Gridded Data Validator GUI User's Manual Version 1.0

By Saurav Bhattarai, Rocky Talchabhadel and Nawaraj Pradhan

PURPOSE

The purpose of this user's guide is to provide background methods and implementation guidance on the Gridded Data Validator GUI, a comprehensive desktop application designed for researchers and climate analysts to fetch, analyze, and visualize climate data from multiple sources including ground stations (Meteostat), ERA5 reanalysis, DAYMET, and PRISM datasets across the United States.

INTRODUCTION

Research and development of effective methods for analyzing climate data from multiple sources is critical for understanding climate patterns, validating model results, and making informed decisions (Bhattarai and Talchabhadel, 2024). The Gridded Data Validator GUI simplifies this process by providing a unified interface for accessing, comparing, and visualizing precipitation data from various sources. The tool supports temporal analysis (daily, monthly, yearly, seasonal) and extreme value analysis (both low and high extremes), and generates publication-quality visualizations to help identify discrepancies and patterns across different climate datasets.

Climate data analysis often requires accessing and comparing data from multiple sources, each with its own format, resolution, and data access methods. This process can be time-consuming and error-prone when done manually or through custom scripts. The Gridded Data Validator GUI addresses these challenges by providing a consistent, user-friendly interface that streamlines the entire workflow from data acquisition to visualization.

This application was developed at the U.S. Army Engineer Research and Development Center (ERDC) by Saurav Bhattarai as an ORISE fellow under the mentorship of Dr. Nawaraj Pradhan (ERDC) and Dr. Rocky Talchabhadel (Jackson State University). The current version focuses on precipitation data for the continental United States, but future versions will incorporate additional variables such as temperature and extend support to other global regions and datasets.

THEORETICAL BACKGROUND

The Gridded Data Validator GUI incorporates several statistical metrics to quantify the agreement between ground station observations and gridded datasets. These metrics are essential for understanding the reliability and biases of different data sources.

Statistical Metrics

The following statistical metrics are calculated for comprehensive comparison:

Coefficient of Determination (R^2)

R^2 measures how well the gridded data corresponds to the ground observations, with values ranging from 0 to 1 (perfect fit):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{1}$$

where O_i is the observed value, P_i is the predicted (gridded) value, and $\bar{O}$ is the mean of observed values.

Root Mean Square Error (RMSE)

RMSE quantifies the average magnitude of prediction errors:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{2}$$

Bias

Bias represents the average tendency of the gridded data to overestimate or underestimate relative to ground observations:

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i) \tag{3}$$

Mean Absolute Error (MAE)

MAE measures the average magnitude of errors without considering their direction:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{4}$$

Nash-Sutcliffe Efficiency (NSE)

NSE indicates how well the gridded data matches ground observations relative to using the mean of the observations:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{5}$$

Percent Bias (PBIAS)

PBIAS measures the average tendency of the gridded data in percentage:

$$PBIAS = 100 \times \frac{\sum_{i=1}^{n}(P_i - O_i)}{\sum_{i=1}^{n} O_i} \tag{6}$$

Temporal Analysis Approaches

The application employs multiple temporal approaches to analyze climate data:

1. **Daily Analysis**: Compares daily precipitation values to capture high-frequency variations

2. **Monthly Analysis**: Aggregates data to monthly totals to reduce noise and identify seasonal patterns

3. **Yearly Analysis**: Provides long-term trends and annual variability assessment

4. **Seasonal Analysis**: Groups data by meteorological seasons (Winter, Spring, Summer, Fall) to identify seasonal biases

Extreme Value Analysis

The application performs specialized analysis for extreme precipitation values:

1. **Low Extremes (10th percentile)**: Evaluates how well gridded datasets capture low precipitation events, which are important for drought monitoring

2. **High Extremes (90th percentile)**: Assesses performance during intense precipitation events, critical for flood forecasting and infrastructure design

Data Aggregation Methods

When aggregating daily data to coarser temporal resolutions, the application employs the following methods:

Monthly Aggregation

Monthly precipitation is calculated as the sum of daily values, requiring at least 80% data availability for a valid month:

$$P_{monthly} = \sum_{i=1}^{n} P_{daily,i} \quad \text{where} \quad n_{valid} \geq 0.8 \times n_{days} \tag{7}$$

Yearly Aggregation

Yearly totals require at least 9 valid months:

$$P_{yearly} = \sum_{i=1}^{12} P_{monthly,i} \quad \text{where} \quad n_{valid\_months} \geq 9 \tag{8}$$

MODEL DEVELOPMENT

The Gridded Data Validator GUI is developed in Python using the PyQt5 framework for the graphical user interface. The application follows a modular architecture with clear separation of concerns for maintainability and extensibility.

Architecture

The application implements a layered architecture with four main components:

1. **UI Layer**: Contains all user interface components built with PyQt5
2. **Controller Layer**: Manages application flow and coordinates between UI and business logic
3. **Business Logic Layer**: Implements core functionality for data fetching, analysis, and visualization
4. **Data Layer**: Handles data storage, loading, and configuration management

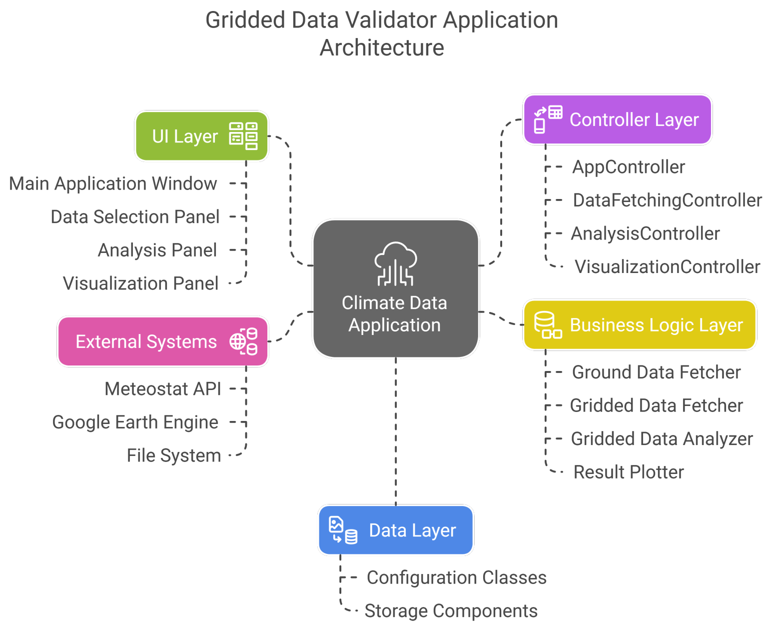

Figure 1: Architecture diagram of the Climate Data Fetcher GUI application showing the relationships between UI, Controller, Business Logic, and Data layers.

Key Components

The following key components comprise the application:

UI Layer

- **ClimateDataApp**: Main application window containing the tab widget and menu bar
- **DataSelectionPanel**: Panel for selecting and downloading climate data
- **AnalysisPanel**: Panel for analyzing downloaded data
- **VisualizationPanel**: Panel for generating and viewing visualizations

Controller Layer

- **AppController**: Main controller that coordinates all aspects of the application
- **DataFetchingController**: Manages data fetching operations
- **AnalysisController**: Manages analysis operations
- **VisualizationController**: Manages visualization generation

Business Logic Layer

- **GroundDataFetcher**: Fetches data from ground stations using the Meteostat API
- **GriddedDataFetcher**: Fetches data from gridded datasets using Google Earth Engine
- **GriddedDataAnalyzer**: Analyzes and compares datasets using statistical methods
- **ResultPlotter**: Generates visualizations of analysis results

Data Layer

- **DataConfig**: Base configuration for data fetching
- **GroundDataConfig**: Configuration for ground data fetching
- **GriddedDataConfig**: Configuration for gridded data fetching

Data Flow

The application implements the following data flow:

1. User inputs parameters in the Data Selection panel
2. Application fetches data from selected sources
3. Downloaded data is stored in the Data directory
4. User configures and runs analysis in the Analysis panel
5. Analysis results are stored in the Results directory
6. User generates visualizations in the Visualization panel
7. Visualizations are stored in the Plots directory

Threading Model

To ensure a responsive user interface during long-running operations, the application implements a threading model:

1. Main thread handles UI rendering and user interactions
2. Worker threads handle data fetching, analysis, and visualization generation
3. Communication between threads is managed through PyQt's signals and slots mechanism

DATA SOURCES AND CONFIGURATION

The Gridded Data Validator GUI supports multiple data sources for comprehensive comparison and analysis.

Supported Data Sources

Table 1 provides a summary of the data sources currently supported by the application.

Table 1: Supported Data Sources in Gridded Data Validator GUI

Source	Type	Resolution	Coverage	Period
Meteostat	Ground stations	Point-based	Global	1980-2024
ERA5	Reanalysis	0.25° × 0.25°	Global	1980-2024
DAYMET	Gridded	1 km	North America	1980-2024
PRISM	Gridded	4 km	CONUS	1980-2024

Ground Station Data

Ground station data is accessed through the Meteostat Python library, which provides a consistent interface to weather station data from various sources worldwide. The application supports filtering by state and specified time periods.

ERA5 Reanalysis Data

ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate. The application accesses ERA5 data through Google Earth Engine, which hosts the ”ECMWF/ERA5_LAND/DAILY_AGGR” collection. The daily precipitation is converted from meters to millimeters for consistency.

DAYMET Data

DAYMET provides gridded estimates of daily weather parameters for North America. The application accesses DAYMET Version 4 data through Google Earth Engine's ”NASA/ORNL/DAYMET_V4” collection.

PRISM Data

PRISM (Parameter-elevation Regressions on Independent Slopes Model) provides gridded climate data for the continental United States. The application accesses PRISM data through Google Earth Engine's ”OREGONSTATE/PRISM/AN81d” collection.

Configuration

The application uses a hierarchical configuration system to manage settings for data fetching, analysis, and visualization.

Data Fetching Configuration

Ground data and gridded data fetching is configured through dedicated configuration classes:

```python
@dataclass
class DataConfig:
    """Base configuration for data fetching"""
    start_year: int = 1980
    end_year: int = 2024
    data_dir: str = "Data"

@dataclass
class GroundDataConfig(DataConfig):
    """Configuration for ground data fetching"""
    states: Optional[List[str]] = None  # None means entire US
    metadata_filename: str = "stations_metadata.csv"
    data_filename: str = "ground_daily_precipitation.csv"

@dataclass
class GriddedDatasetConfig:
    """Configuration for a single gridded dataset"""
    name: str
    collection_name: str
    variable_name: str
    conversion_factor: float = 1.0
    enabled: bool = False
```

Listing 1: Example Data Configuration Classes

Earth Engine Configuration

Google Earth Engine access is managed through a dedicated Earth Engine panel that allows users to authenticate and configure the Earth Engine project ID.

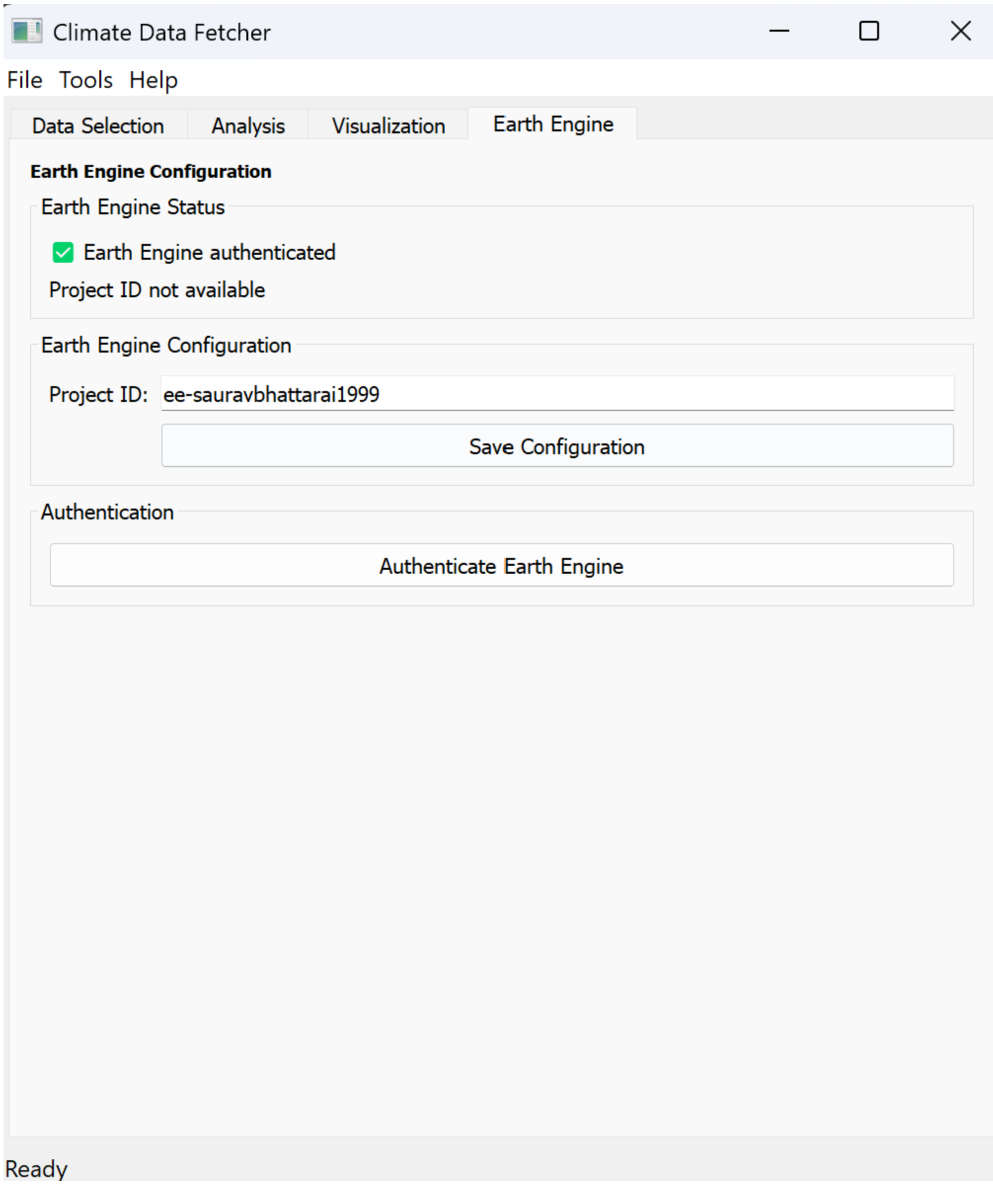

Figure 2: Earth Engine Configuration panel showing authentication status and project configuration options.

INSTALLATION AND SETUP

The Gridded Data Validator GUI can be installed and run on Windows, macOS, and Linux platforms.

Prerequisites

Before installing the application, ensure you have the following prerequisites:

- Python 3.10 or higher

- pip (Python package installer)
- Git (optional, for cloning the repository)
- Google Earth Engine account (for accessing gridded datasets)

Installation Steps

Follow these steps to install the Gridded Data Validator GUI:

1. Clone or download the repository:

```
git clone https://github.com/Saurav-JSU/GeeData-GroundData-validator.git
cd GeeData-GroundData-validator
```

2. Create and activate a virtual environment:

```
# On Windows
python -m venv venv
venv\Scripts\activate

# On macOS/Linux
python -m venv venv
source venv/bin/activate
```

3. Install dependencies:

```
pip install -r requirements.txt
```

4. Set up Earth Engine authentication (if needed):

```
earthengine authenticate
```

Running the Application

Start the GUI application using one of the following methods:

1. Run directly with Python:

```
python main.py
```

2. On Windows, use the provided batch file:

```
launch_app.bat
```

Directory Structure

The application creates the following directories for data storage:

- **Data/**: Raw data files downloaded from sources
- **Results/**: Statistical analysis results
- **Plots/**: Generated visualizations

USER INTERFACE AND WORKFLOW

The Gridded Data Validator GUI is organized into three main tabs that guide users through the workflow from data selection to visualization.

Main Application Window

The main application window contains a tab widget for navigating between different functionality and a menu bar for application-level operations.

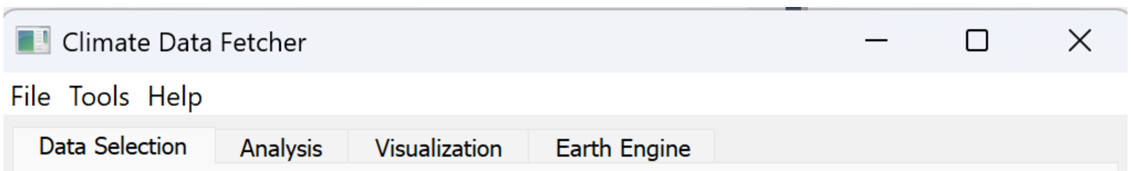

Figure 3: Main application window of the Gridded Data Validator GUI showing the tab widget and menu bar.

Data Selection Panel

The Data Selection panel allows users to select data sources, specify time periods, and select geographical areas for data fetching.

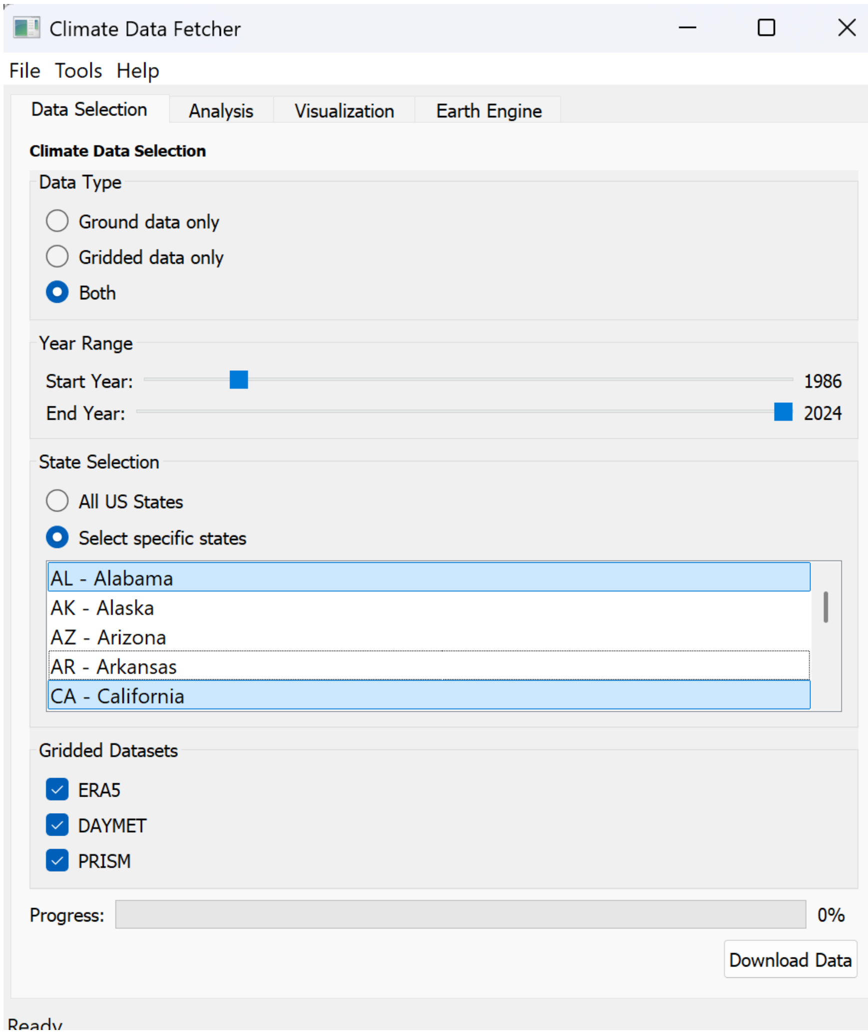

Figure 4: Data Selection panel showing options for selecting data sources, time periods, and geographical areas.

Data Type Selection

Users can select one of the following data types:

- Ground data only
- Gridded data only
- Both ground and gridded data

Year Range Selection

Users can specify the start and end years for data fetching using sliders. The available range is from 1980 to 2024.

State Selection

Users can select either all US states or specific states from a list. When selecting specific states, users can choose multiple states by holding Ctrl (or Cmd on macOS) while clicking.

Gridded Dataset Selection

Users can select one or more gridded datasets to include in the analysis:

- ERA5
- DAYMET
- PRISM

Downloading Data

After configuring the options, users can click the ”Download Data” button to start the data fetching process. A progress bar indicates the download progress, and status messages provide updates on the process.

Analysis Panel

The Analysis panel allows users to configure and run statistical analysis on the downloaded data.

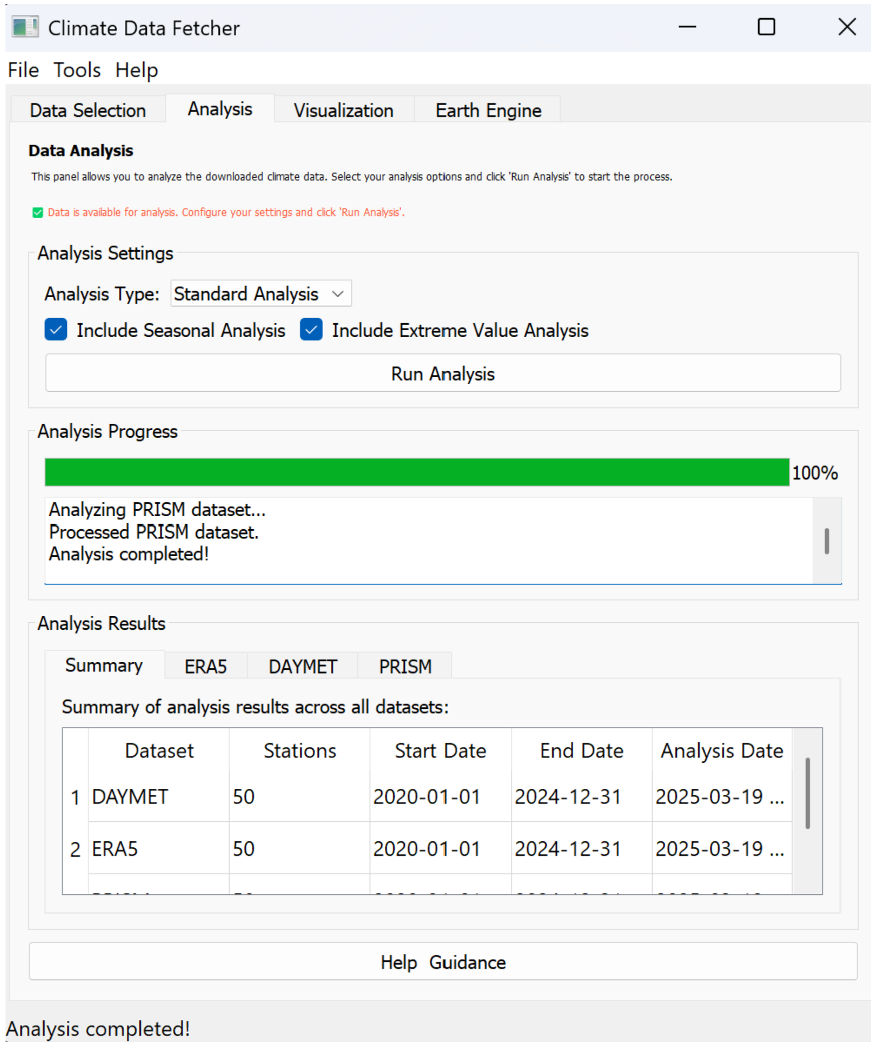

Figure 5: Analysis panel showing configuration options and results tabs.

Analysis Settings

Users can configure the following analysis settings:

- Analysis Type: Standard or Custom analysis
- Include Seasonal Analysis: Toggle for including seasonal analysis
- Include Extreme Value Analysis: Toggle for including extreme value analysis

Running Analysis

After configuring the settings, users can click the ”Run Analysis” button to start the analysis process. A progress bar indicates the analysis progress, and a status text area provides detailed status updates.

Viewing Results

Analysis results are displayed in tabs organized by dataset (ERA5, DAYMET, PRISM). Each tab shows statistical metrics for different temporal scales (daily, monthly, yearly) and seasonal comparisons.

Visualization Panel

The Visualization panel allows users to generate and view visualizations of the analysis results.

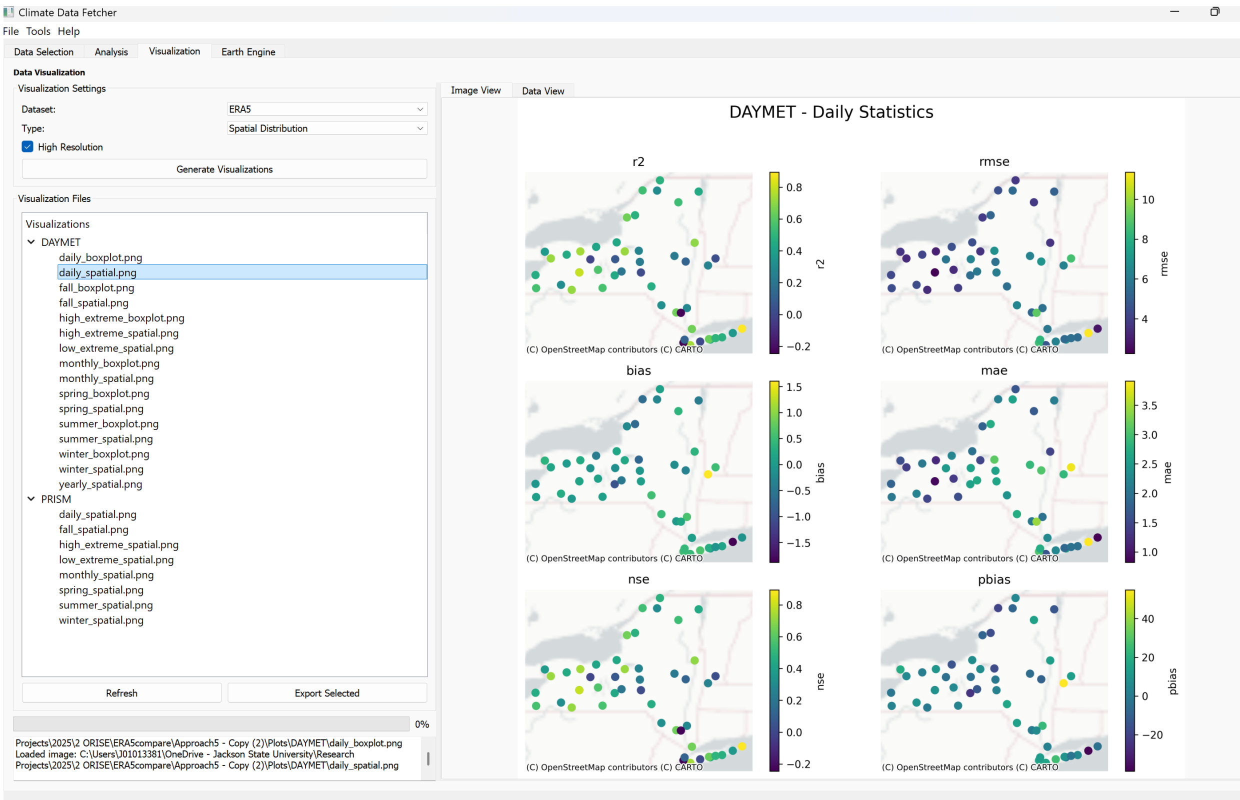

Figure 6: Visualization panel showing configuration options and visualization preview.

Visualization Settings

Users can configure the following visualization settings:

- Dataset: Select the dataset to visualize (ERA5, DAYMET, PRISM)
- Type: Select the visualization type (Spatial Distribution, Box Plots, Seasonal Comparison, etc.)
- High Resolution: Toggle for generating high-resolution images

Generating Visualizations

After configuring the settings, users can click the ”Generate Visualizations” button to start the visualization generation process. A progress bar indicates the generation progress, and a status text area provides detailed status updates.

Viewing and Exporting Visualizations

Generated visualizations are displayed in the image viewer and listed in the file tree. Users can select a visualization from the file tree to view it in the image viewer and export it to a location of their choice.

FUTURE DEVELOPMENT

The Gridded Data Validator GUI is under active development with planned enhancements to expand its capabilities and utility.

Additional Climate Variables

Future versions will incorporate additional climate variables beyond precipitation, including:

- Temperature (minimum, maximum, and mean)
- Relative humidity
- Wind speed and direction
- Solar radiation
- Evapotranspiration

Additional Data Sources

Support for additional gridded datasets will be added in future versions:

- GPM IMERG
- CHIRPS
- MERRA-2
- NLDAS-2

Enhanced Statistical Analysis

Future versions will include additional statistical analysis capabilities:

- Climate index calculation and comparison
- Trend analysis and significance testing
- Frequency analysis
- Uncertainty quantification

Improved Visualization

Visualization capabilities will be expanded to include:

- Interactive maps and charts
- Time series animations
- Comparison sliders
- Custom report generation

Performance Enhancements

Planned performance improvements include:

- Distributed computing support for large datasets
- Caching mechanisms for faster data access
- Incremental data fetching

CONCLUSION

The Gridded Data Validator GUI provides a comprehensive solution for accessing, analyzing, and visualizing climate data from multiple sources. The application streamlines the workflow from data acquisition to visualization, enabling researchers and climate analysts to efficiently compare and evaluate different datasets.

The current version supports precipitation data from ground stations, ERA5 reanalysis, DAYMET, and PRISM datasets across the United States, with planned extensions to include additional variables, datasets, and enhanced analysis capabilities in future versions. The application's modular architecture and user-friendly interface make it a valuable tool for climate research, model validation, and decision support.

Publications, models, and data products that make use of this software should include proper citation and acknowledgement. Any additional documents on user guidelines of future Gridded Data Validator GUI enhancements will be an addendum to this original user's manual.

DISCLAIMER

The software is provided without warranty of any kind in its use.

ADDITIONAL INFORMATION

This Coastal and Hydraulics Engineering technical note (CHETN) was prepared by Saurav Bhattarai, saurav.bhattarai@students.jsums.edu, PhD student at Jackson State University, as part of an ORISE fellowship at the US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory. The work was conducted under the mentorship of Dr. Nawaraj Pradhan (ERDC) and Dr. Rocky Talchabhadel (Jackson State University).

This CHETN should be cited as follows:

Bhattarai, S. 2025. Gridded Data Validator GUI User's Manual Version 1.0. ERDC/CHL CHETN-XX-X. Vicksburg, MS: US Army Engineer Research and Development Center.

REFERENCES

Bhattarai, S., and R. Talchabhadel. 2024. "Comparative Analysis of Satellite-Based Precipitation Data across the CONUS and Hawaii: Identifying Optimal Satellite Performance." *Remote Sensing*, 16(16): 3058. https://doi.org/10.3390/rs16163058.

Beck, H. E., N. Vergopolan, M. Pan, V. Levizzani, A. I. J. M. van Dijk, G. P. Weedon, L. Brocca, F. Pappenberger, G. J. Huffman, and E. F. Wood. 2017. "Global-Scale Evaluation of 22 Precipitation Datasets Using Gauge Observations and Hydrological Modeling." *Hydrology and Earth System Sciences*, 21(12): 6201–6217. https://doi.org/10.5194/hess-21-6201-2017.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris. 2008. "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation across the Conterminous United States." *International Journal of Climatology*, 28(15): 2031–2064. https://doi.org/10.1002/joc.1688.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, et al. 2020. "The ERA5 Global Reanalysis." *Quarterly Journal of the Royal Meteorological Society*, 146(730): 1999–2049. https://doi.org/10.1002/qj.3803.

Thornton, P. E., M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook. 2018. "Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3." ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1328.