of findings

The project successfully developed an ARIMA model for predicting hurricane intensity.

### 8.2 Limitations of the current approach

Limitations include the reliance on historical data and the model's inability to account for unforeseen environmental factors.

### 8.3 Potential improvements and extensions

Potential improvements include incorporating additional environmental variables, using machine learning techniques, and expanding the dataset.

## References

- [1] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M., *Time Series Analysis: Forecasting and Control*. 5th ed. Hoboken, NJ: Wiley, 2015.
- [2] Emanuel, K.A., "The dependence of hurricane intensity on climate," *Nature*, vol. 326, no. 6112, 1987, pp. 483–485.
- [3] Elsner, J.B., Kossin, J.P., and Jagger, T.H., "The increasing intensity of the strongest tropical cyclones," *Nature*, vol. 455, no. 7209, 2008, pp. 92–95.
- [4] Hyndman, R.J., and Athanasopoulos, G., *Forecasting: Principles and Practice*. 3rd ed. Melbourne: OTexts, 2021.
- [5] NOAA (National Oceanic and Atmospheric Administration), "Hurricane Research Division: Frequently Asked Questions." [Online]. Available at: [www.aoml.noaa.gov/hrd-faq/](http://www.aoml.noaa.gov/hrd-faq/) (Accessed: April 7, 2025).
- [6] Zhang, G.P., "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, no. C, 2003, pp. 159–175.