USER GUIDE

Guide to make full use of the R shiny app

Climateviz

Contents

Introduction

This document outlines the user guide for the Shiny app built by ISSS-608 2024/2025 January Term, developed by SG ClimateViz - Group 7. It is important to read through the user guide first so that users will be able to understand the visual analytics displays, which aim to facilitate user understanding and draw insights on how climate change affects weather parameters in Singapore, and how it impacts our daily lives through energy consumption and public health without the need to use statistical software or rely on R or Python.

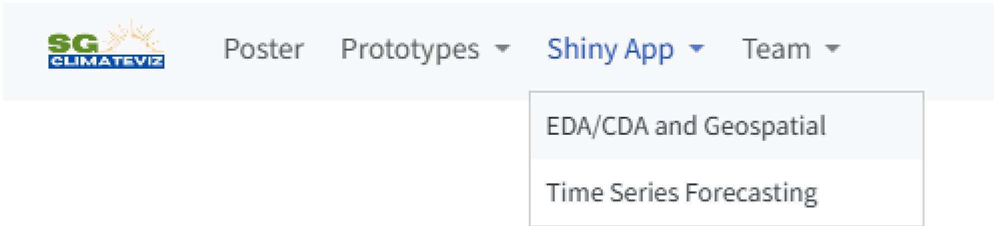
With increasing rainfall and higher mean temperatures, there has been a rise in electricity consumption in households and the number of dengue cases reported in recent years.

As described on our project website, the Shiny app is built using data from the past 10 years on weather parameters, household electricity consumption, and reported dengue cases. To ensure a smooth user experience, our group has aggregated the data in both monthly and yearly formats for all modules—namely Exploratory Data Analysis (EDA), Confirmatory Data Analysis (CDA), time series analysis, and geospatial analysis.

We have also included a description in each of the tabs to better facilitate the user experience.

1.0 Navigating to the shiny app page:

1. Go to the project website at the following link: https://isss608jan25group7.netlify.app/
2. Hover over to “Shiny App”
3. Select the desired visual analytic module which you wish to view
    a. “EDA /CDA and Geospatial” or
    b. “Time Series Forecasting”

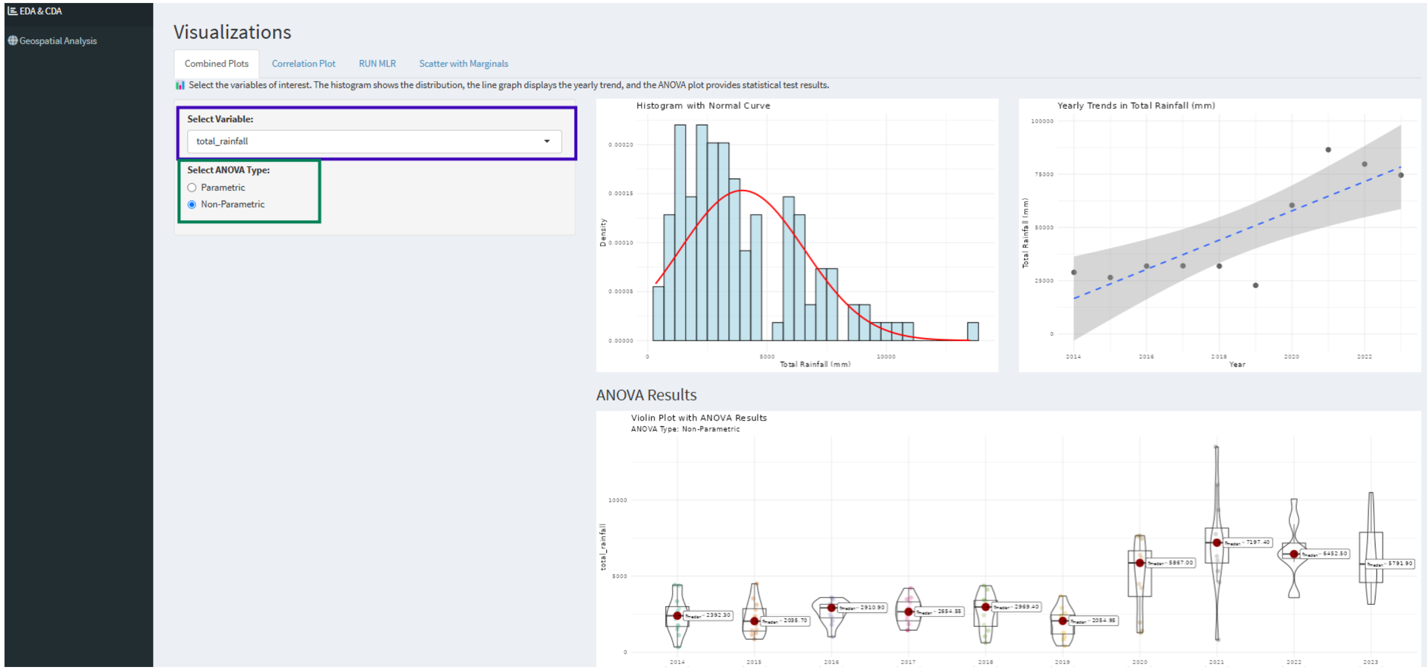


2.0 Exploratory Data Analysis (EDA) and Confirmatory Data Analysis (CDA)

2.1 Combined Plots Tab

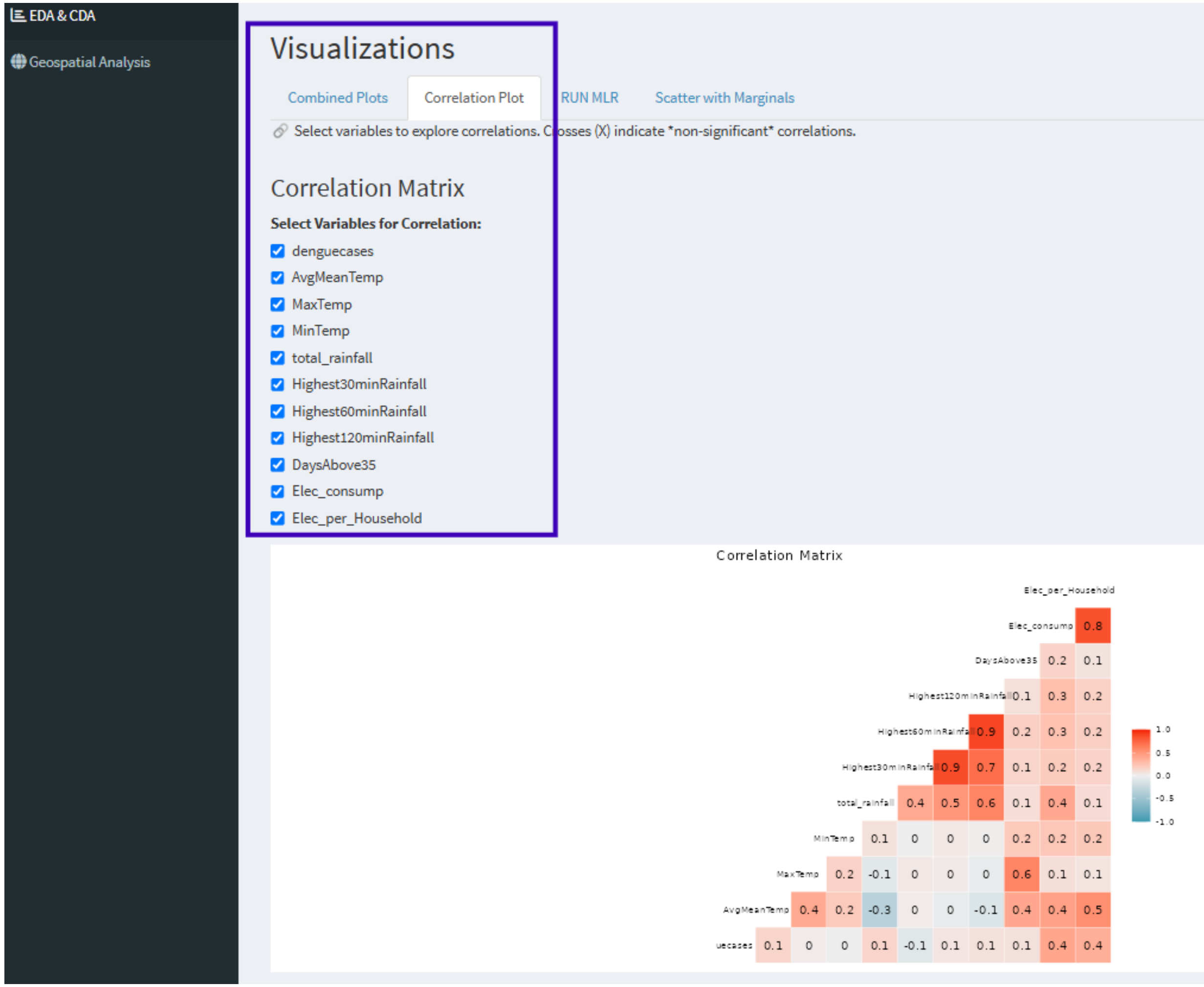
This **Combined Plot** tab shows the histogram with the normal distribution curve on the top right-hand side, the scatter plot of the aggregated yearly trend data on the top left-hand side, and the analysis of variance with a violin plot displayed at the bottom of the tab.

You may choose the weather parameters from the dropdown menu and select the type of ANOVA test you would like to perform on the parameters in the right panel.



2.2 Correlation Plot

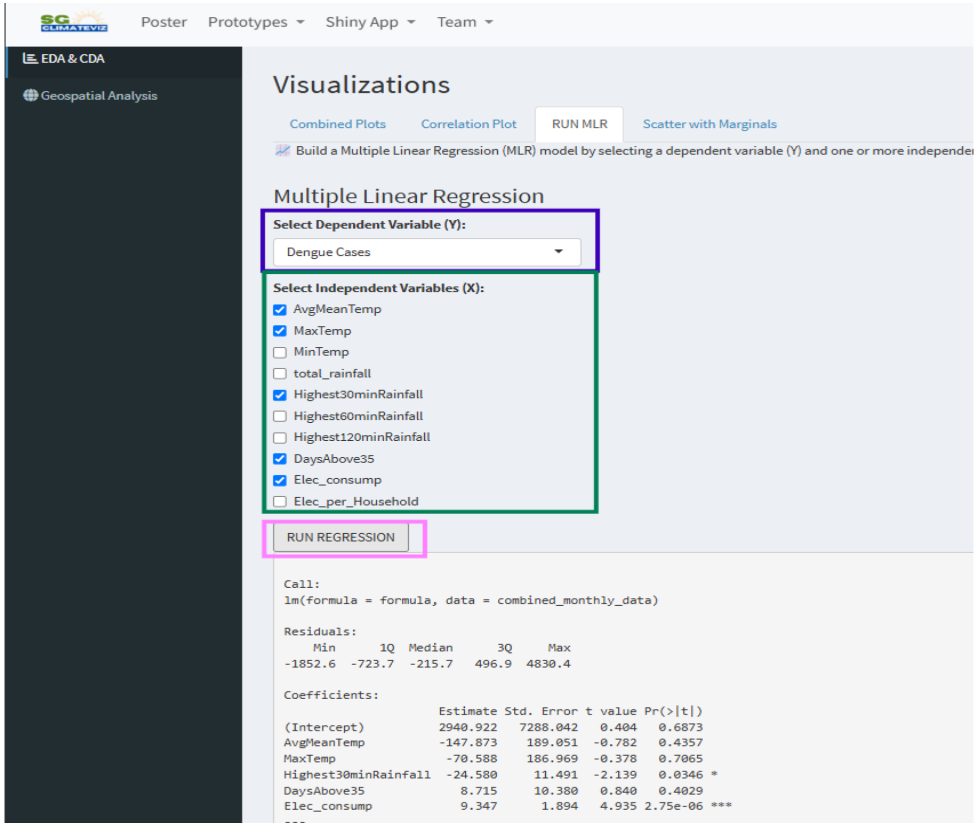
The **Correlation Plot** tab displays the correlation plot. You may select variables of interest, or choose all available variables in the dataset, to explore the correlations between them.



2.3 Run MLR (multi-linear regression model)

The **Run MLR** tab allows you to run Multiple Linear Regression (MLR) as part of the CDA module. The benefit of doing so is that it helps you gain insights into which weather parameters affect dengue cases and electricity consumption.

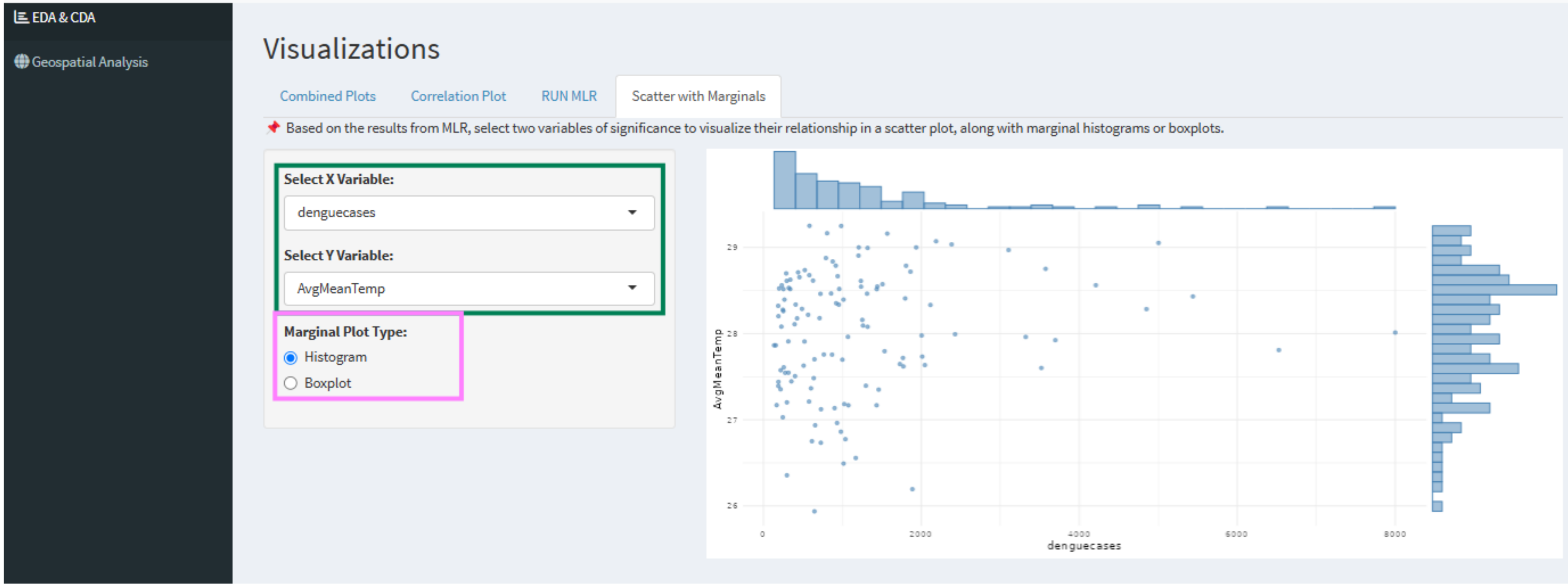
1. Select the Y variable (dependent), namely **Dengue Cases** or **Household Electricity Consumption**.
2. Select the X variables (independent), you want to include in the MLR model.
3. Click **Run Regression**.



2.4 Scatter plots with Marginal

The Scatter Plot with Marginal tab displays the Scatter plot with either a Histogram or Boxplot on the side.

1. After reviewing the outcome of the MLR model, select your variables of interest as the X and Y variables to carry out bivariate analysis.
2. You may change the marginal plot to either a Histogram or Boxplot to be displayed alongside the scatter plot.



3.0 Geospatial Analysis

The geospatial analysis module aims to provide users identify meaningful local patterns and understand how weather conditions may be spatially concentrated across Singapore.

3.1 Introduction to Spatial Interpolation

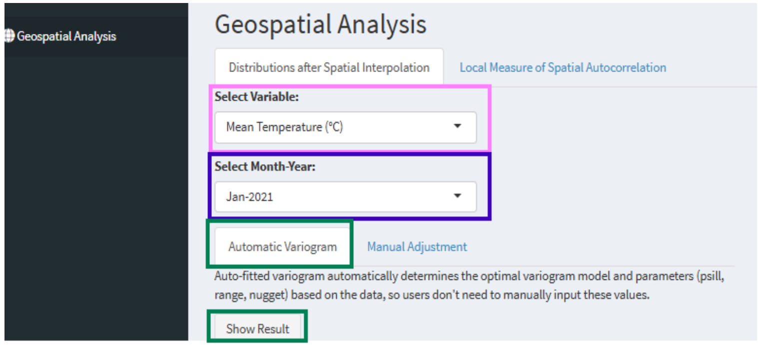
Spatial interpolation is the process of predicting values for unmeasured locations based on known values from nearby locations.

Kriging is a method of spatial interpolation that uses statistical models to predict the value of a variable at unmeasured points, considering the spatial correlation between data points.

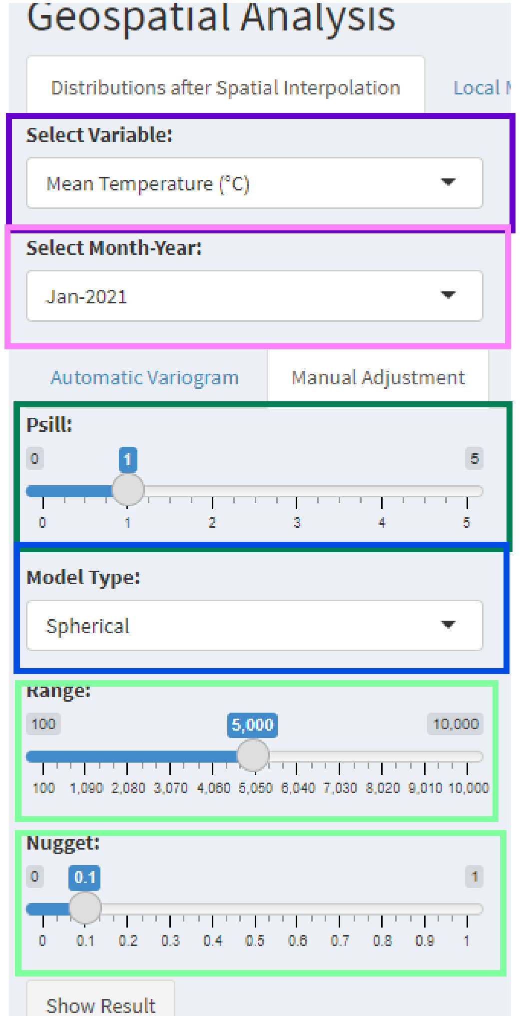
3.2 Distribution After Spatial Interpolation

The **Distributions after Spatial Interpolation** tab shows how a selected weather variable—like mean temperature—is spread across different areas of Singapore for a specific month and year.

1. Select the weather parameters of interest
2. Select the month-year.
3. If you select Automatic Variogram, the system will automatically determine the optimal variogram model and parameters (psill, range, nugget) based on the data, so users don't need to manually input these values.



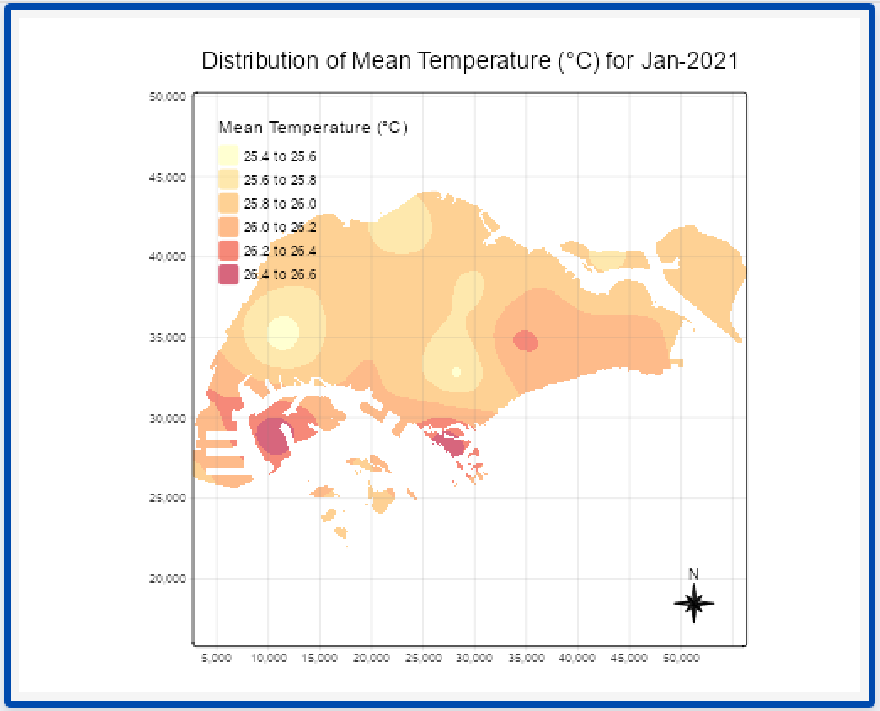
4. If you select Manual Adjustment, you can set the Psill level, Model Type, Range, and Nugget values manually.

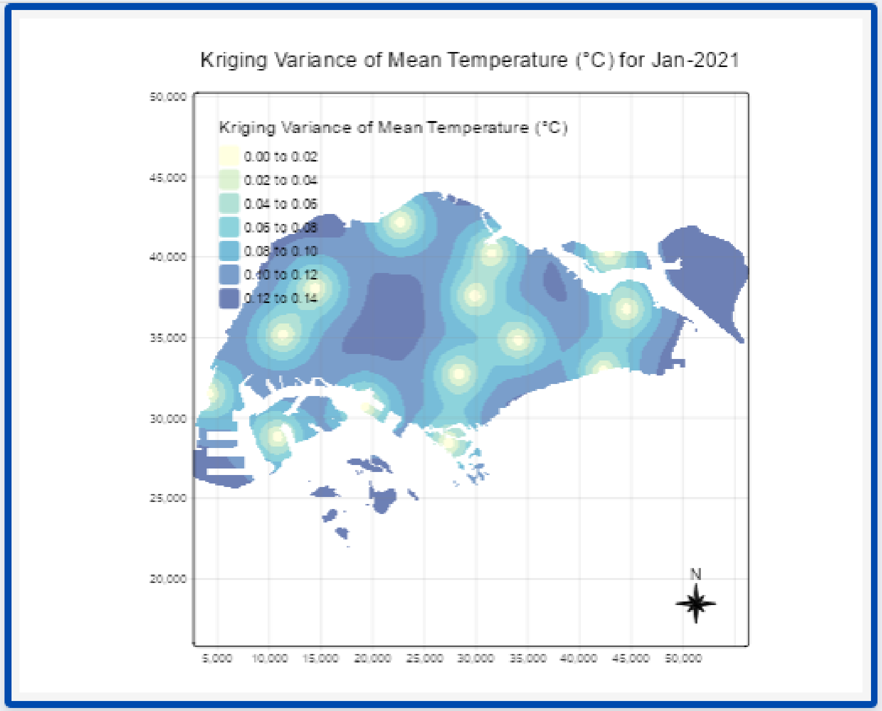


3.2.2 How to Interpret the Plots

The first plot represents the predicted values (e.g., temperature, rainfall, or wind speed) for the selected variable at different locations.

The second plot shows the Kriging variance, which represents the uncertainty of the predictions. Higher variance indicates less confidence in the prediction.

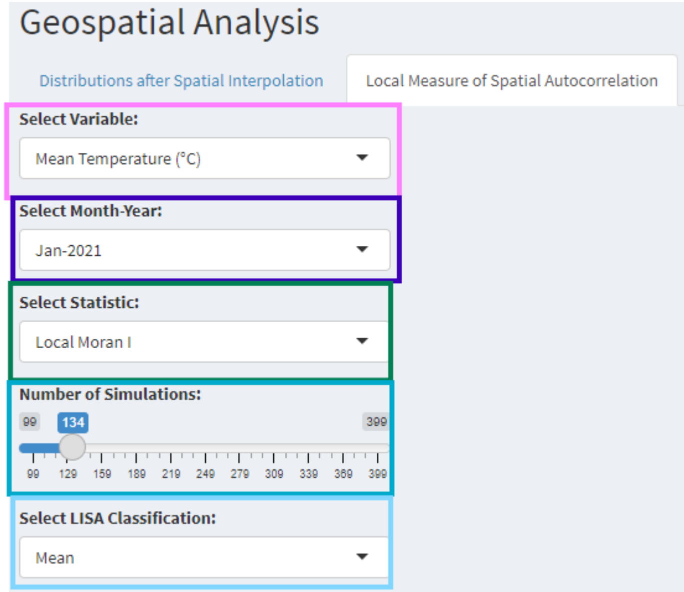




3.3 Local Measure of Spatial Autocorrelation

This tab allows users to explore local spatial patterns and clustering of a selected weather parameter using **Local Indicators of Spatial Association (LISA)** and **Local Moran's I**.

1. Select Variable: Choose the weather parameter of interest (Mean Temperature).
2. Select Month-Year: Pick the month and year for the analysis.
3. Select Statistic: Choose the statistic to compute (Local Moran's I, P-Value, Std Deviation, Expectation).
4. Number of Simulations: Adjust the number of simulations used to calculate statistical significance. A higher number provides more reliable results but may take longer to compute.
5. Select LISA Classification: Choose the classification method (e.g., Mean) for detecting local spatial clusters.

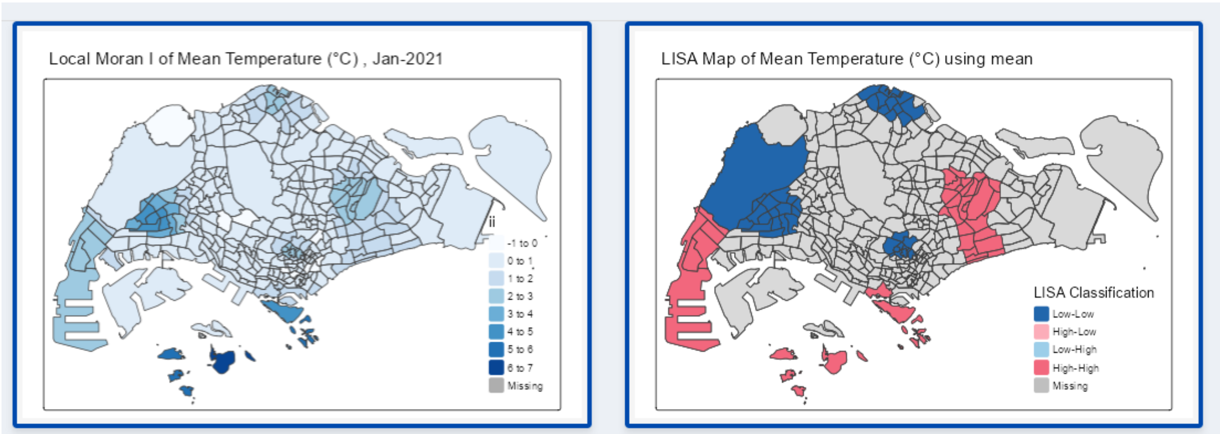


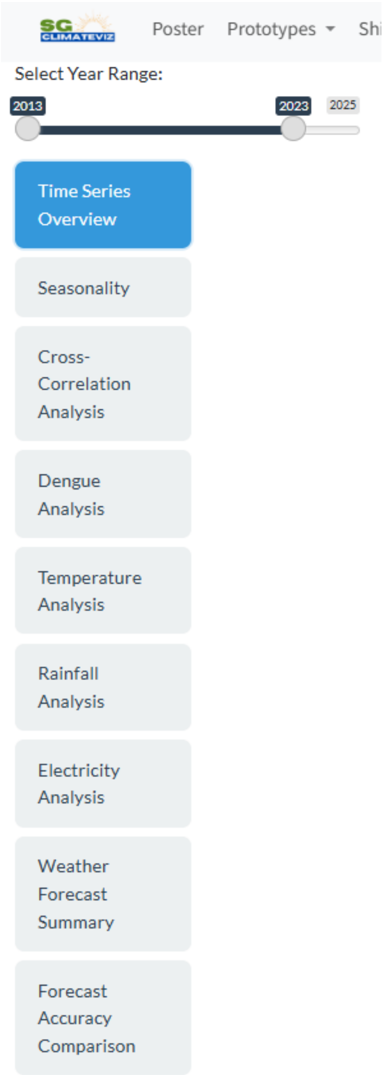
3.3.1 Interpretation of Maps of Local Measure of Spatial Autocorrelation:

The first map displays the selected statistic (e.g., Local Moran's I, P-value, Std Deviation, etc.) for the chosen variable (e.g., Total Rainfall, Mean Temperature, Mean Wind Speed). Each colour in the map represents the value of the statistic for a particular geographic region, where darker or lighter colours indicate higher or lower values of the statistic, respectively.

The second map is the LISA (Local Indicators of Spatial Association) map, which shows significant clusters of similar values for the selected variable. LISA highlights areas where high values are clustered together, and similarly, low values are grouped in specific locations:

- **High-High**: Areas with high values surrounded by other high values.
- **Low-Low**: Areas with low values surrounded by other low values.
- **High-Low**: Areas with high values surrounded by low values.
- **Low-High**: Areas with low values surrounded by high values.





4.0 Time Series Analysis

For the **Time Series Analysis** module, the aim is to explore seasonality and trends, and to forecast changes in weather parameters and their relationship to the impact on our daily lives—specifically in terms of public health (represented by dengue cases) and electricity consumption.

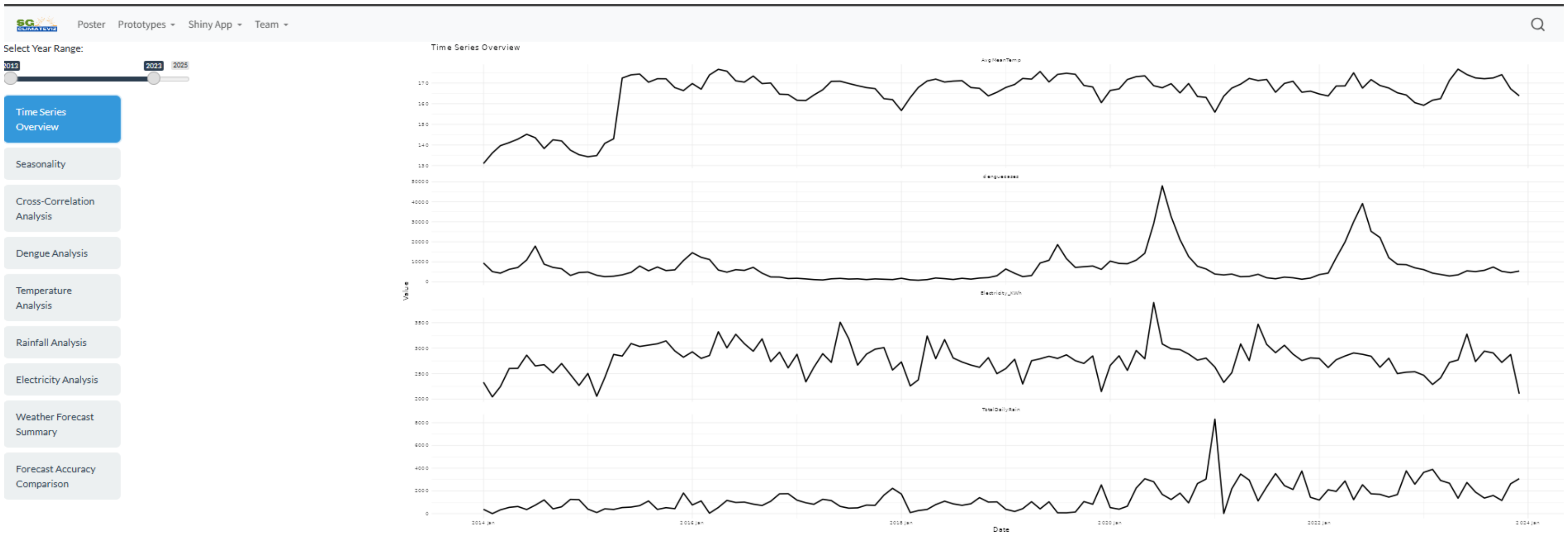
Users may select the **year range** using the time slider located at the top of the page. Like the previous two modules, the data is available from **2013 to 2025** for weather parameters, namely **average mean temperature**, **total rainfall**, **number of dengue cases**, and **electricity consumption**.

Users can click through the individual tabs on the right panel to explore the full range of time series features available.

It is important to note that the selected year range will be applied across all tabs within the module.

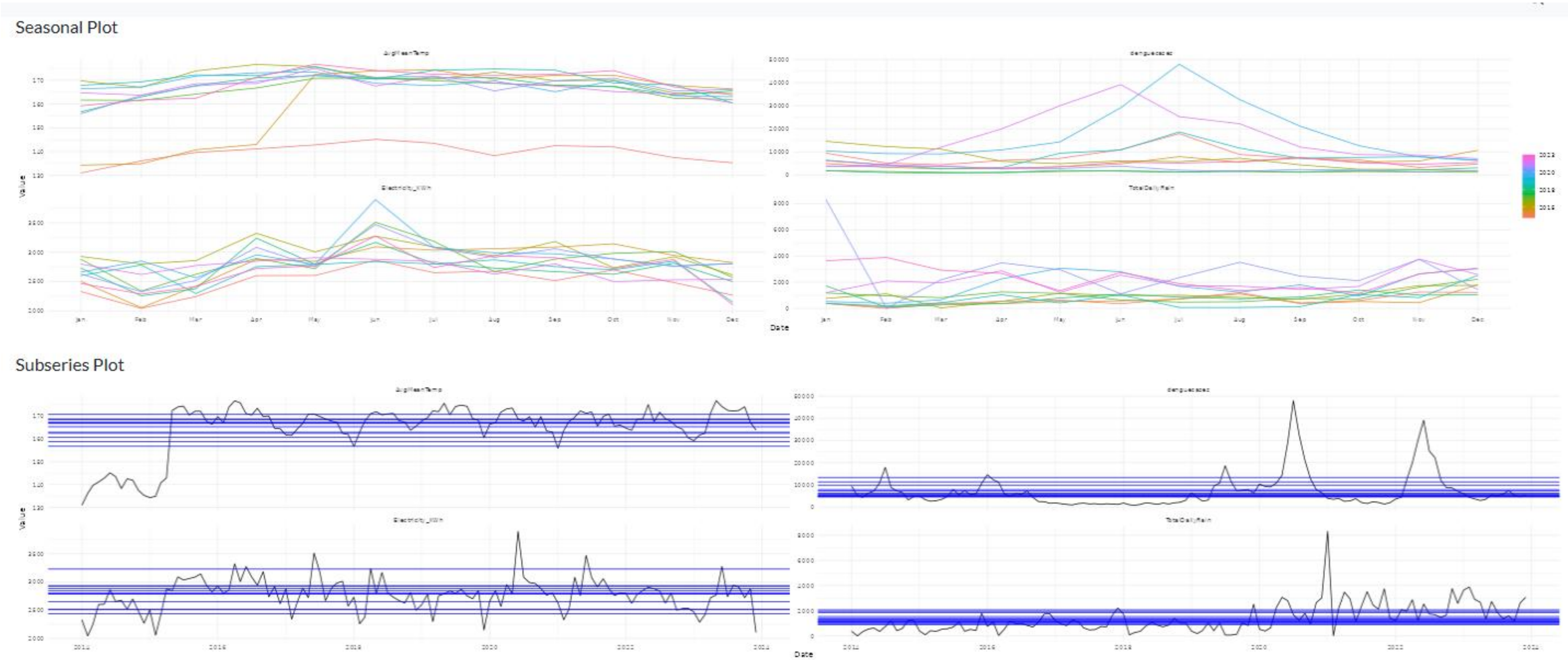
4.1 Time Series Overview

The Time Series overview provides a general overview of the selected time series variables across the chosen year range. It allows users to get a quick sense of how weather parameters namely average mean temperature, total daily rain, dengue cases, and electricity consumption have changed over time.



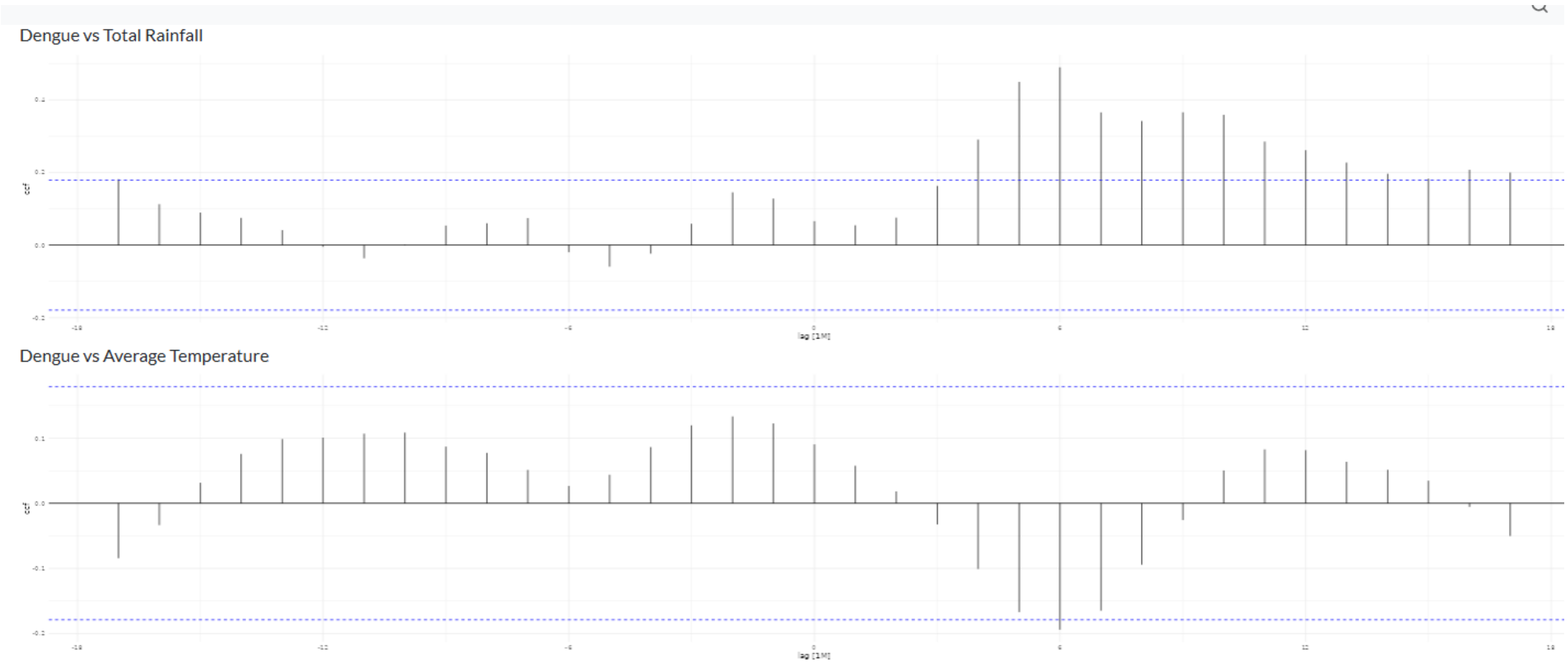
4.2 Seasonality

The Seasonality tab explores the seasonal patterns of the selected variables. Users can visually identify recurring monthly or yearly trends that may exist in dengue cases, electricity usage, or weather parameters such as average mean temperature, total daily rain.



4.3 Cross-Correlation Analysis

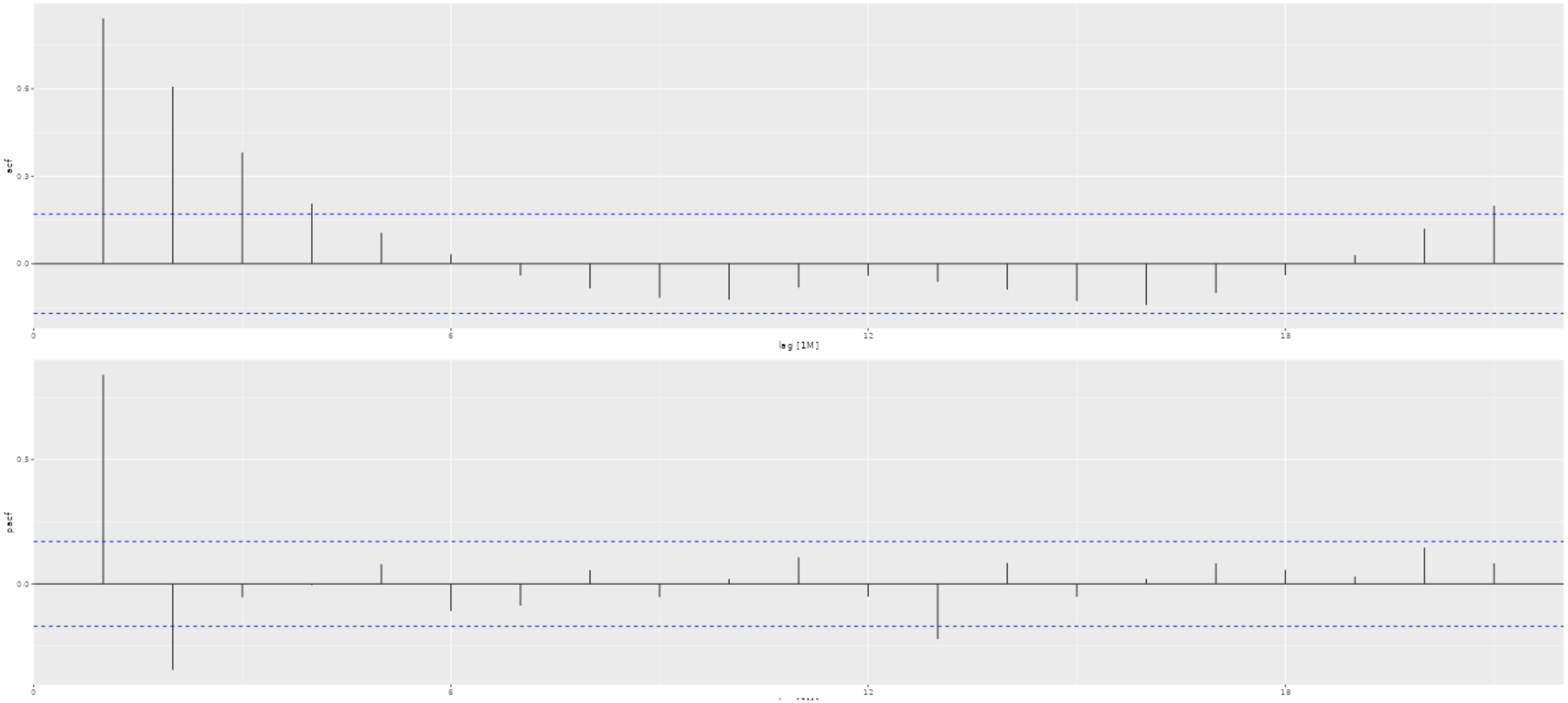
This tab examines the relationship between two different variables over time, such as whether changes in average mean temperature and total rainfall are followed by changes in dengue cases or electricity usage. It helps identify lag effects or delayed relationships.



4.4 Dengue Analysis

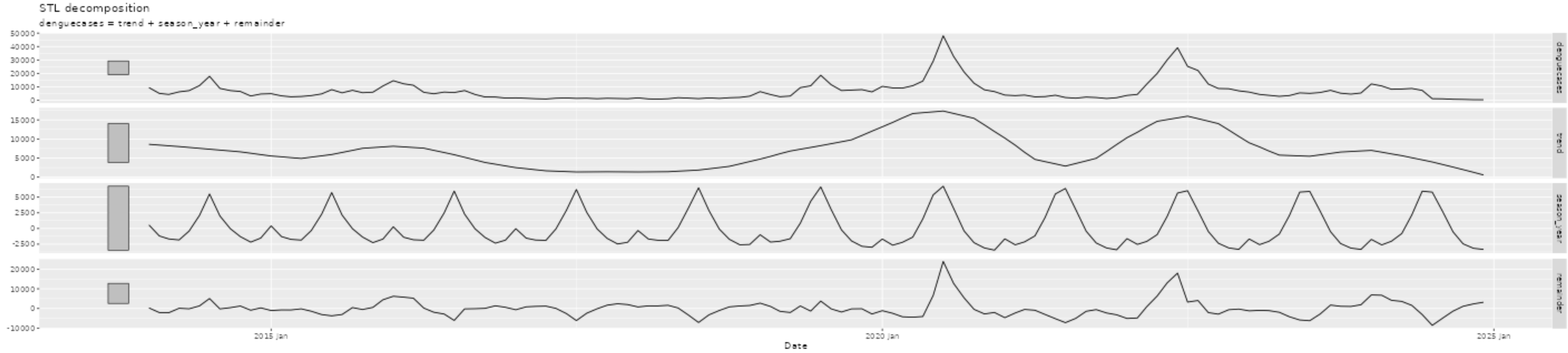
The Dengue Analysis tab focuses specifically on the time series of dengue cases. Users can explore trends, detect seasonal spikes, and observe how cases have fluctuated across the selected years through Time Series Decomposition using methods such as ACF, STL, and Classical Decomposition, as well as perform Forecasting using ETS and ARIMA models.

ACF & PACF

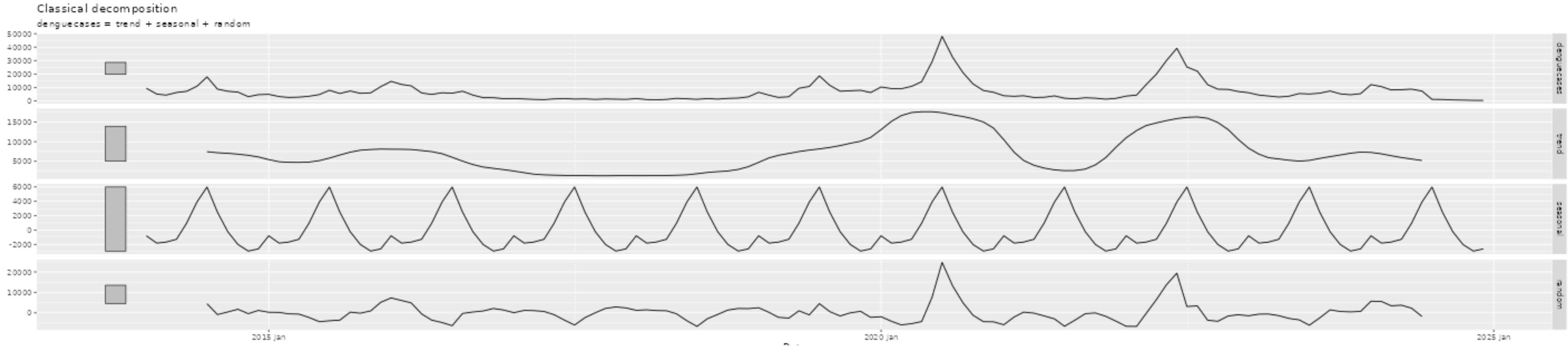


lag [1M]

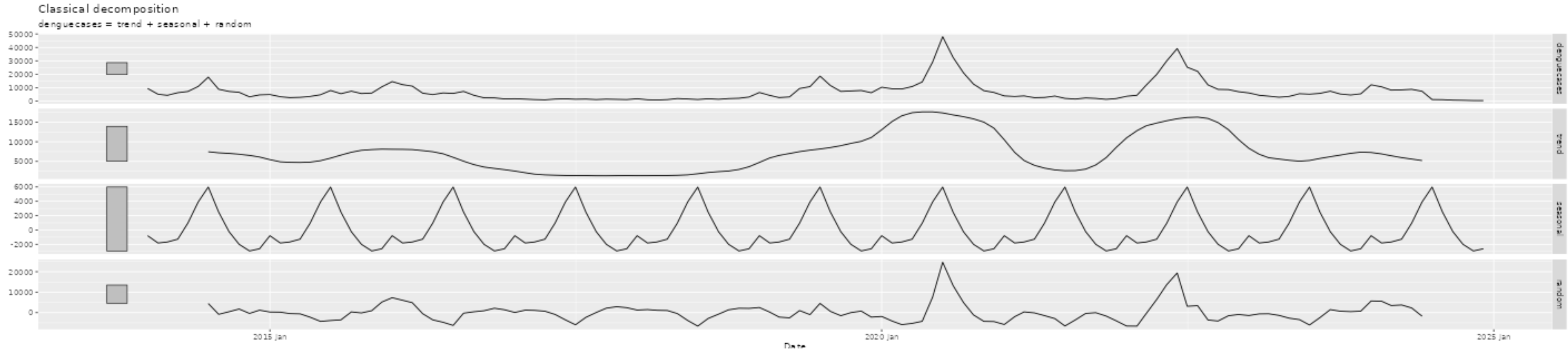
STL Decomposition



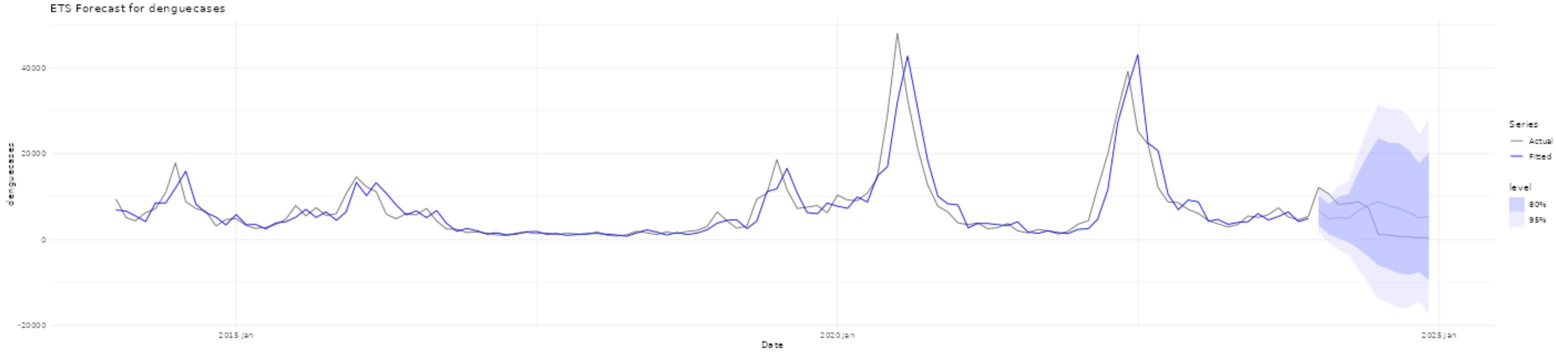
Classical Decomposition



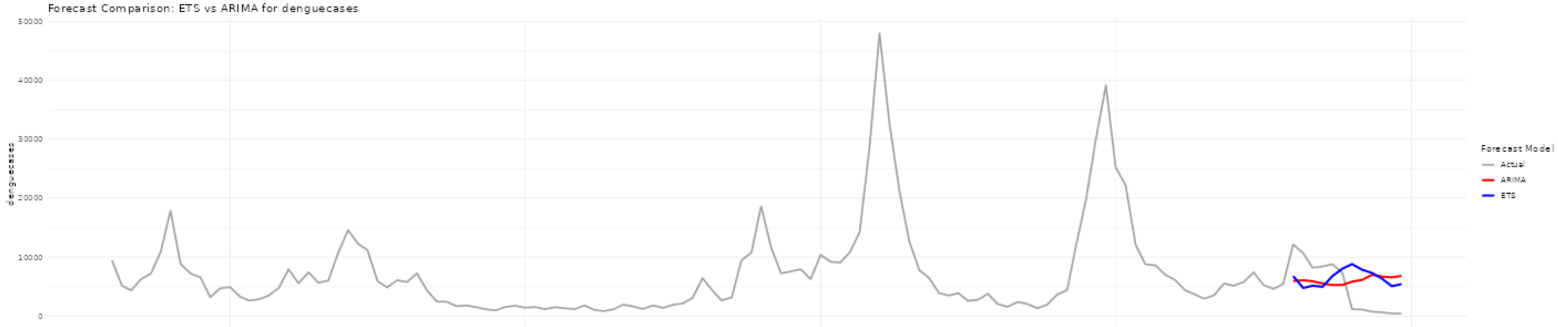
Classical Decomposition



ETS Forecast



ETS vs ARIMA Forecast Comparison



The visual displays for the other variables such as temperature, rainfall and electricity consumption will be the same.

4.5 Temperature Analysis

The **Temperature** Analysis tab focuses specifically on the time series of **Temperature** data. Users can explore trends, detect seasonal spikes, and observe how temperature has fluctuated across the selected years through Time Series Decomposition using methods such as ACF, STL, and Classical Decomposition, as well as perform Forecasting using ETS and ARIMA models.

4.6 Rainfall Analysis

The **Rainfall** Analysis tab focuses specifically on the time series of **Rainfall** Analysis. Users can explore trends, detect seasonal spikes, and observe how rainfall has fluctuated across the selected years through Time Series Decomposition using methods such as ACF, STL, and Classical Decomposition, as well as perform Forecasting using ETS and ARIMA models.

4.7 Electricity Consumption Analysis

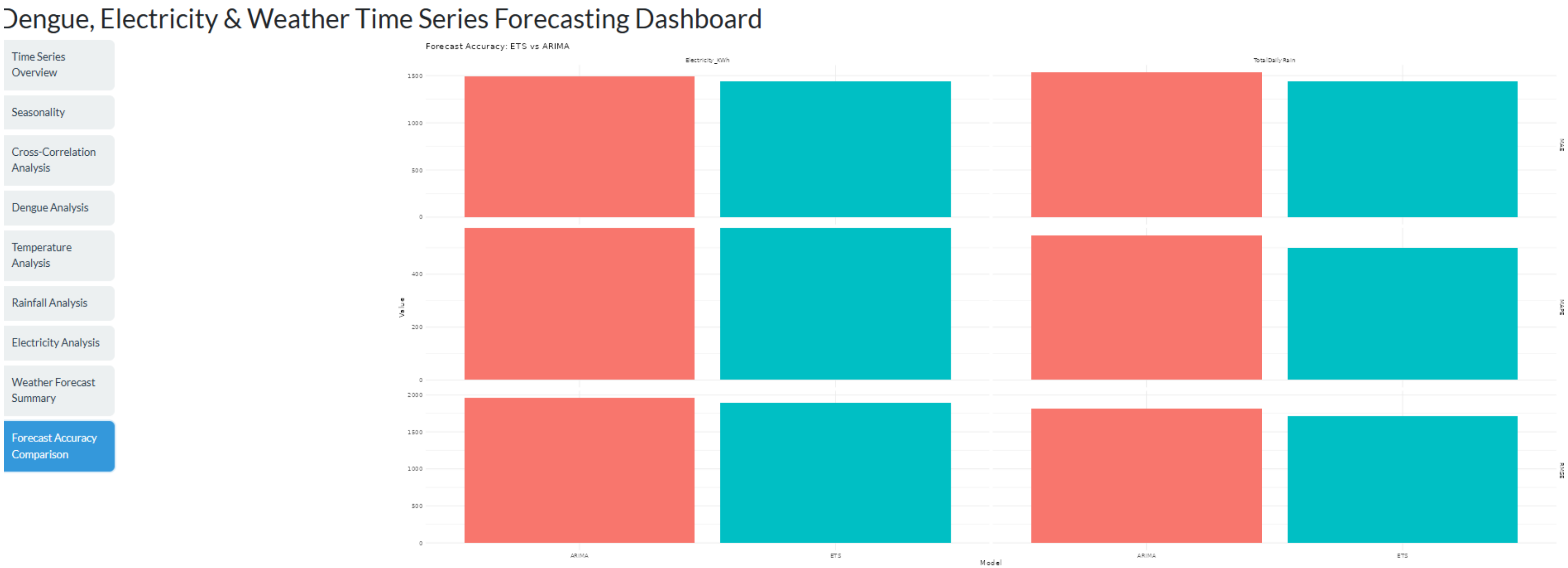
The Electricity Consumption Analysis tab focuses specifically on the time series of Electricity Consumption. Users can explore trends, detect seasonal spikes, and observe how have fluctuated across the selected years through Time Series Decomposition using methods such as ACF, STL, and Classical Decomposition, as well as perform Forecasting using ETS and ARIMA models.

4.8 Weather Forecast Summary

This tab provides the **forecast results** for key weather parameters such as **temperature** and **rainfall**, along with **electricity consumption**. Users can view projected trends for the next 12 months based on past data using **ETS (Exponential Smoothing)** model outputs.

4.9 **Forecast Accuracy** Comparison

This tab allows users to compare the accuracy of different forecasting models (ARIMA, ETS). The bar plot presents a side-by-side comparison of forecast accuracy between ETS and ARIMA models for rainfall and electricity consumption, evaluated using three common error metrics: RMSE, MAE, and MAPE.



End of User Guide