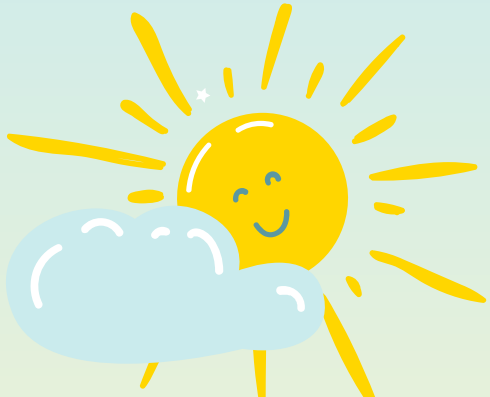
Visual Analytics for Singapore's Weather

User Guide

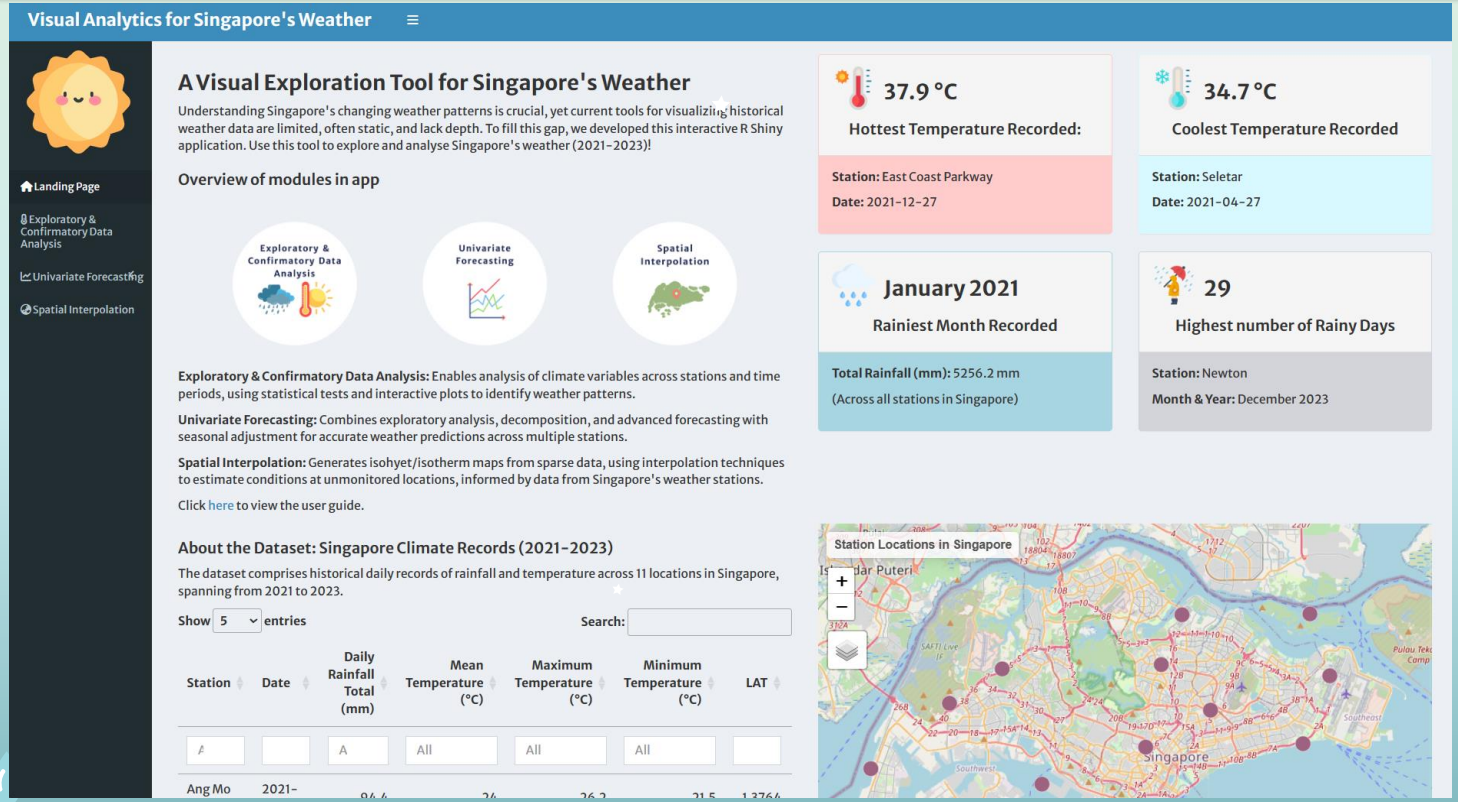
01

Landing page



Landing Page

This page provides an overview of the application. It describes the key modules and the dataset used. It also shares some interesting information about the dataset like the hottest temperature recorded in the dataset. Lastly, there is a map showing the location of the weather stations from which the data was collected from.

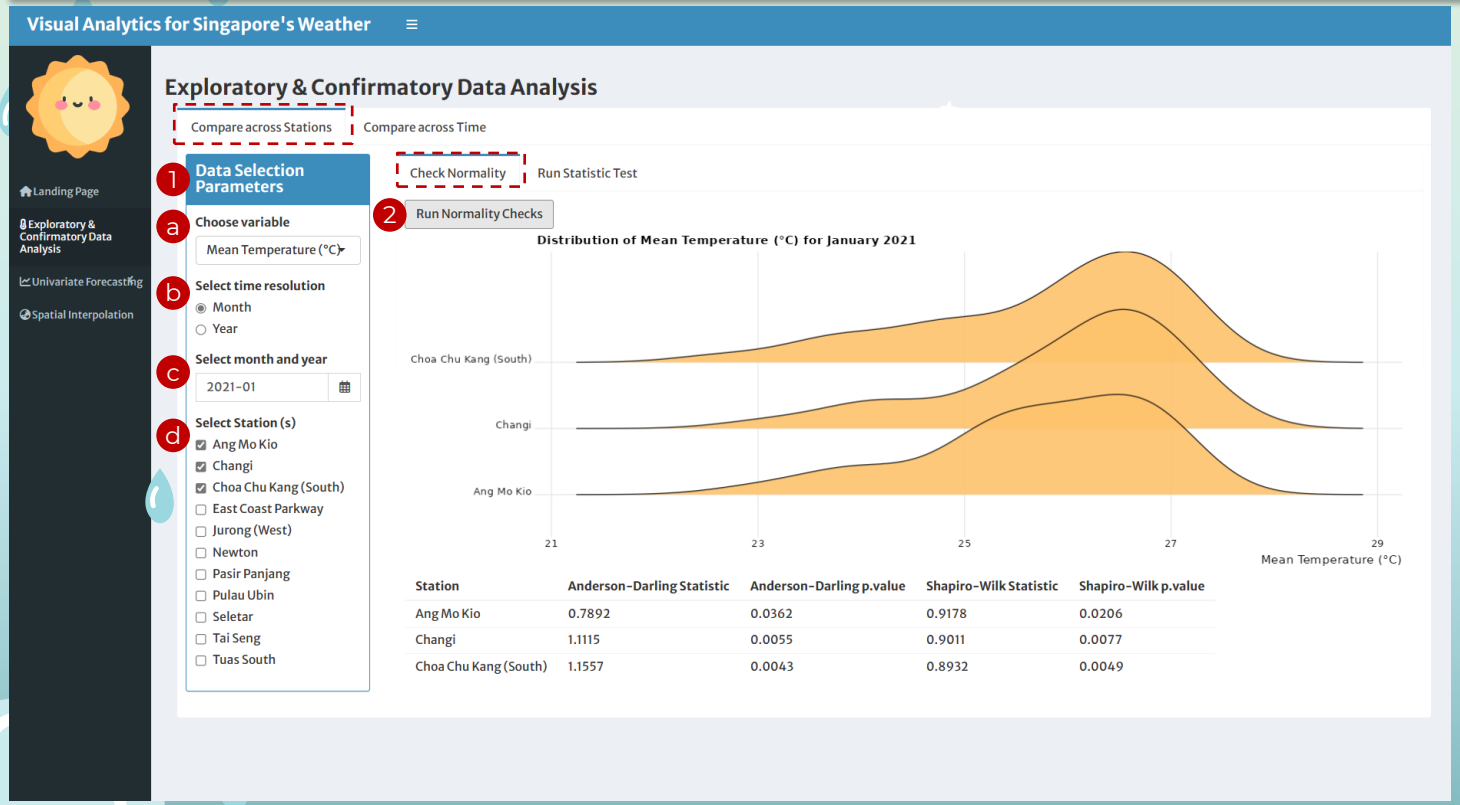


02

Exploratory & Confirmatory Data Analysis

Exploratory & Confirmatory Data Analysis

Compare across Stations: This page allows users to conduct normality assumption test of daily rainfall or temperature across different stations. The outcomes of the test guide the selection of suitable statistical methods for further confirmatory data analysis.



Tab: Check normality

1. Data Selection Parameters:

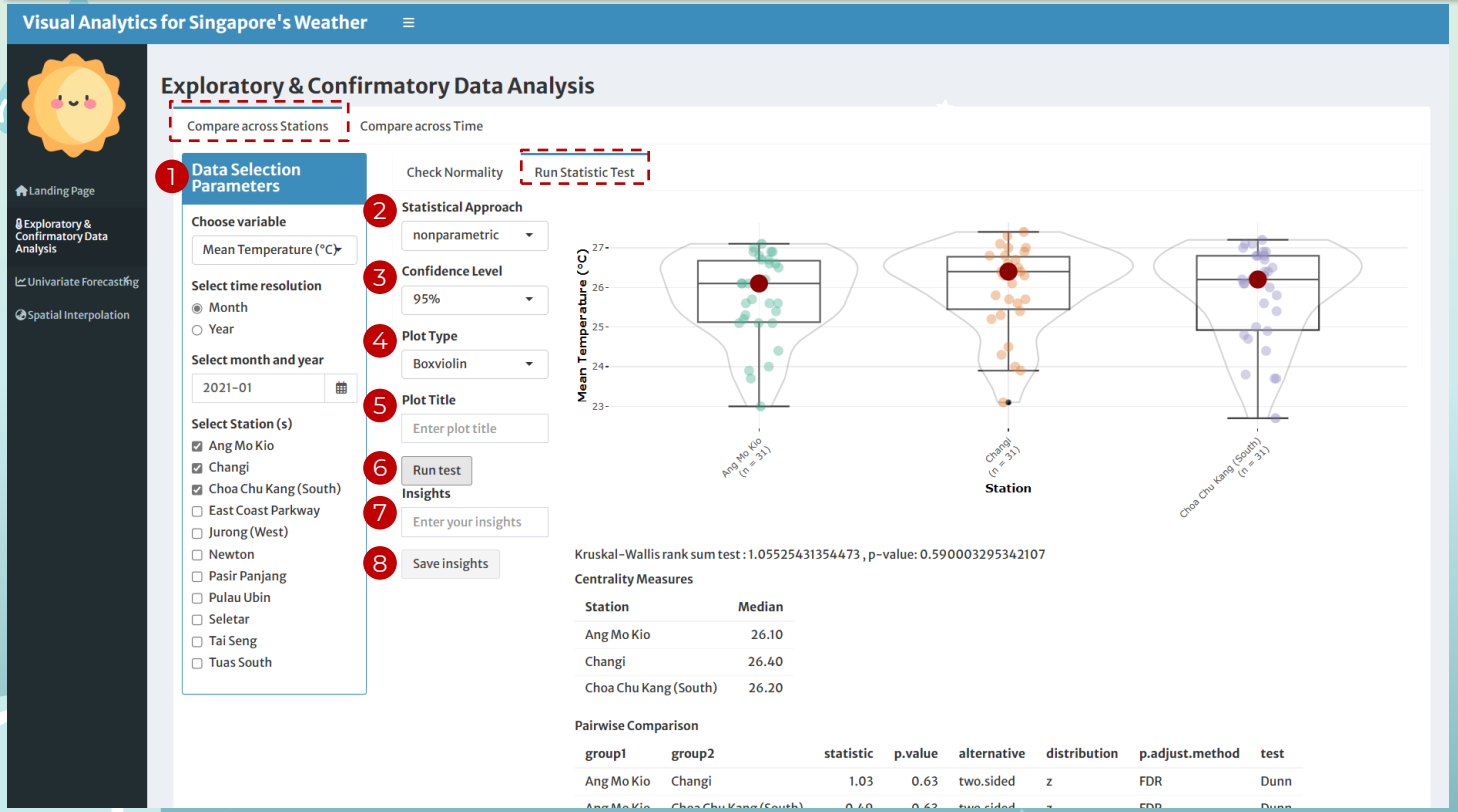
a) Use the dropdown box to select a desired variable
b) Use the radio button to select time resolution
c) Use the month and year picker to specify the period
d) Use the checkbox to select multiple stations for the normality test

2. Click 'Run normality checks' button to display the distribution plot and normality test results.

Tips: If at least one result shows **p-value < 0.05**, it indicates that the sample failed to confirm normality. Therefore, non-parametric test statistic will be used for confirmatory data analysis.

Exploratory & Confirmatory Data Analysis

Compare across Stations : This page allows users to compare daily rainfall or temperature across different stations. The comparison will enable users to identify patterns, trends, and irregularities over time that facilitate a deeper understanding of weather dynamics in Singapore.



Tab: Run Statistic Test

1. Data Selection Parameters:
The input for this section automatically aligns with what user selected for the check normality tab

2. Use the dropdown boxes to select the statistical approach for the plot

3. Use the dropdown boxes to select the confidence level

4. Use the dropdown boxes to select a plot type

5. Input the plot title

6. Click on 'Run test' button to generate the plot

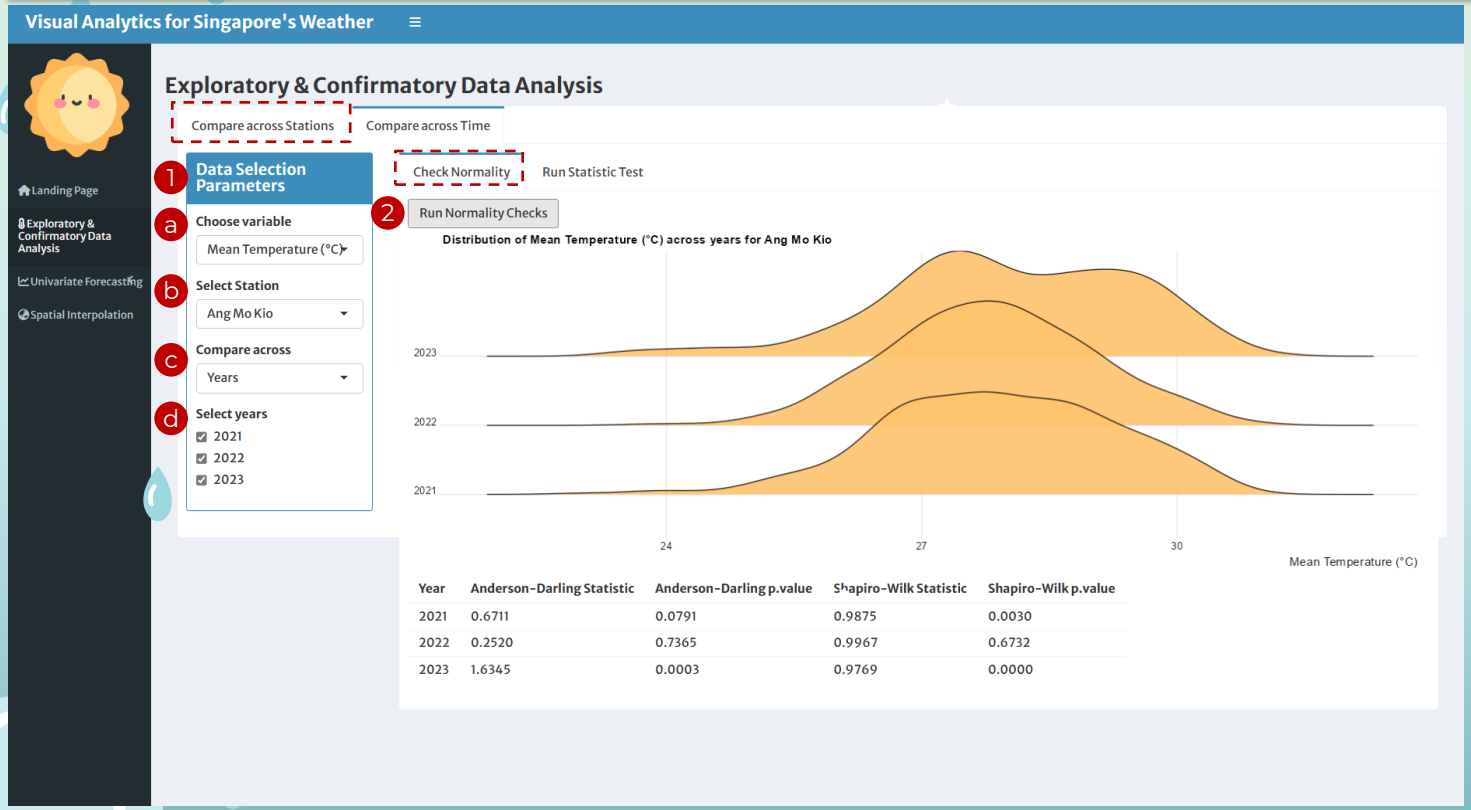
7. Input the insights

8. Click on 'Save insights' button to generate the insight text box under the graph

Tips: At the confidence level 95%, if p-value of the statistic test is less than 0.05, it implies that the mean of selected variables for different stations are not equal.

Exploratory & Confirmatory Data Analysis

Compare across Time : This page allows users to conduct normality assumption test of daily rainfall or temperature across different time-period. The outcomes of the test guide the selection of suitable statistical methods for further confirmatory data analysis.



Tab: Check normality

1. Data Selection Parameters:

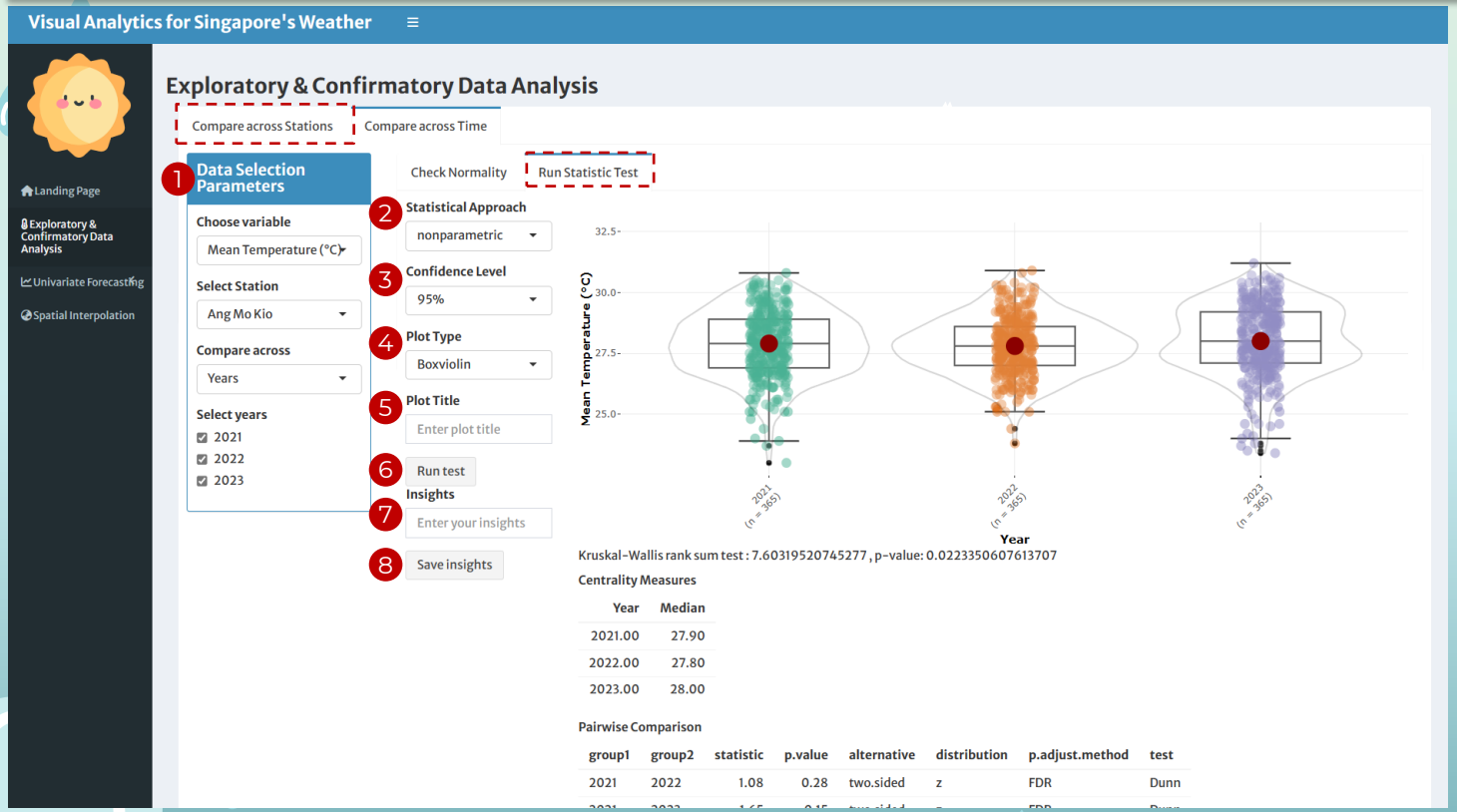
a) Use the dropdown box to select a desired variable
b) Use the dropdown box to select a station for the normality test
c) Use the dropdown box to select one type of period (years, months, months for specified a year, months of different years)
d) Use checkbox to select a specified period.

2. Click 'run normality checks' button to display the distribution plot and normality test results.

Tips: If at least one result shows **p-value < 0.05**, it indicates that the sample failed to confirm normality. Therefore, non-parametric test statistic will be used for confirmatory data analysis.

Exploratory & Confirmatory Data Analysis

Compare across Time : This page allows users to compare daily rainfall or temperature across different time-period. The comparison will enable users to identify patterns, trends, and irregularities over time that facilitate a deeper understanding of weather dynamics in Singapore.



Tab: Run Statistic Test

1. Data Selection Parameters:
The input for this section automatically aligns with what user selected for the check normality tab

2. Use the dropdown boxes to select the statistical approach for the plot

3. Use the dropdown boxes to select the confidence level

4. Use the dropdown boxes to select a plot type

5. Input the plot title

6. Click on 'Run test' button to generate the plot

7. Input their own insights

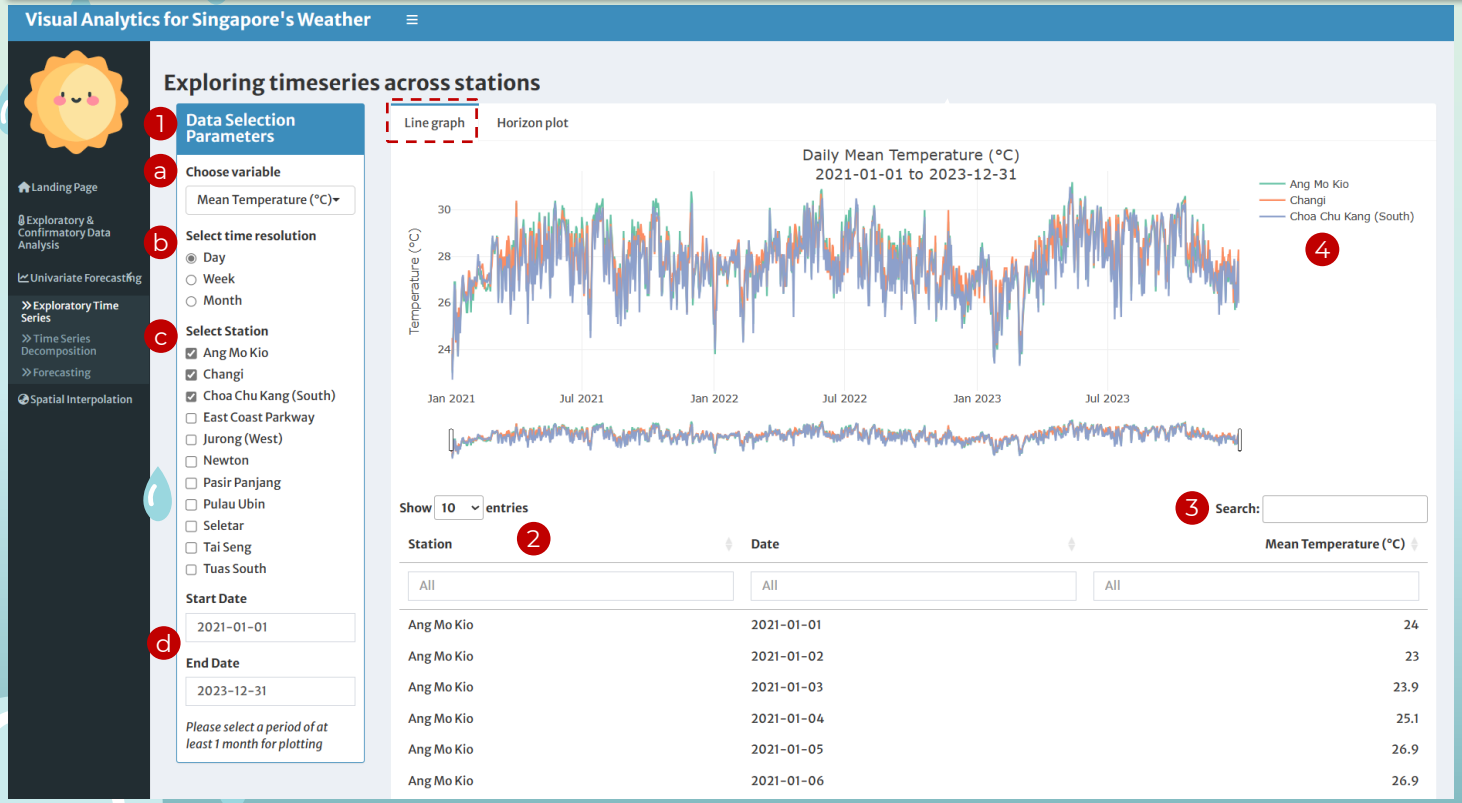
8. Click on 'Save insights' button to generate the insight text box under the graph

Tips: At the confidence level 95%, if p-value of the statistic test is less than 0.05, it implies that the mean of selected variables for different time-periods are not equal.


03
Univariate Forecasting

Univariate Forecasting: Exploratory Time Series

Line graph: This exploration page enables users to dynamically compare weather data from different stations on the same line graph, supplemented by a detailed data table for in-depth analysis.



1. Data Selection Parameters:

a) Use the dropdown box to select a desired variable
b) Use the radio button to select the time resolution
c) Use the checkbox to select multiple stations for comparison
d) Use the date picker to specify the start date and end date

2. Apply a filter to view data as stations selected in (c) for a specific station

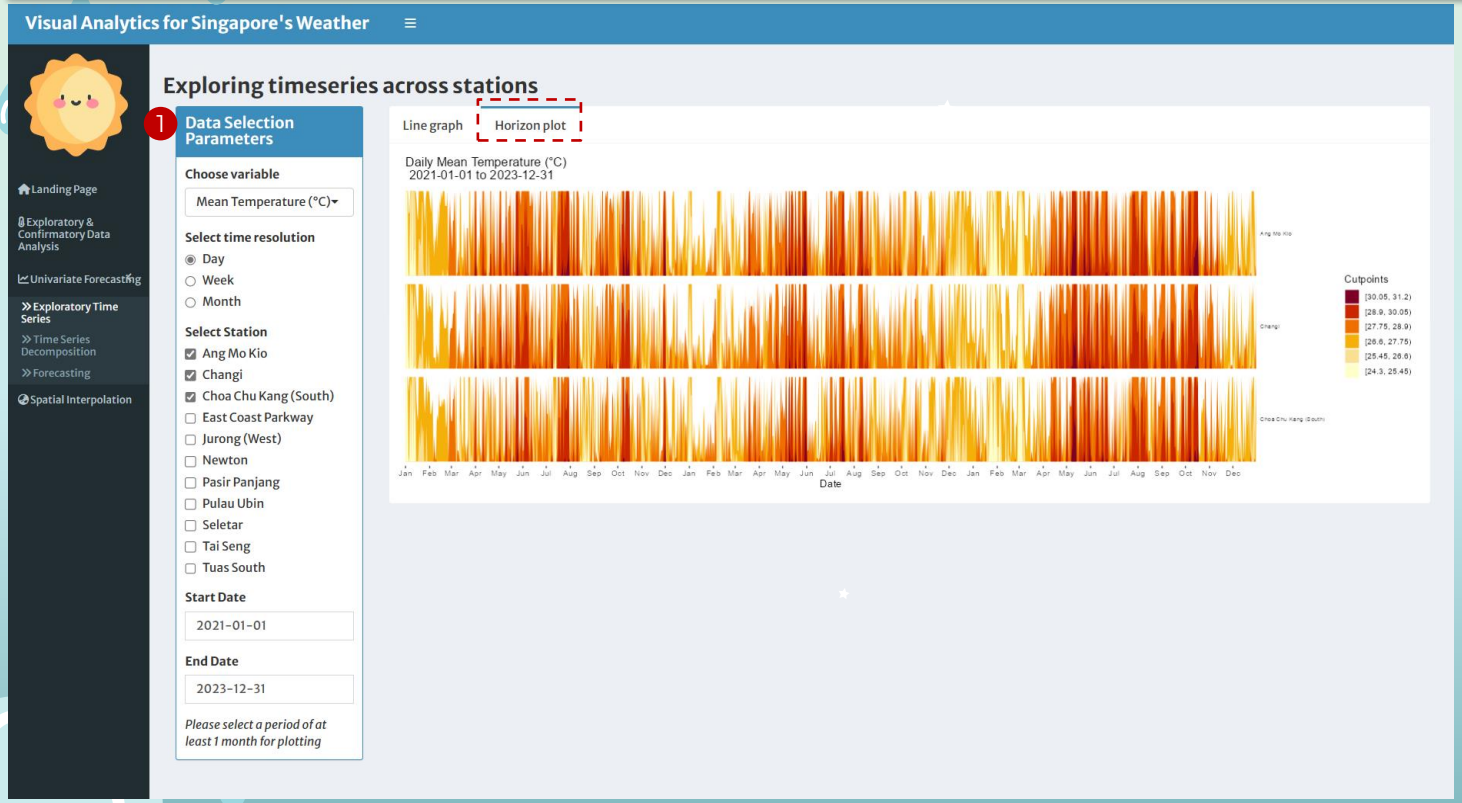
3. Sort the data table in ascending or descending order

4. Click on the legend to toggle the visibility of specific stations' lines on the graph

Tips for date selection: Note that the minimum period is one month. If users select an end date that is earlier than the start date, the start date will automatically adjust to one month before the selected end date.

Univariate Forecasting: Exploratory Time Series

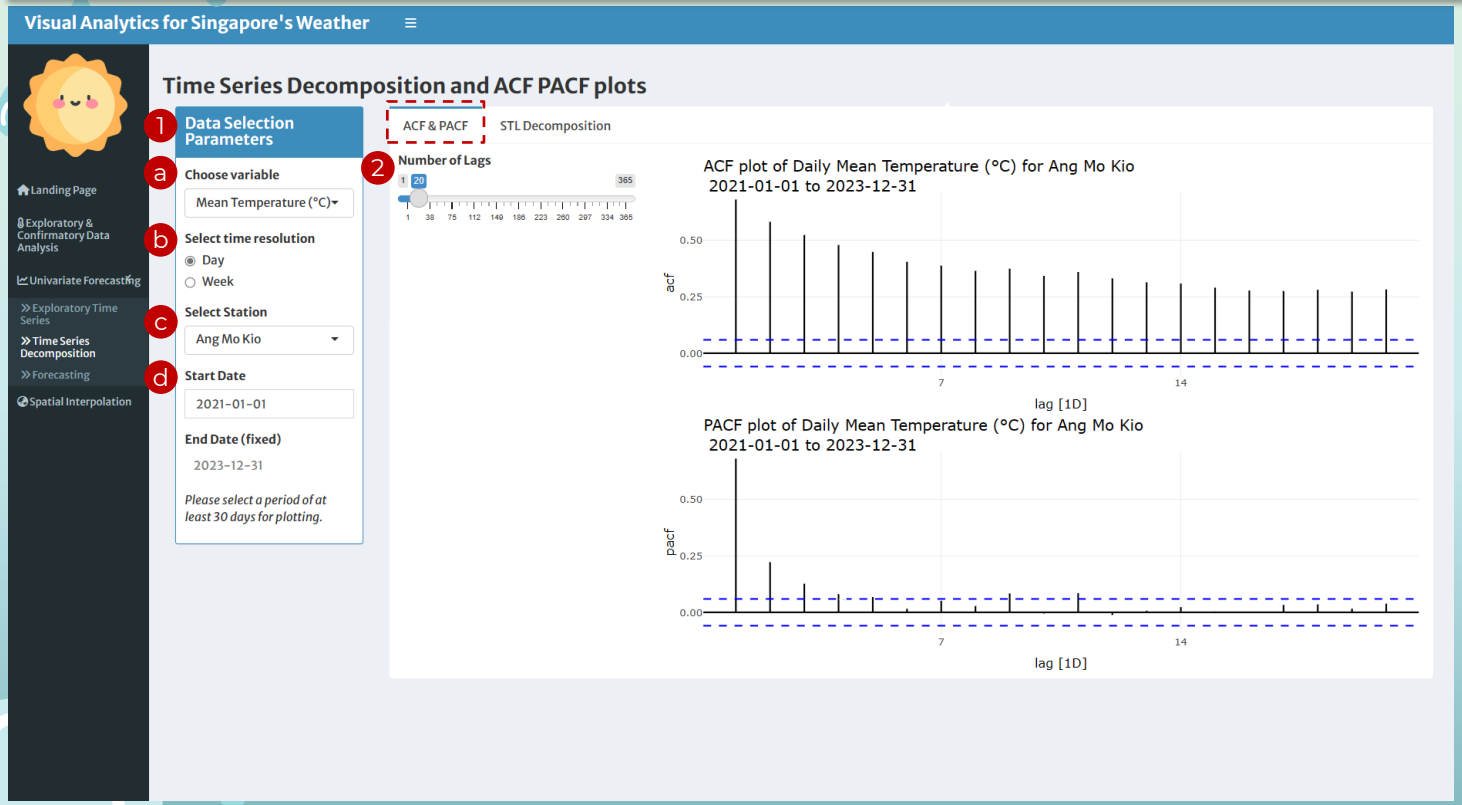
Horizon Plot: It shows rain and temperature changes over time using colors. Blue shades represent rainfall, with darker blues indicating more rain, while a range from yellow to darker red colors shows temperature changes, with darker reds indicating higher temperature. If multiple stations are selected, the plot switches to a facet view, presenting separate charts for each station for easy comparison.



1. Data Selection Parameters:
The input for this section automatically aligns with what user selected for the line graph tab

Univariate Forecasting: Time Series Decomposition

ACF & PCF: This page allows users to analyze the time-dependent characteristics of selected time series data through ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function). ACF measures the linear relationship between lagged values of the series, while PACF identifies the direct correlation between observations at different lags, with the influence of intermediary observations removed.



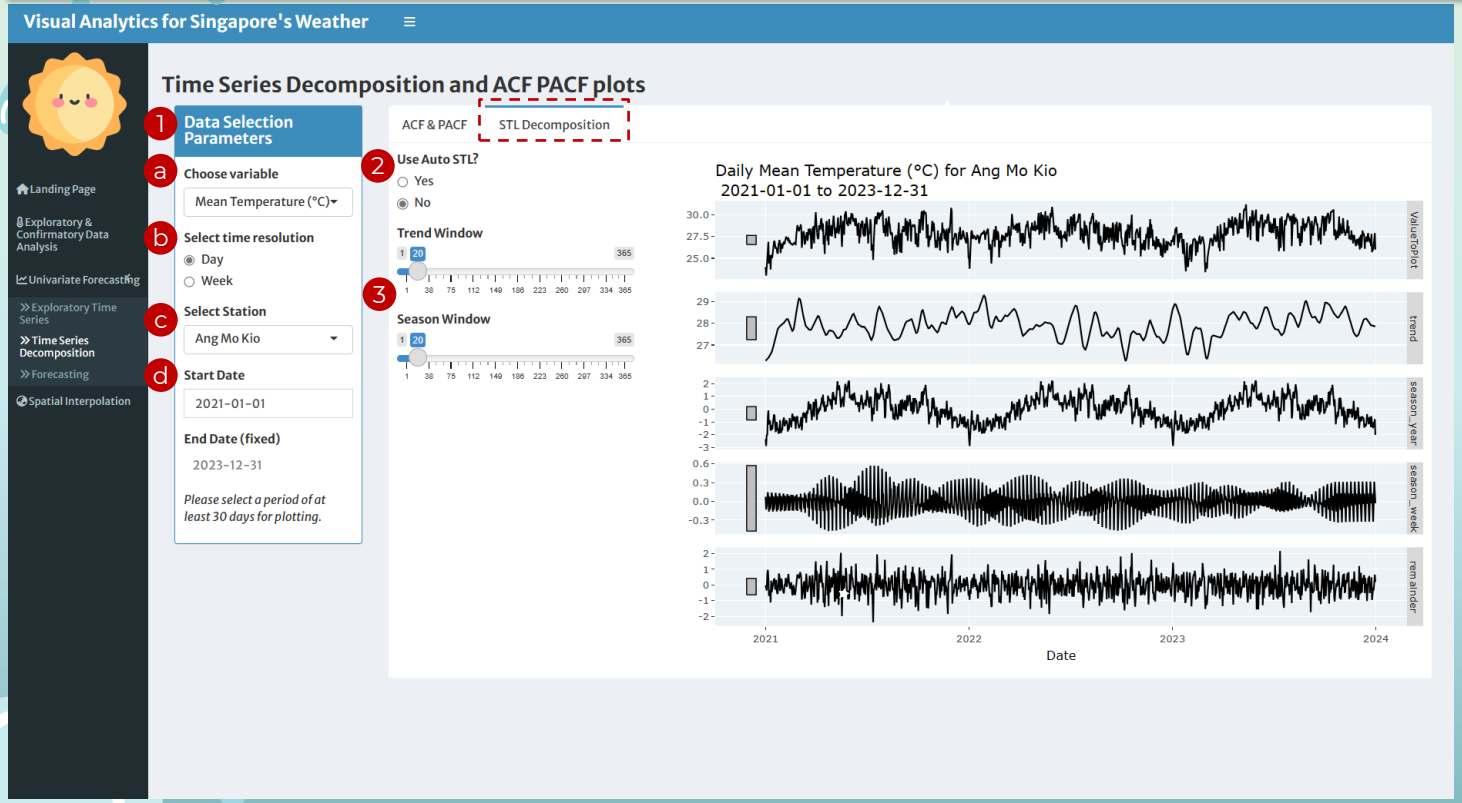
1. Data Selection Parameters:

a) Use the dropdown box to select a desired variable
b) Use the radio button to select the time resolution
c) Use the dropdown box to select a specific station
d) Use the date picker to specify the start date

2. Use the slider to specify the number of lags

Univariate Forecasting: Time Series Decomposition

STL Decomposition: This page enables users to decompose a time series from a specific station into three key components using STL: trend, seasonality, and residuals. The 'remainder' component, displayed in the last panel, represents what remains after the seasonal and trend-cycle components have been removed from the data.



1. Data Selection Parameters:

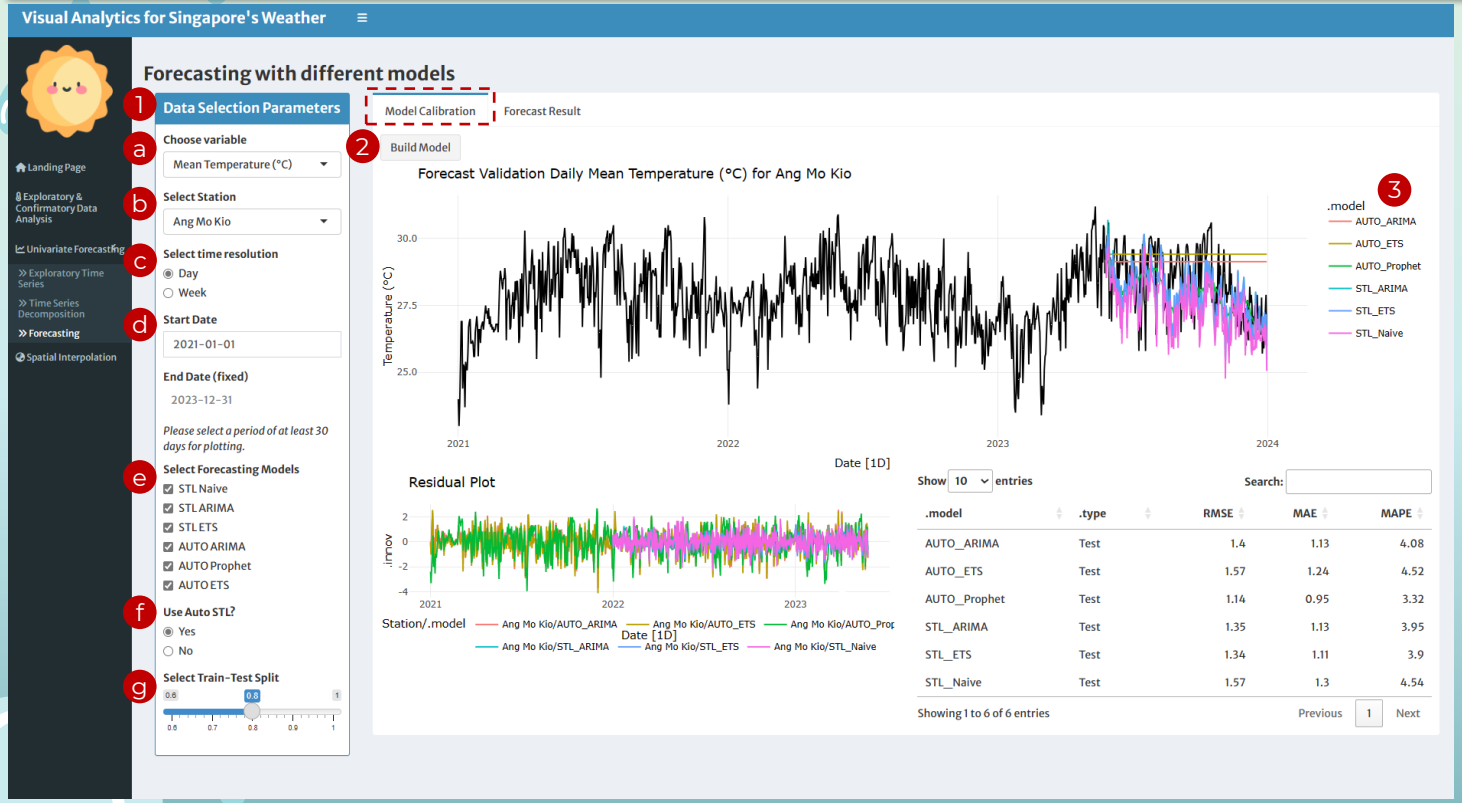
a) Use the dropdown box to select a desired variable
b) Use the radio button to select the time resolution
c) Use the dropdown box to select a specific station
d) Use the date picker to specify the start date

2. Choose 'Auto STL' by selecting 'Yes' for automatic settings or 'No' for manual configuration.

3. If 'No' is selected for Auto STL, specify the 'trend' and 'season' windows to customize your decomposition.

Univariate Forecasting: Forecasting

Model Calibration: This page allows users to select a station and compare multiple forecast models, showcasing a forecast validation plot, error plot, and key metrics (RMSE, MAE, MAPE) in an error table for comprehensive model assessment.



1. Data Selection Parameters:

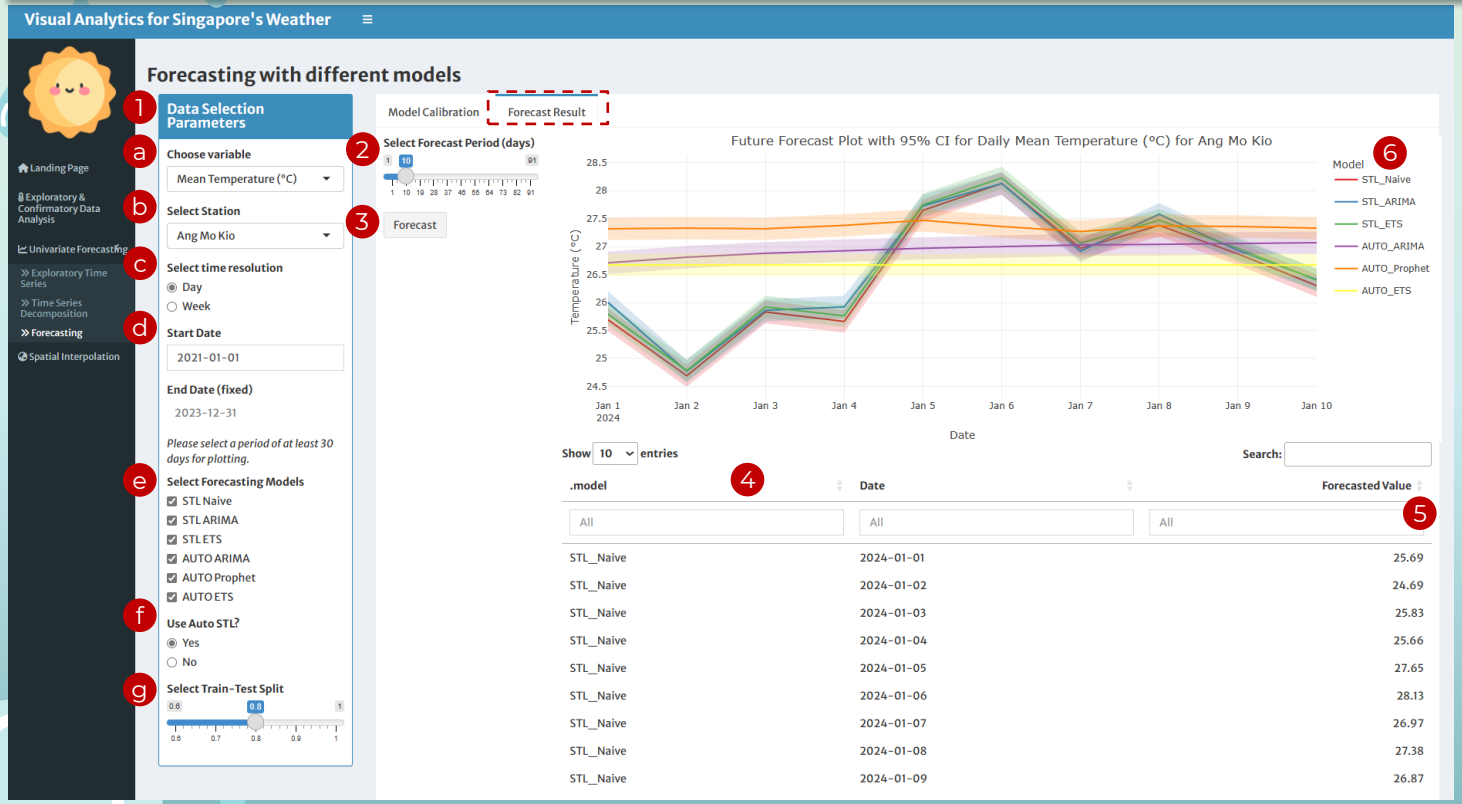
a) Use the dropdown box to select a desired variable
b) Use the radio button to select the time resolution
c) Use the dropdown box to select a specific station
d) Use the date picker to specify the start date
e) Use the checkbox to select multiple models for comparison
f) Choose 'Auto STL' by selecting 'Yes' for automatic settings or 'No' for manual configuration. If 'No' is selected for Auto STL, specify the 'trend' and 'season' windows to customize your decomposition.
g) Use the slider to specify the train-test split

2. Click the "Build Model" button to update the plot and table with the results

3. Click on the legend to toggle the visibility of specific models' lines on the graph

Univariate Forecasting: Forecasting

Forecast Result: This page displays future forecasts for a selected station by plotting multiple forecast models on the same line graph and providing a data table of forecasted values. This setup allows users to visually compare the forecasts from different models, offering a straightforward way to assess their relative outcomes.



1. Data Selection Parameters:

a) Use the dropdown box to select a desired variable
b) Use the radio button to select the time resolution
c) Use the dropdown box to select a specific station
d) Use the date picker to specify the start date
e) Use the checkbox to select multiple models for comparison
f) Choose 'Auto STL' by selecting 'Yes' for automatic settings or 'No' for manual configuration. If 'No' is selected for Auto STL, specify the 'trend' and 'season' windows to customize your decomposition.
g) Use the slider to specify the train-test split

2. Use the slider to specify the forecast period.

3. Click the "Forecast" button to update the plot and table with the results.

4. For the models selected in (e), apply a filter to view data for a specific model.

5. Sort the data table in ascending or descending order

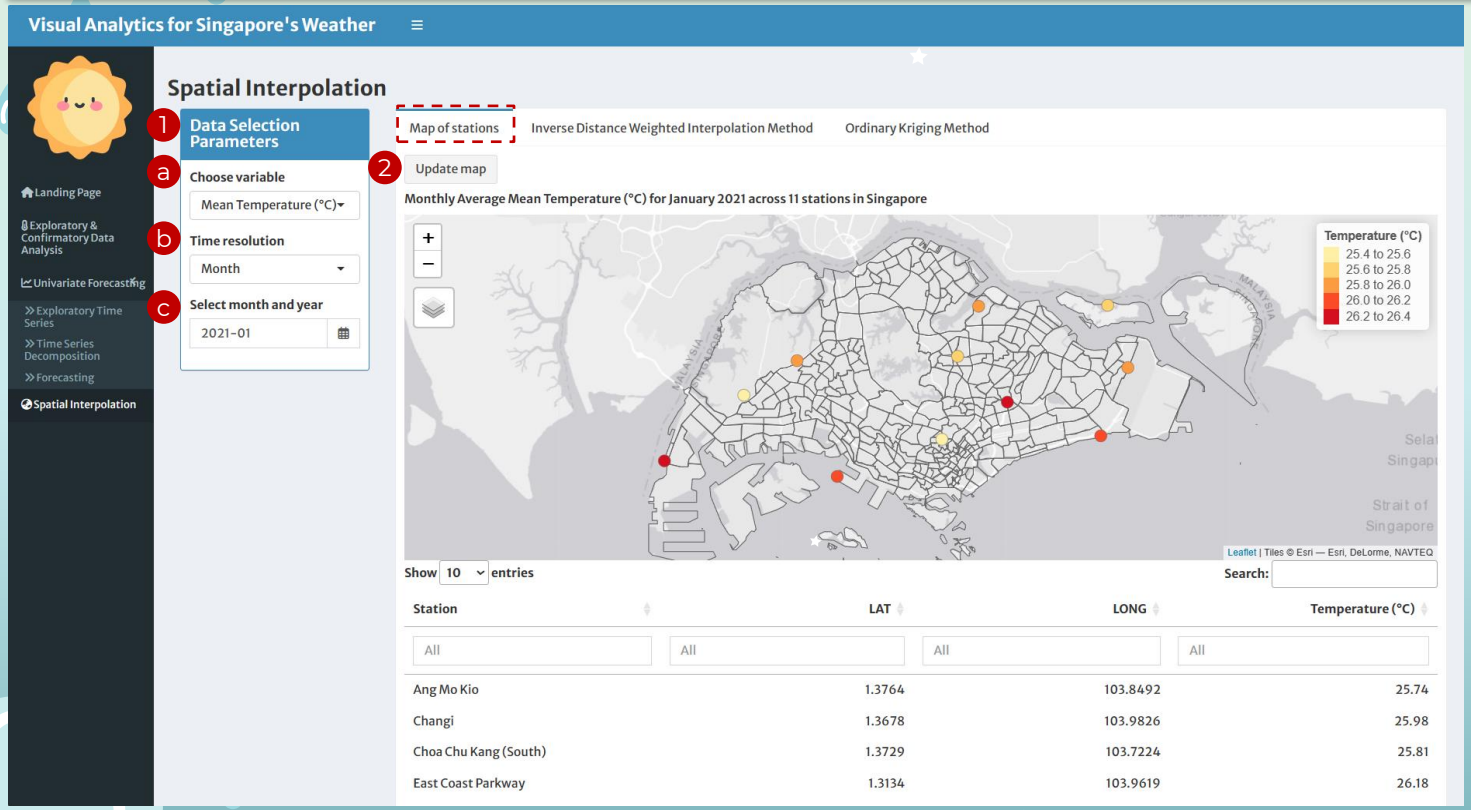
6. Click on the legend to toggle the visibility of specific models' lines on the graph

04

Spatial Interpolation

Spatial Interpolation

Map of stations: This page allows users to visually explore the geographical distribution of weather stations across Singapore, displaying precise locations where temperature and rainfall data are collected. The data can be displayed for specific days or aggregated over months or years.



1. Data Selection Parameters:

a) Use the dropdown box to select a desired variable

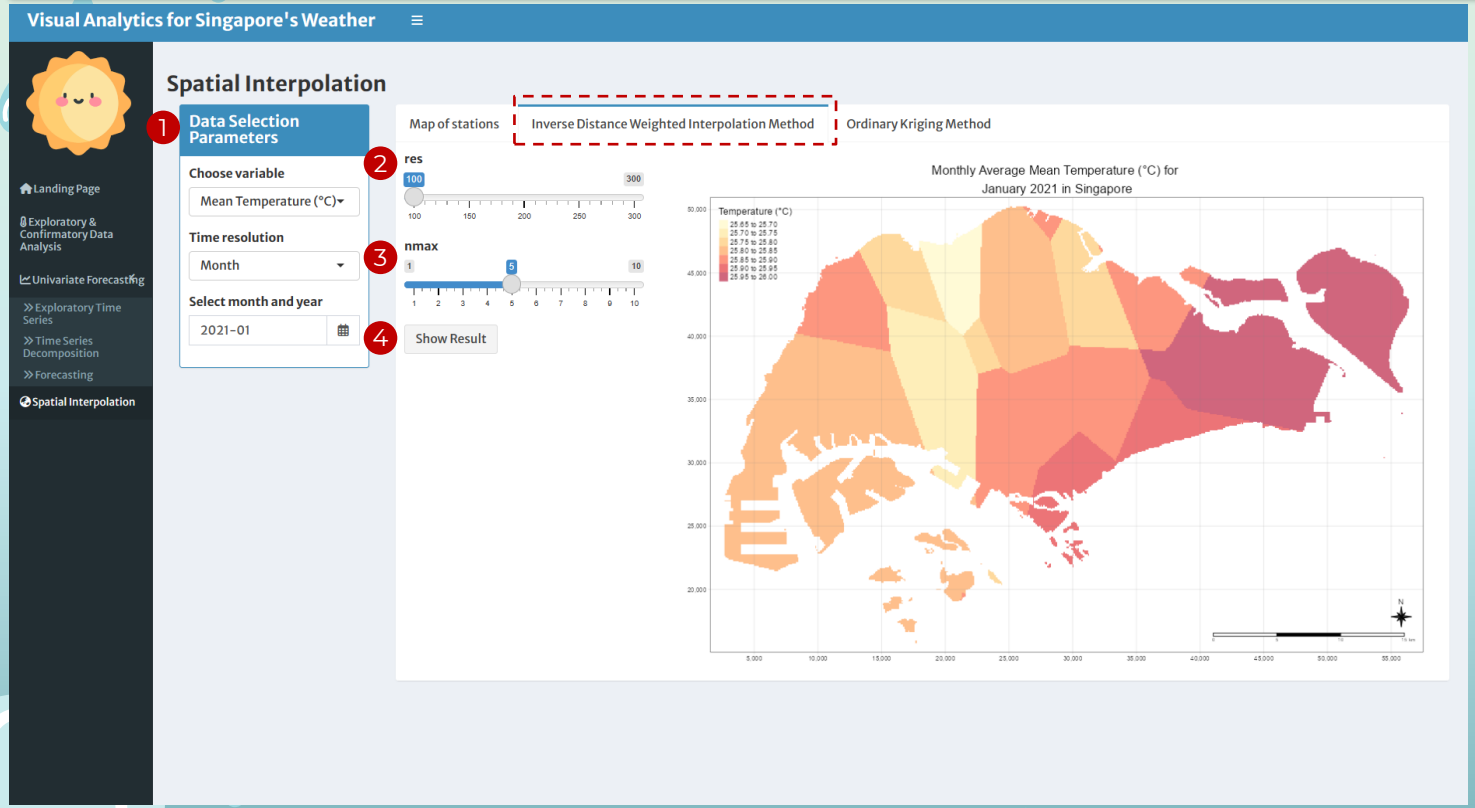
b) Use the dropdown box to select the time resolution to display the data.

c) Use the date picker to specify the period of data to display

2. Click 'Update map' button to display the data on the interactive map

Spatial Interpolation

Inverse Distance Weighted Interpolation Method: This page allows users to apply the inverse distance weighted interpolation method to generate an isohyet/isotherm map. This method calculates values based on the proximity and influence of nearby monitored points.



1. Data Selection Parameters: The input for this section automatically aligns with the user's selections from the "Map of Stations" tab.

2. Use the slider to specify *res* for the generated plot

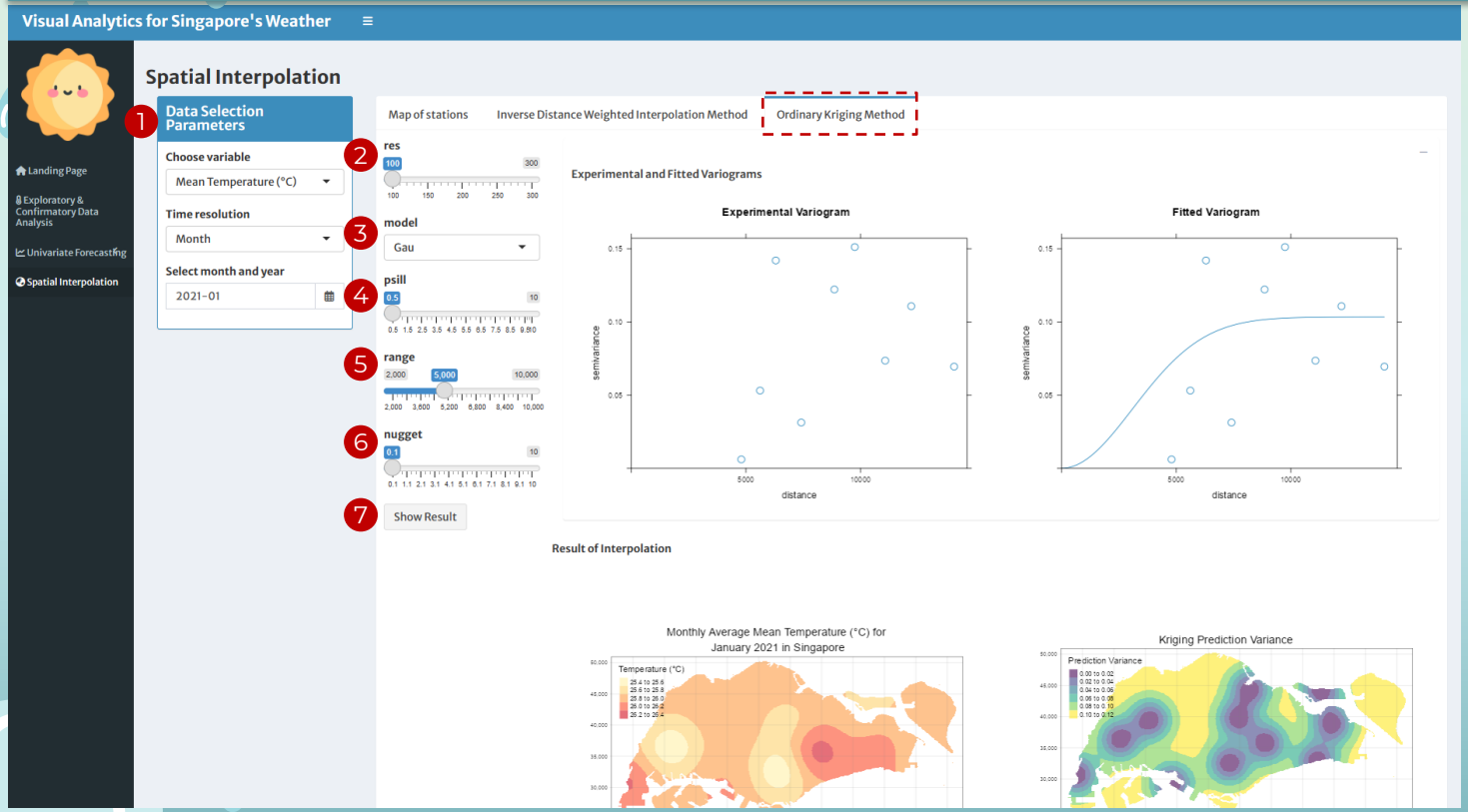
3. Use the slider to specify *nmax*, which is the number of nearest point observations to be used for the model

4. Click 'Show Result' button to display the isohyet / isotherm map

Tips: A lower *res* value will make the plot more detailed and clearer because the raster layer would be created with a finer grid for higher resolutions.

Spatial Interpolation

Ordinary Kriging Method: This page allows users to apply ordinary kriging, to generate an isohyet/isotherm map. This method can generate more accurate estimates of weather conditions at unmonitored locations, because it considers distance and spatial relationship between points.



1. Data Selection Parameters:
The input for this section automatically aligns with the user's selections from the "Map of Stations" tab.
2. Use the slider to specify *res*
3. Use dropdown box to select variogram *model* for kriging
4. Use the slider to specify *psill* (partial sill).
5. Use the slider to specify *range*, which is the range parameter of the variogram.
6. Use the slider to specify *nugget*, which represents the variance at zero distance.
7. Click 'Show Result' button to display the plots of experimental variogram, fitted variogram, isohyet/isotherm map and prediction variance.

Tip: The *range* parameter is likely to have the most impact on the resultant map. It affects how far the influence of a known data point reaches in the interpolation process, which can alter the smoothness and extent of interpolated values across the plot