

# Spatio-Temporal Analysis of Earthquake Magnitudes in Japan

## 1. Introduction

Earthquakes are a significant natural hazard in Japan, and understanding their magnitudes is crucial for risk mitigation. This project utilizes spatio-temporal statistical models to analyze and predict earthquake magnitudes using a dataset of geographic, temporal, and event-specific variables.

Spatio-temporal models consider both spatial and temporal correlations, making them suitable for data where dependencies exist across space (locations) and time (periods). These models are more robust than traditional regression models as they explicitly model spatial and temporal structures, often observed in natural processes like earthquakes.

The objectives are:

1. To explore spatial and temporal trends in earthquake magnitudes.
2. To apply spatial interpolation methods such as Inverse Distance Weighting (IDW) and Gaussian interpolation to predict earthquake magnitudes across a spatial grid.
3. To visualize the results and evaluate the model's performance.

## 2. Data Explanation

The dataset contains the following variables:

- **Magnitude (mag):** The response variable, representing the earthquake's intensity.
- **Month, Day:** Temporal indicators of when the earthquake occurred.
- **Latitude and Longitude:** Geographical coordinates of the earthquake's epicenter.
- **Depth:** Depth of the earthquake in kilometers, which can influence its' surface impact.
- **Gap:** Measurement gap, indicating the quality of the recorded data.
- **Dmin:** Minimum distance from a seismic station to the earthquake's epicenter.
- **RMS:** Root mean square of the recorded measurement.

### Data Features:

- **Spatial variation:** Earthquakes are geographically distributed with different magnitudes across regions.
- **Temporal variation:** Earthquakes occur at various times, potentially exhibiting trends.

We aim to model these variations and predict earthquake magnitudes using spatio-temporal methods.

## 2.1 Dataset Overview

The dataset contains 564 observations and 9 variables, each describing different aspects of earthquake events in Japan. Here's a summary of the key variables:

### 1. Temporal Variables:

- month: Month of the earthquake, ranging from January (1) to December (12)
- day: Day of the earthquake

### 2. Spatial Variables:

- latitude: Latitude of the earthquake's epicenter, ranging from 23.53 to 50.59 degrees
- longitude: Longitude of the earthquake's epicenter, ranging from 124.6 to 156.8 degrees

### 3. Earthquake Characteristics:

- depth: Depth of the earthquake in kilometers, ranging from 2.73 to 510.33 km, with a median of 35.00 km
- mag: Magnitude of the earthquake (response variable), ranging from 4.5 to 6.6, with an average of 4.76
- gap: Measurement gap, indicating the quality of data, ranging from 14 to 214
- dmin: Minimum distance to the seismic station, ranging from 0.139 to 18.781
- rms: Root mean square of the measurement, ranging from 0.27 to 1.54 (Root Mean Square is used to represent the average strength of seismic waves by taking the square root of the average of the squared amplitudes of the wave signal, essentially providing a more consistent measure of ground motion compared to just looking at peak amplitudes)

Table 1. Descriptive Statistics

Variable	Description	Min	1st Quartile	Median	Mean	3rd Quartile	Max
depth	Depth of the earthquake (km)	2.73	10.00	35.00	47.84	50.60	510.33
mag	Magnitude of the earthquake	4.5	4.5	4.6	4.763	4.9	6.6
gap	Measurement gap (quality indicator)	14	77.75	114	107.86	137	214
dmin	Minimum distance to seismic station (km)	0.139	1.150	1.915	2.437	3.271	18.781
rms	Root mean square of measurement	0.27	0.65	0.78	0.8071	0.94	1.54

## Key Observations

1. **No Missing Values:** The dataset is complete, with no missing entries in any variable.
2. **Magnitude Variability:** Earthquake magnitudes are concentrated between 4.5 and 6.6, with a slightly right-skewed distribution (mean > median).
3. **Depth Range:** The depth varies widely, with extreme outliers exceeding 500 km, suggesting the need to inspect depth's effect on magnitude.
4. **Data Quality:** Variables like gap and dmin measure the precision of the data collection, which might influence the reliability of magnitude predictions.

## 2.2 Exploratory Data Analysis

The goal of this analysis is to explore the spatio-temporal patterns in the earthquake dataset. We aim to map the spatial locations of earthquakes in Japan.

### Spatial Distribution

The spatial distribution of earthquake magnitudes in Japan is illustrated in the map below. The points represent earthquake events, with their size corresponding to magnitude and their color indicating depth. The map highlights regions with frequent seismic activity, primarily along the tectonic boundaries near Japan.

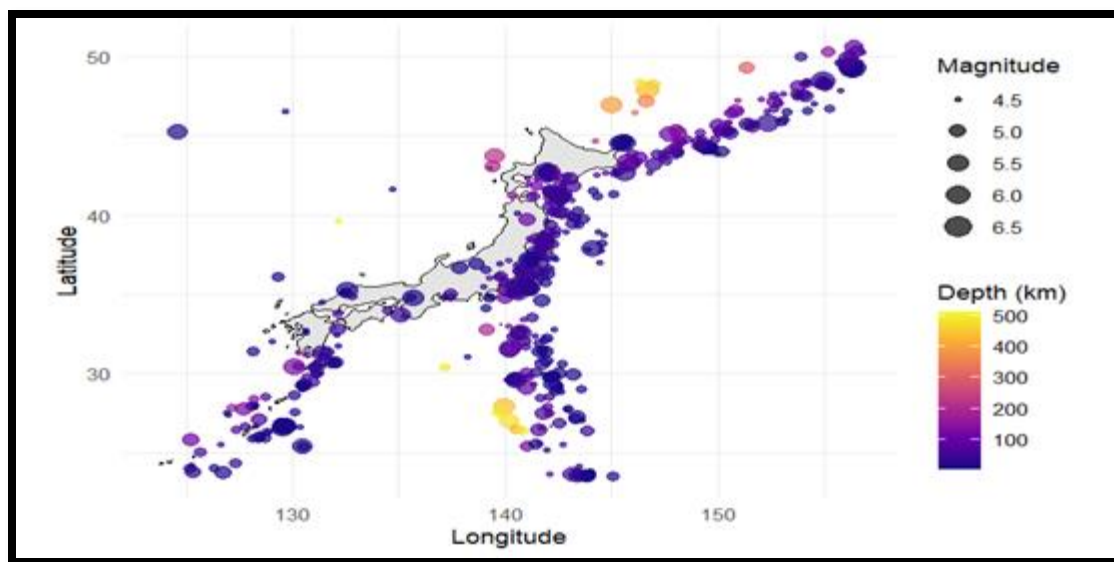


Figure 1. Spatial Distribution of the Earthquakes in Japan

### Boxplot Trends

The boxplot depicts the distribution of earthquake magnitudes across different months in Japan. Each box represents the interquartile range (IQR) of magnitudes, with the horizontal line within the box indicating the median magnitude for the month. The whiskers extend to the range of non-outlier data, while the red points represent outliers.

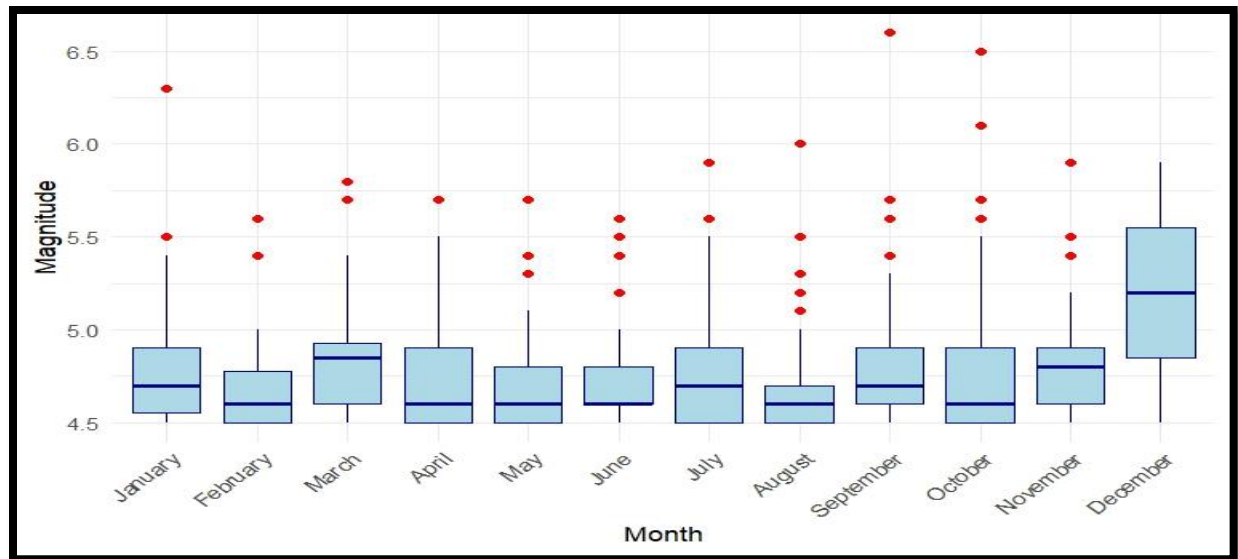


Figure 2. Monthly Earthquake Magnitudes by Boxplot

### Key Observations:

1. **Median Magnitude:**
  - The median earthquake magnitude is relatively consistent across most months.
  - **December** shows a slightly higher median compared to other months.
2. **Variability:**
  - The interquartile range (IQR), represented by the height of the boxes, varies across months.
  - **December** appears to have the largest IQR, indicating greater variability in earthquake magnitudes during this month.
  - Other months, such as **June and August**, have smaller IQRs, indicating less variability and suggesting more consistent magnitudes.
3. **Outliers:**
  - Outliers, shown as red dots, are present in every month, except **December**. These represent earthquakes with unusually high magnitudes relative to the typical range for that month. Months like **August, September, and October** have a notably higher number of outliers.
4. **Whisker Length:**
  - The whiskers extend further in months like **December**, suggesting a broader range of non-outlier data. Other months, such as **February and November**, have shorter whiskers, indicating a more concentrated distribution of magnitudes.

### Seasonal Variations in the Spatial Distribution of Earthquake Magnitudes in Japan

This plot provides a seasonal perspective of earthquake magnitudes across Japan.

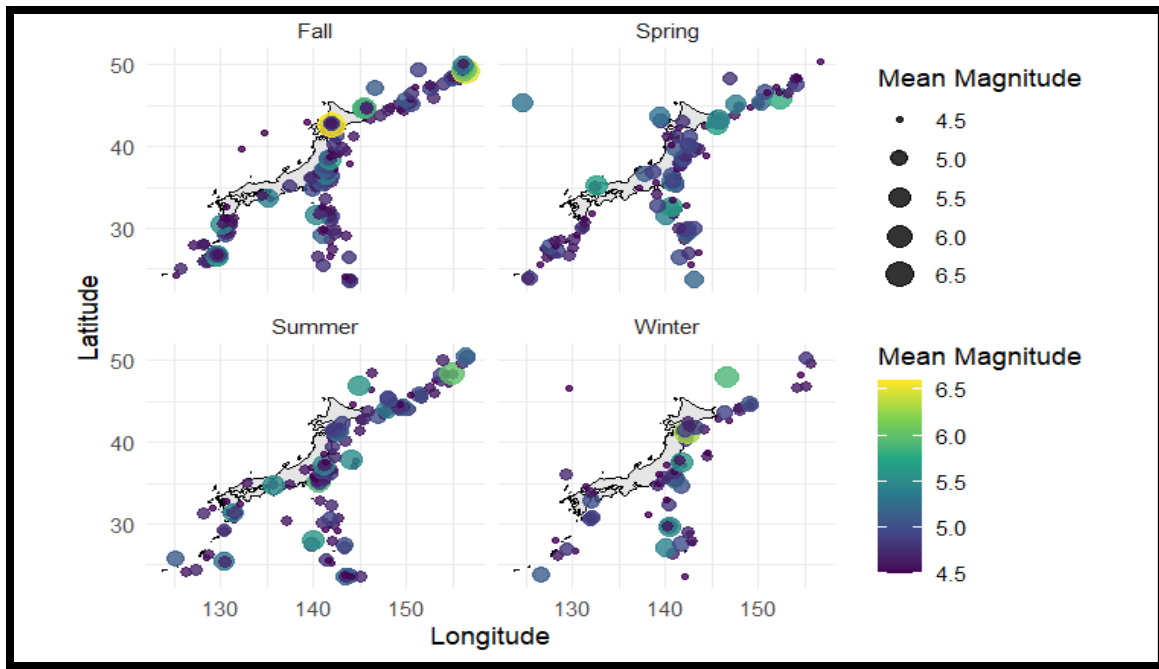


Figure 3. Mean Earthquake Magnitudes by Season in Japan

Here's a breakdown of what it shows:

1. **Seasonal Breakdown:**

- The plot is divided into four panels, each representing a season: Spring, Summer, Fall, and Winter.
- This allows us to observe how earthquake activity, in terms of magnitude, varies spatially and seasonally.

2. **Spatial Distribution:**

- Earthquakes are shown on the map of Japan with their respective locations (latitude and longitude).
- The distribution suggests that seismic activity occurs across a wide geographic area, with some clustering in specific regions.

3. **Magnitude Representation:**

- The size and color of the points represent the **mean earthquake magnitude** for each location during the given season.
- Larger and darker points indicate higher mean magnitudes, while smaller and lighter points represent lower mean magnitudes.

4. **Seasonal Observations:**

- **Fall:** Seems to have slightly more intense magnitudes in the northern regions.
- **Spring:** Earthquake magnitudes appear more evenly distributed, with moderate activity across Japan.
- **Summer:** Shows slightly lower magnitudes overall compared to other seasons.
- **Winter:** Some significant points are visible, particularly in the northern and eastern regions.

### Spatial Density Map of Earthquakes

This contour map illustrates the spatial density distribution of earthquake occurrences across the study region, represented in terms of latitude and longitude. The density levels are visualized using contour lines and a gradient color scale.

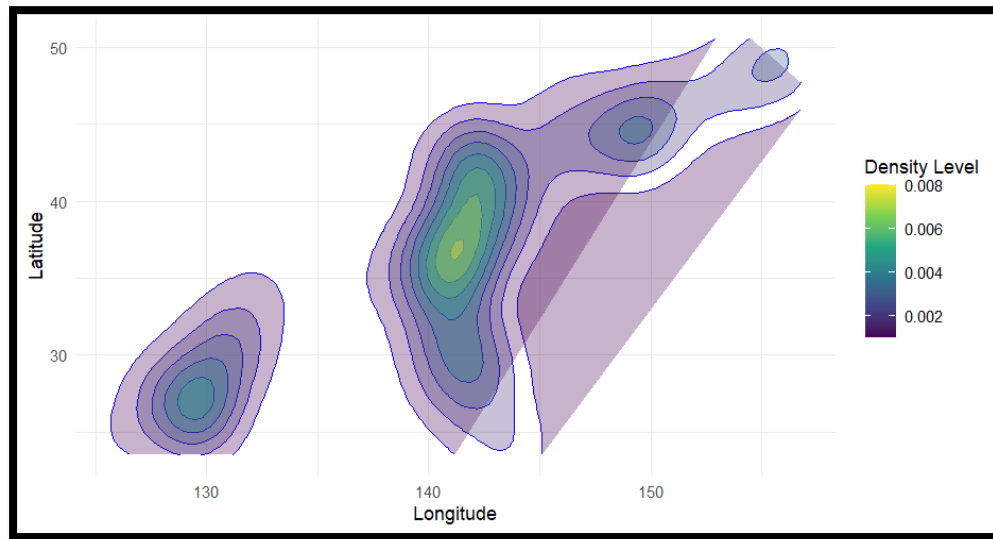


Figure 4. Spatial Density Map of Earthquakes

#### Density Distribution:

- The contours indicate regions with varying earthquake densities, where areas with denser contours (yellow to green) correspond to higher earthquake frequencies.
- The areas with lighter colors, such as yellow, signify the highest density of earthquakes, while darker areas (blue to purple) show lower density.

#### Spatial Insights:

- Two major zones of high density are apparent, centered near longitudes around **130** and **140**, which align with tectonic activity zones in Japan.
- This spatial density map is a useful visualization for identifying earthquake hotspots and assessing the spatial distribution of seismic activity.

## Chapter 3: Spatial and Temporal Analysis of Earthquake Magnitudes

### 3.1 Introduction

In this chapter, we analyze the spatial and temporal distribution of earthquake magnitudes observed in Japan during 2018. The primary focus is to evaluate and compare two spatial interpolation methods: Inverse Distance Weighting (IDW) and Gaussian interpolation. Leave-One-Out Cross-Validation (LOOCV) is employed to optimize the parameters and assess the performance of these methods.

### 3.2 Spatial Interpolation Using IDW

IDW interpolation predicts earthquake magnitudes across a spatial grid for each month by weighting observations inversely proportional to their spatial distance. The method assumes closer observations exert greater influence on the predicted values. An optimal parameter ( $\alpha = 4$ ) for IDW was determined using LOOCV, resulting in a minimum cross-validation error of 0.167.

#### Visualization of Predicted Magnitudes using IDW

The results of the IDW interpolation are presented in a faceted map for each month. The spatial patterns of predicted magnitudes were visualized using a color gradient ranging from blue (low magnitudes) to red (high magnitudes).

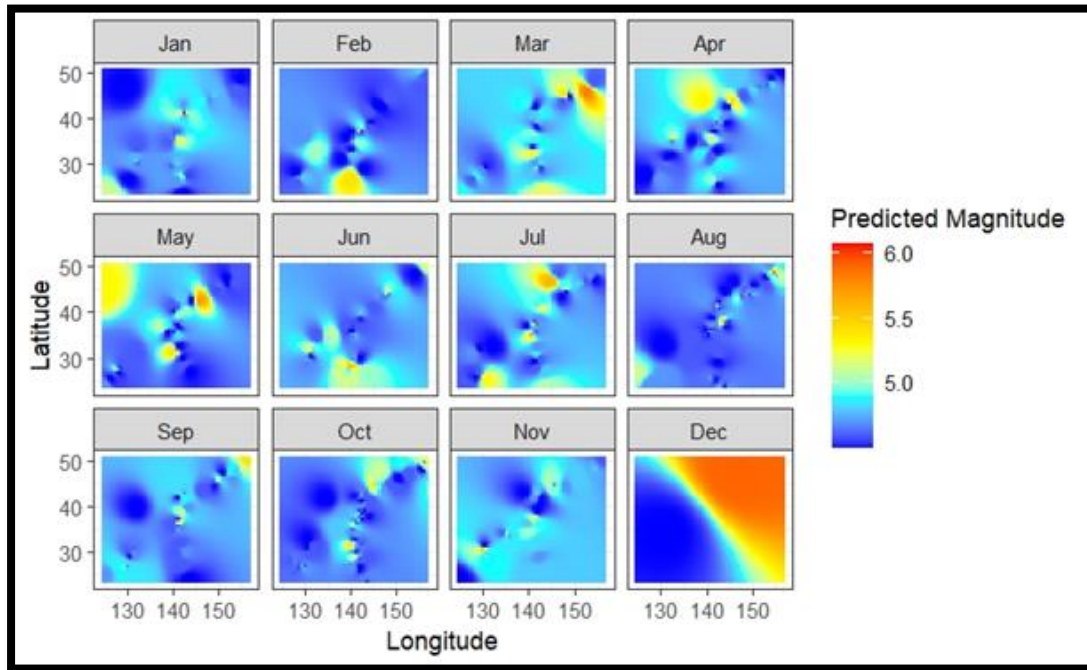


Figure 5. Faceted Map of Predicted Magnitude with IDW Method

#### Summary Statistics of Predicted Magnitudes (IDW)

Summary statistics for the predicted magnitudes revealed the following:

- Minimum: 4.50
- Median: 4.72
- Mean: 4.78
- Maximum: 5.90

These results demonstrate a relatively small range of earthquake magnitudes across the year, with a slight increase in magnitudes during certain months.

### 3.3 Gaussian Interpolation and Parameter Optimization

Gaussian interpolation, another spatial prediction technique, models the spatial influence using a Gaussian weighting function. This method provided a better fit for the dataset, as indicated by its' lower cross-validation error (0.117) with an optimal parameter ( $\theta = 2.1$ ).



### Summary Statistics of Predicted Magnitudes (Gaussian)

Summary statistics for Gaussian-predicted magnitudes showed:

- Minimum: 4.65
- Median: 4.76
- Mean: 4.79
- Maximum: 5.80

These results reflect a consistent range of earthquake magnitudes across the region, with localized areas experiencing higher magnitudes.

### Visualization of Predicted Magnitudes using Gaussian

The figure shows the spatial-temporal distribution of predicted earthquake magnitudes across Japan using Gaussian interpolation. Most months exhibit lower magnitudes (blue), while localized higher magnitudes (yellow to red) appear particularly in December. This highlights consistent earthquake activity with occasional high-magnitude events in specific regions.

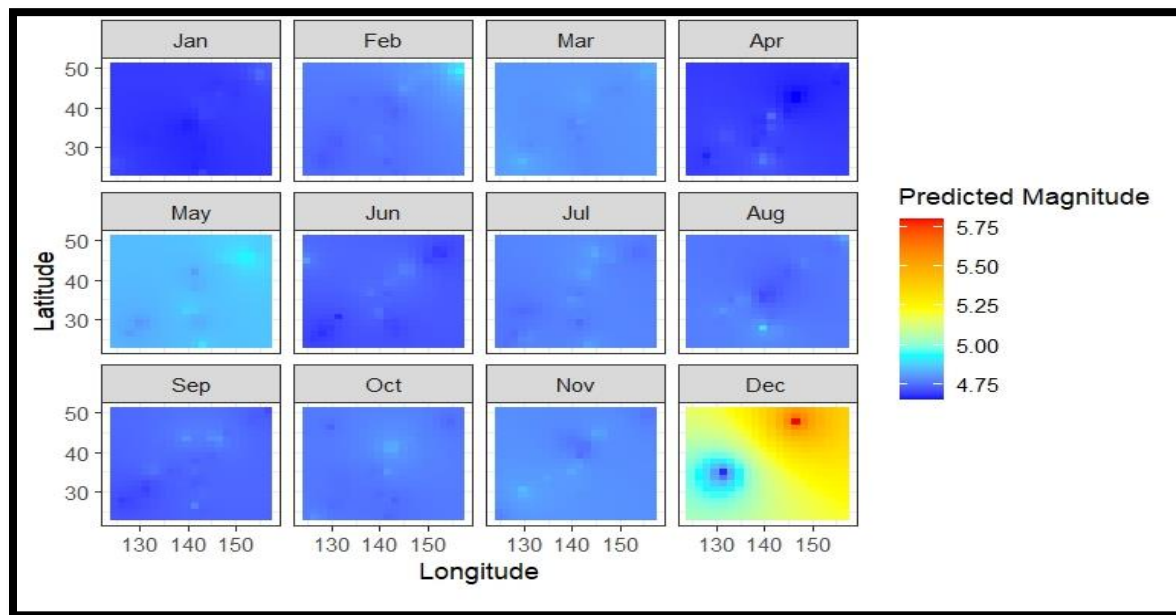


Figure 6. Faceted Map of Predicted Magnitude with Gaussian Method

### 3.4 Cross-Validation Analysis

LOOCV was employed to optimize the parameters for both interpolation methods. Gaussian interpolation consistently demonstrated lower errors compared to IDW, affirming its suitability for this dataset. The relationship between parameters ( $\alpha$  for IDW and  $\theta$  for Gaussian) and LOOCV errors is visualized in comparative plots.



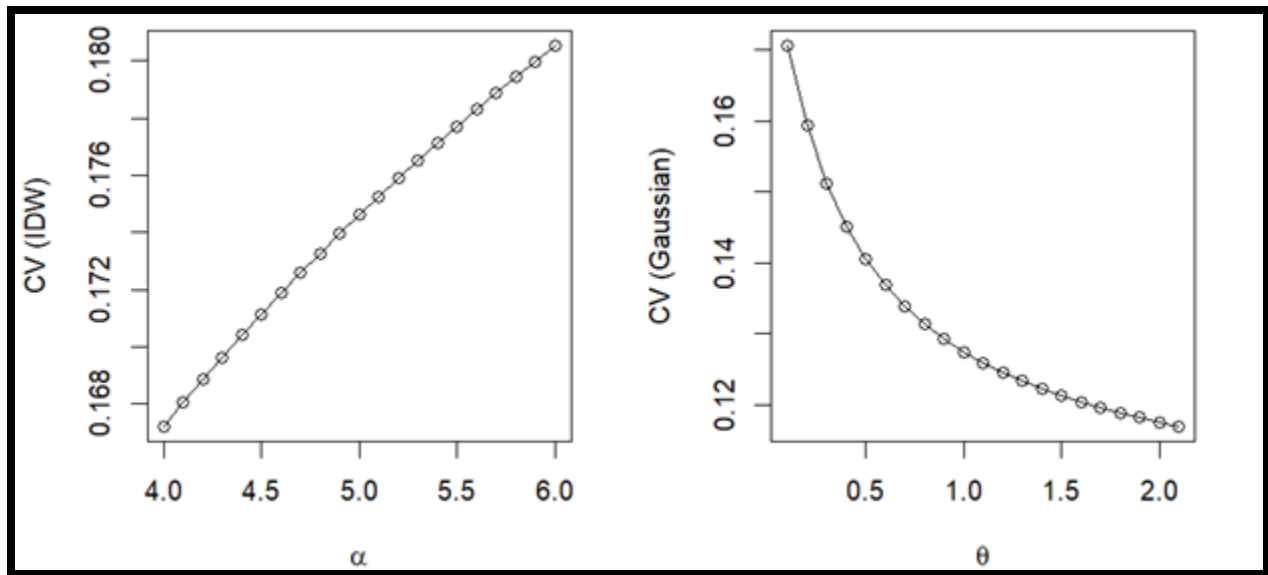


Figure 7. CV Error Comparison of IDW and Gaussian Methods

### 3.5 Comparison of Methods

Gaussian interpolation outperformed IDW, with a lower minimum cross-validation error (0.117 vs. 0.167). This indicates that Gaussian interpolation provides more accurate spatial predictions for the earthquake dataset.

## 4. Conclusion

This analysis has provided valuable insights into the spatio-temporal patterns of earthquake magnitudes in Japan. By employing robust statistical techniques, we observed key seasonal and monthly trends in the magnitudes of earthquakes. Spatial interpolation methods, particularly Gaussian interpolation, proved to be effective in predicting earthquake magnitudes, outperforming IDW due to its lower cross-validation errors and better handling of spatial dependencies.

These findings underscore the importance of spatio-temporal modeling for understanding and predicting earthquake characteristics. The combination of statistical rigor and visualizations contributes to a deeper understanding of seismic activity, providing essential tools for disaster preparedness and risk management. Future work can expand on these models by integrating additional seismic variables or exploring real-time prediction applications.