Gridded Data Validator Graphical User Interface: A PyQt5-based GUI for Comparative Analysis of Precipitation Products

Saurav Bhattarai[1,*], Rocky Talchabhadel[1], and Nawa Raj Pradhan[2]

[1]Department of Civil and Environmental Engineering, Jackson State University, Jackson, USA
[2]U.S. Army Corps of Engineers, Engineer Research and Development Center, Vicksburg, Mississippi, USA
[*]Corresponding author: saurav.bhattarai@students.jsums.edu

April 30, 2025

Abstract

This technical note presents the Gridded Data Validator Graphical User Interface (GUI), a comprehensive Python application for retrieving, analyzing, and visualizing precipitation data from multiple sources. The tool facilitates systematic comparison between ground station measurements and various gridded precipitation products (ERA5, DAYMET, PRISM, CHIRPS, FLDAS, GSMAP, and GLDAS). Developed using PyQt5, the application implements a Model-View-Controller architecture integrating Earth Engine API for efficient data acquisition. Users can select areas of interest through state boundaries, Hydrologic Unit Code (HUC) watersheds, or custom polygons, filter by date ranges, and generate statistical comparisons across temporal scales (daily, monthly, yearly, and seasonal). The application automates the calculation of performance metrics including coefficient of determination (R^2), root mean square error (RMSE), bias, mean absolute error (MAE), and Nash-Sutcliffe efficiency (NSE), producing publication-quality visualizations such as spatial distribution maps, time series comparisons, box plots, and radar charts. This tool streamlines precipitation analysis workflows, enabling researchers to efficiently evaluate the performance of gridded products in diverse geographic contexts.

Contents

1 Introduction

1.1 Background

Precipitation is a fundamental component of the hydrological cycle and serves as a critical input for numerous hydrological models, climate studies, and water resource management applications. While ground-based precipitation measurements provide valuable point observations, their spatial distribution is often limited, particularly in remote or complex terrain areas. To address this limitation, various gridded precipitation products have been developed, each with its own spatial resolution, temporal coverage, and derivation methodology.

However, the accuracy and reliability of these gridded products can vary significantly by region, season, elevation, and precipitation intensity. Comprehensive evaluation of these products against ground observations is essential for understanding their strengths and limitations before application in hydrological studies or engineering projects.

Currently, researchers must implement custom workflows to download, process, analyze, and visualize these datasets - a time-consuming and error-prone process that requires expertise in multiple programming languages and APIs. The Gridded Data Validator GUI addresses this challenge by providing a comprehensive, user-friendly platform for the systematic comparison of precipitation products.

1.2 Objective

The primary objective of the Gridded Data Validator GUI is to provide a unified, interactive platform that streamlines the process of retrieving, analyzing, and visualizing precipitation data from multiple sources. Specifically, the tool aims to:

- Facilitate efficient acquisition of precipitation data from ground stations and multiple gridded products
- Enable systematic statistical comparison between ground measurements and gridded products
- Generate comprehensive visualizations to evaluate product performance across different spatial and temporal scales
- Provide a user-friendly interface accessible to researchers without extensive programming expertise

1.3 Approach

The Gridded Data Validator GUI implements a Model-View-Controller (MVC) architecture using PyQt5 for the graphical interface. The tool integrates several key technologies:

- **Python** as the core programming language, providing access to rich data analysis and visualization libraries
- **PyQt5** for developing the graphical user interface with interactive components

- **Earth Engine API** for efficient access to multiple gridded precipitation products
- **Meteostat** for accessing ground station precipitation records
- **Pandas and NumPy** for data processing and statistical analysis
- **Matplotlib and Seaborn** for generating visualizations

The application streamlines the entire workflow from data selection and retrieval through analysis and visualization, enabling researchers to concentrate on interpreting results rather than implementing complex data handling procedures.

1.4 Scope

The Gridded Data Validator GUI currently focuses on:

- **Geographic scope:** Continental United States, with flexible area selection options
- **Temporal scope:** 1980-2024, depending on dataset availability
- **Precipitation products:** Ground stations (via Meteostat), ERA5, DAYMET, PRISM, CHIRPS, FLDAS, GSMAP, and GLDAS (Historical and Current)
- **Analysis metrics:** R^2, RMSE, bias, MAE, NSE, PBIAS at daily, monthly, yearly, and seasonal timescales
- **Visualization types:** Spatial distribution maps, time series comparisons, box plots, dataset comparisons, and radar charts

The application provides a foundation that can be extended to include additional datasets, analysis methods, and visualization techniques in future releases.

2 Methodology and Implementation

2.1 Software Architecture

The Gridded Data Validator GUI follows the Model-View-Controller (MVC) architectural pattern, which separates the application into three interconnected components. This design promotes modular development, code reusability, and maintainability. Figure 1 illustrates the high-level architecture of the application.

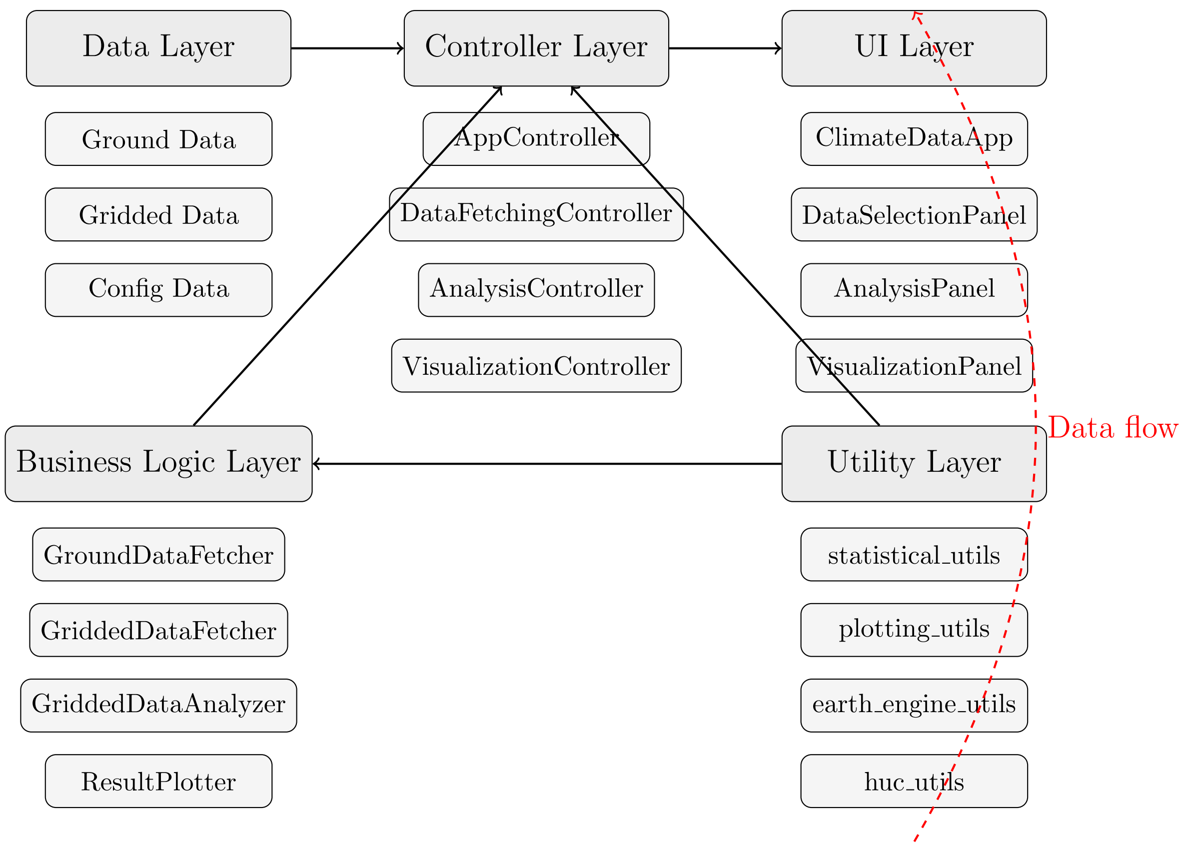

Figure 1: Model-View-Controller architecture of the Gridded Data Validator GUI

The key components of the architecture are:

1. **User Interface (UI) Layer (View):** Contains graphical user interface components built with PyQt5. This layer is responsible for displaying information to the user and interpreting user inputs.

2. **Controller Layer:** Manages the application flow and coordinates between the UI and business logic. Controllers handle user actions, trigger data operations, and update the UI.

3. **Business Logic Layer:** Implements core functionality for data fetching, analysis, and visualization. This layer contains the algorithms and processes that operate on the data.

4. **Data Layer:** Handles data storage, loading, and configuration management. This includes raw data files, analysis results, and application settings.

5. **Utility Layer:** Provides common functionality used across different parts of the application, such as statistical calculations, plotting functions, and integration with external services.

2.2 Data Sources and Acquisition

The application retrieves precipitation data from two primary sources:

2.2.1 Ground Station Data: Meteostat

Meteostat is an open-source Python library that provides access to historical weather observations from thousands of weather stations worldwide. The library interfaces with the Meteostat API, which aggregates data from various meteorological organizations including NOAA, Environment Canada, and the German Weather Service (DWD).

Key characteristics of Meteostat data include:

- **Temporal coverage:** Varies by station, with some records extending back to the early 1900s
- **Spatial coverage:** Global, with particularly dense coverage in North America and Europe
- **Variables:** Daily, hourly, and sub-hourly precipitation, temperature, humidity, wind, and other meteorological variables
- **Quality control:** Basic quality control measures applied, including outlier removal and consistency checks
- **Metadata:** Comprehensive station metadata including latitude, longitude, elevation, and operational periods

The application implements the following workflow for ground data acquisition:

1. User selects an area of interest using one of three methods:
   - US states selection (individual states or all US states)
   - HUC watershed selection (based on Hydrologic Unit Codes)
   - Custom polygon drawing on an interactive map
2. The application retrieves station metadata for stations within the selected area
3. For each station, daily precipitation data is downloaded for the specified time range
4. The data is compiled into a unified DataFrame with stations as columns and dates as the index
5. Quality control is performed to filter stations with insufficient data

2.2.2 Hydrologic Unit Codes (HUC) Integration

The application incorporates the United States Geological Survey (USGS) Watershed Boundary Dataset (WBD) through Google Earth Engine to enable watershed-based analysis. The HUC system is a hierarchical classification that divides the United States into hydrologic units based on surface hydrologic features:

- **HUC-2:** 21 major geographic regions (e.g., Mississippi River Basin)
- **HUC-4:** 222 subregions
- **HUC-6:** 370 basins
- **HUC-8:** 2,270 subbasins, typically covering 700-1,000 square miles
- **HUC-10:** 22,000 watersheds
- **HUC-12:** 160,000 subwatersheds

The application uses HUC-8 subbasins as the primary watershed selection unit because they provide an appropriate scale for regional precipitation analysis—large enough to contain multiple weather stations but small enough to represent relatively homogeneous hydrologic characteristics.

The HUC integration workflow includes:

1. Loading and caching HUC metadata from Earth Engine
2. Providing a user interface to browse and select HUCs by region
3. Retrieving the boundary geometry for the selected HUC
4. Filtering ground stations based on spatial intersection with the HUC boundary

2.2.3 Gridded Precipitation Products

The application integrates with Google Earth Engine (GEE) to access multiple gridded precipitation products. This integration enables efficient retrieval of spatially and temporally consistent precipitation estimates at various scales. Table 1 summarizes the gridded products available in the application.

Table 1: Gridded precipitation products available in the Gridded Data Validator GUI

Product	**Temporal Resolution**	**Period**	**Spatial Resolution & Coverage**
ERA5	Hourly→Daily	1950-present	11 km, Global
DAYMET	Daily	1980-present	1 km, North America
PRISM	Daily	1981-present	4 km, Continental US
CHIRPS	Daily	1981-present	5.5 km, 50°S-50°N
FLDAS	Monthly	1982-present	11 km, Africa, Middle East
GSMAP	Hourly→Daily	1998-present	11 km, 60°S-60°N
GLDAS-Historical	3-hourly→Daily	1948-2014	28 km, Global
GLDAS-Current	3-hourly→Daily	2000-present	28 km, Global

Each gridded product has unique characteristics:

- **ERA5:** Produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5 is a global reanalysis product combining model data with observations.

- **DAYMET:** Developed by Oak Ridge National Laboratory, DAYMET provides high-resolution, gridded estimates of daily weather parameters over North America, interpolated from ground station observations.

- **PRISM:** The Parameter-elevation Regressions on Independent Slopes Model (PRISM) produces high-resolution precipitation estimates for the continental US using station data, elevation, and other spatial factors.

- **CHIRPS:** The Climate Hazards Group InfraRed Precipitation with Station data combines satellite imagery with ground station data to create a quasi-global rainfall dataset.

- **FLDAS:** The Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System integrates satellite and ground-based observations to support drought monitoring.

- **GSMAP:** The Global Satellite Mapping of Precipitation provides high-resolution precipitation estimates based on satellite microwave and infrared observations.

- **GLDAS:** The Global Land Data Assimilation System(GLDAS) integrates satellite and ground-based observations to support land surface modeling and water resources applications.

The gridded data acquisition process follows these steps:

1. User selects desired gridded products and a time range
2. The application initializes Earth Engine with the user's project ID
3. For each product, the application:
   - Creates an Earth Engine ImageCollection with the appropriate filters
   - For sub-daily products, aggregates to daily values in Earth Engine
   - Extracts precipitation values at ground station locations
   - Applies product-specific conversion factors to ensure consistent units (mm/day)
   - Creates a DataFrame with the same structure as the ground data (stations as columns, dates as index)
4. The resulting DataFrames are saved to CSV files for subsequent analysis

2.2.4 Earth Engine Integration and Authentication

Access to the gridded precipitation products described above is facilitated through Google Earth Engine (GEE), which serves as the underlying infrastructure for the Gridded Data Validator GUI. This integration enables efficient retrieval of spatially-distributed precipitation estimates without requiring users to manually download large datasets. The application implements a dedicated authentication workflow through a specialized Earth Engine panel.

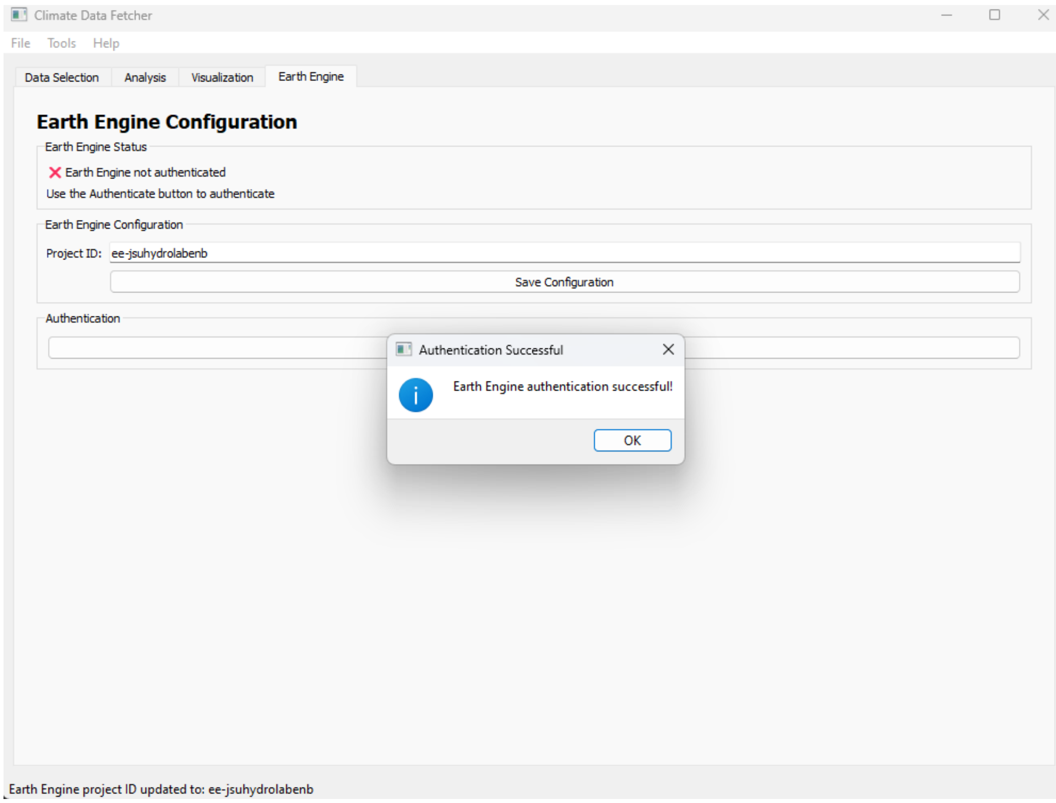

Figure 2: Earth Engine configuration panel showing authentication status, project ID configuration, and authentication controls. This interface streamlines access to gridded precipitation products through a clear, guided process.

The Earth Engine integration follows a systematic workflow:

1. **Status Verification:** Upon opening the panel, the application automatically checks Earth Engine authentication status and displays current connection information
2. **Project Configuration:** Users enter their Google Cloud Project ID (e.g., "ee-username123"), which is required for Earth Engine API access
3. **Credential Storage:** After clicking "Save Configuration," the Project ID is stored securely and made available to all data retrieval operations
4. **Authentication Process:** When users click "Authenticate Earth Engine," a browser window opens to Google's OAuth service, allowing secure authentication without exposing sensitive credentials to the application
5. **Integration with Data Fetching:** Once authenticated, the Earth Engine connection is utilized by the data fetching components described previously to extract precipitation values from multiple gridded datasets at station locations

This streamlined process addresses a key challenge in climate data analysis: democratizing access to large-scale datasets without requiring technical expertise in cloud computing or geospatial APIs. By abstracting the complexity of Earth Engine operations, the application allows researchers to focus on scientific analysis rather than data acquisition mechanics.

The Earth Engine implementation incorporates background threading to prevent UI freezing during data retrieval operations. This asynchronous approach is particularly important when processing high-resolution datasets (such as DAYMET's 1km grid) or extracting long time series (e.g., 40+ years of daily data), which may involve thousands of API calls to extract point values at station locations.

2.3 Statistical Analysis Framework

2.3.1 Analysis Workflow

The statistical analysis module compares ground station measurements with each gridded product to evaluate their performance. The analysis follows these steps:

1. Align temporal coverage between ground data and gridded data
2. For each station and gridded product pair:
    - Calculate daily statistics
    - Aggregate to monthly values and calculate monthly statistics
    - Aggregate to yearly values and calculate yearly statistics
    - Segment by season and calculate seasonal statistics
    - Identify extreme precipitation events (high and low percentiles) and calculate statistics for these subsets
3. Compile results into structured CSV files for each gridded product

2.3.2 Performance Metrics

The application calculates several common statistical metrics to evaluate the performance of gridded products relative to ground observations. Each metric provides different insights into product performance:

1. **Coefficient of determination (R^2):** Measures the proportion of variance in ground observations explained by the gridded product. This metric ranges from 0 to 1, with higher values indicating better performance.

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

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

2. **Root Mean Square Error (RMSE):** Measures the average magnitude of errors, giving more weight to larger errors due to the squaring operation. Lower values indicate better performance.

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

3. **Bias:** Measures the average tendency of the gridded product to overestimate or underestimate precipitation. Values closer to zero indicate less systematic bias.

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

4. **Mean Absolute Error (MAE):** Measures the average magnitude of errors without considering their direction. Lower values indicate better performance.

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

5. **Nash-Sutcliffe Efficiency (NSE):** Indicates how well the gridded product predicts ground observations relative to the mean of observed data. NSE ranges from $-\infty$ to 1, with 1 indicating perfect match, 0 indicating performance equal to using the observed mean, and negative values indicating worse performance than using the observed mean.

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

6. **Percent Bias (PBIAS):** Measures the average tendency of the gridded product to be larger or smaller than ground observations, expressed as a percentage. Values closer to zero indicate less systematic bias.

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

These metrics are calculated at different temporal scales (daily, monthly, yearly) and for different data subsets (seasonal, extreme events) to provide a comprehensive evaluation of gridded product performance.

2.3.3 Statistical Outlier Filtering

The application implements a direction-aware filtering mechanism that removes extreme statistical values while preserving good performance metrics. This feature addresses the challenge of outliers disproportionately affecting visualization and summary statistics in precipitation product comparisons.

The filtering approach employs the following principles:

- **Direction-aware filtering:** The implementation only filters metric values in the "bad" direction:
  - For metrics where higher is better (R^2, NSE, correlation), only extremely low values below the specified lower percentile are filtered
  - For metrics where lower is better (RMSE, MAE, bias), only extremely high values above the specified upper percentile are filtered
- **Configurable percentile thresholds:** Users can specify lower and upper percentile bounds (default 1-99%) through the user interface
- **Preservation of good results:** By only filtering in the "bad" direction, stations with good performance metrics are never removed from analysis regardless of how extreme their good values are

This filtering mechanism is particularly valuable when analyzing large geographical regions where individual stations might have unusual characteristics or data quality issues. The application's implementation ensures that summary statistics and visualizations remain representative of overall performance while removing potential anomalies that could distort comparisons between precipitation products.

2.4 Visualization Generation

The visualization module generates a variety of plots to aid in the interpretation of analysis results. These visualizations are designed to highlight different aspects of gridded product performance and are implemented using Matplotlib, Seaborn, GeoPandas, and Contextily. The application supports the following visualization types:

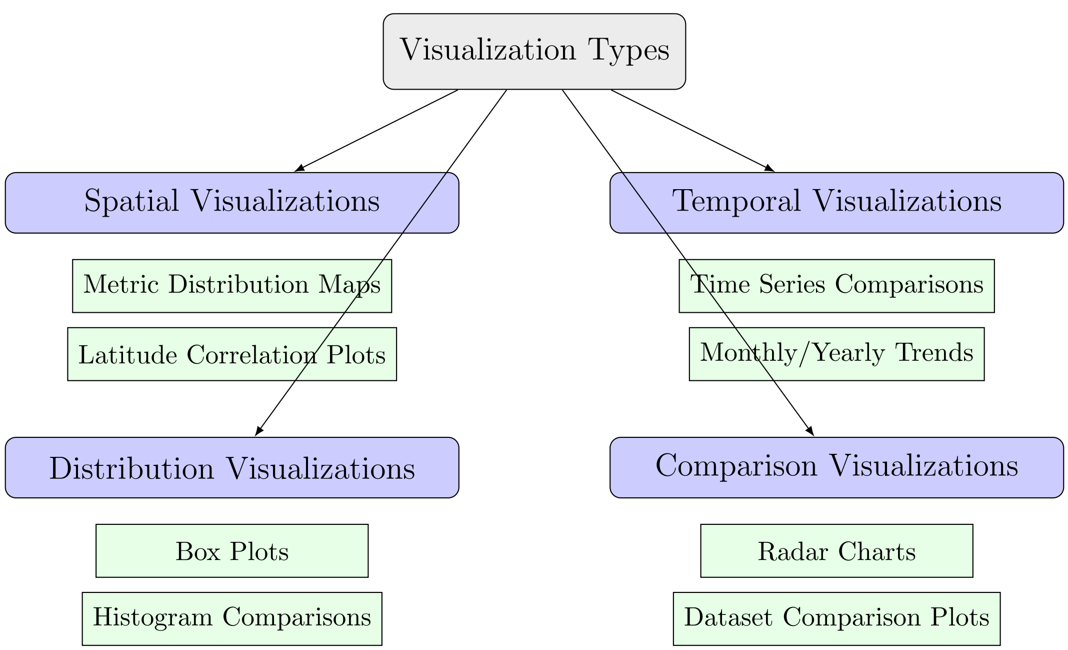

Figure 3: Visualization types supported by the Gridded Data Validator GUI

Each visualization type serves a specific analytical purpose:

1. **Spatial Distribution Maps:** These visualizations display the spatial variation of performance metrics across the study area. Each point on the map represents a ground station, with color coding indicating the value of the selected metric (e.g., R^2, RMSE, bias) at that location. These maps help identify regional patterns in gridded product performance, such as coastal effects, elevation dependencies, or urban/rural differences.

2. **Time Series Comparisons:** These plots show the temporal evolution of precipitation from both ground observations and gridded products. Multiple time scales can be visualized, including:

    - **Daily time series:** Shows day-to-day variability and the ability of gridded products to capture daily precipitation events
    - **Monthly time series:** Highlights seasonal patterns and monthly biases
    - **Yearly time series:** Reveals long-term trends and interannual variability

3. **Box Plots:** These visualizations summarize the distribution of performance metrics across all stations in the study area. They show the median, quartiles, and outliers for each metric, enabling quick comparison between different gridded products. Box plots can be generated for:

    - Each performance metric (R^2, RMSE, bias, MAE, NSE, PBIAS)
    - Different temporal scales (daily, monthly, yearly)
    - Seasonal subsets (winter, spring, summer, fall)
    - Extreme precipitation events (high and low percentiles)

4. **Seasonal Comparison Charts:** These visualizations focus on seasonal variations in gridded product performance. They help identify which products perform better in specific seasons, which is critical for applications that focus on particular times of the year (e.g., flood forecasting during spring, drought monitoring during summer).

5. **Dataset Comparison Plots:** These visualizations directly compare multiple gridded products using a common metric. They allow users to quickly identify which products perform best overall or for specific metrics.

6. **Radar Charts:** These multi-dimensional visualizations show performance across multiple metrics simultaneously. Each axis of the radar chart represents a different metric, and each gridded product is represented by a polygon. Radar charts provide a comprehensive view of product performance, highlighting strengths and weaknesses across multiple dimensions.

7. **Extreme Value Analysis Plots:** These specialized visualizations focus on how well gridded products capture extreme precipitation events (both high and low). They include:

- Scatter plots of extreme values
- Performance metrics calculated only for extreme events
- Return period comparisons

All visualizations are automatically saved as high-resolution PNG files and can be viewed and exported through the application's Visualization Panel.

2.5 Application Workflow

The typical workflow for using the Gridded Data Validator GUI is illustrated in Figure 4 and consists of the following steps:

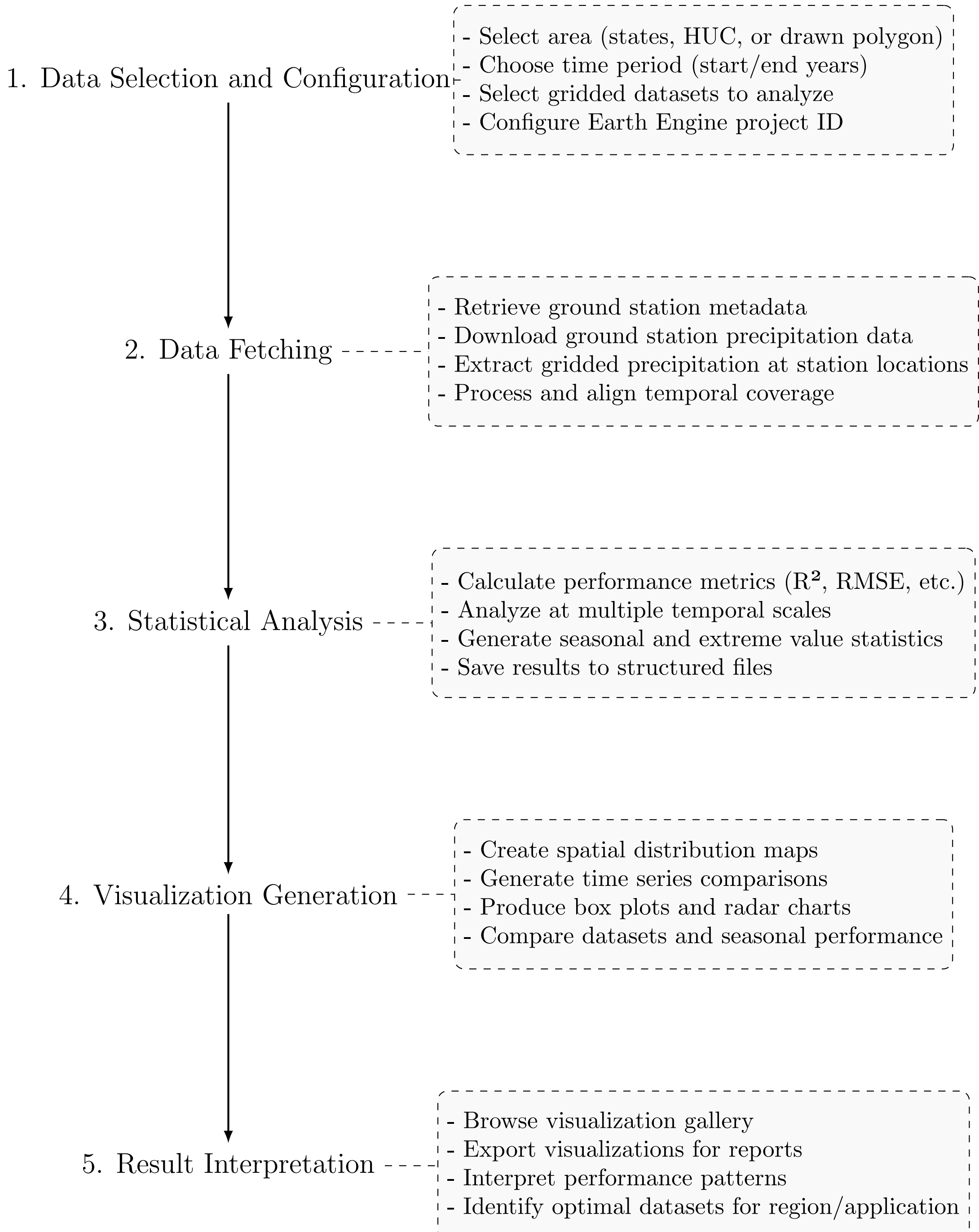

Figure 4: Typical workflow for using the Gridded Data Validator GUI

1. **Data Selection and Configuration:**
    - User selects the area of interest using one of the three available methods
    - User specifies the time range for analysis
    - User selects the gridded products to include in the analysis
    - User configures any additional settings (e.g., Earth Engine project ID)

2. **Data Fetching:**
    - Upon clicking "Download Data," the application retrieves ground station data and gridded product data
    - Progress is displayed in real-time through a progress bar
    - Data is saved to CSV files in the application's data directory
3. **Statistical Analysis:**
    - User navigates to the Analysis panel and configures analysis options
    - Upon clicking "Run Analysis," the application calculates performance metrics
    - Results are displayed in the UI and saved to CSV files
4. **Visualization Generation:**
    - User navigates to the Visualization panel and selects visualization options
    - Upon clicking "Generate Visualizations," the application creates the specified visualizations
    - Generated visualizations are saved as image files
5. **Result Interpretation and Export:**
    - User browses the visualization gallery in the Visualization panel
    - User selects visualizations to view in detail
    - User exports selected visualizations for use in reports or presentations

3 User Interface Snapshots

The Gridded Data Validator GUI provides an intuitive interface organized into three main panels: Data Selection, Analysis, and Visualization. Below are representative screenshots of the application's interface components.

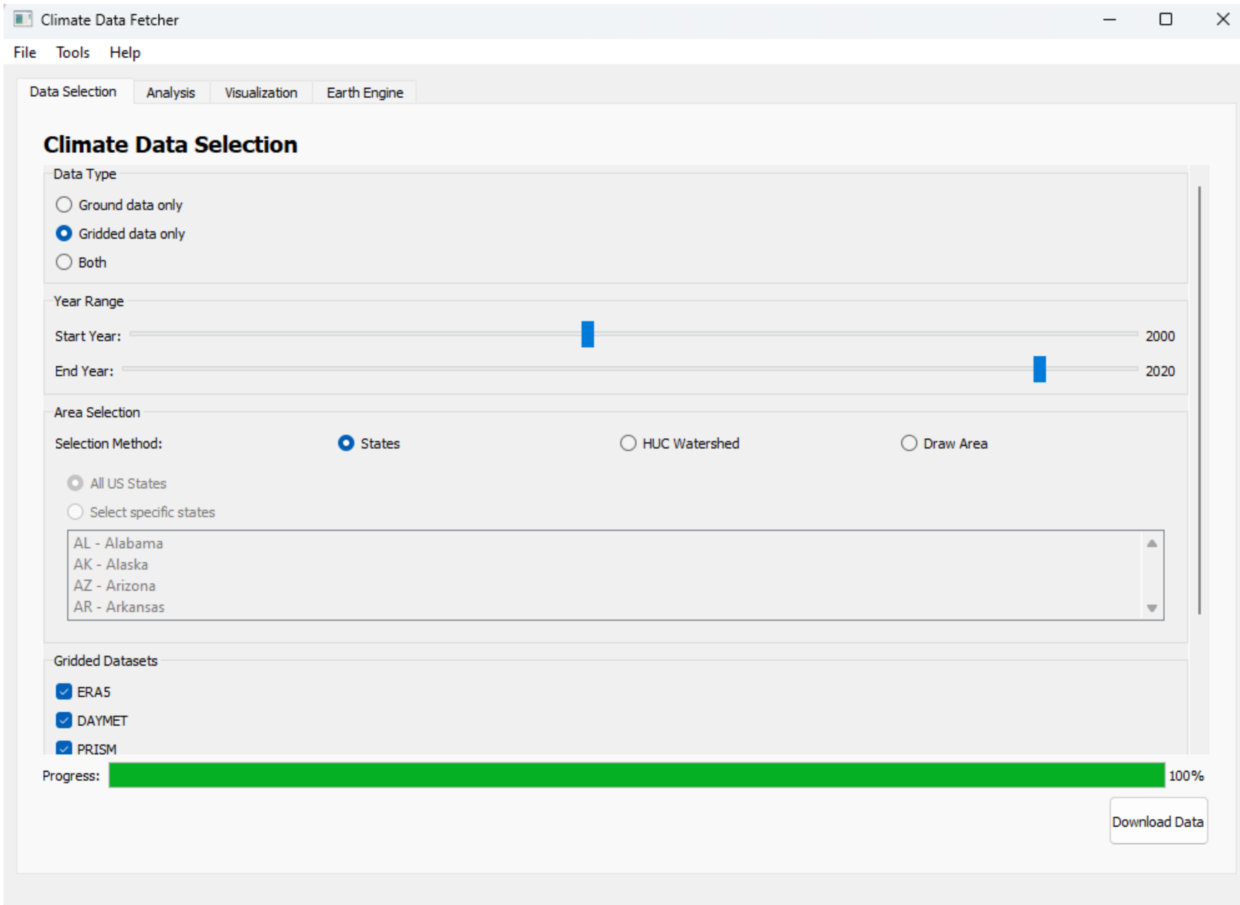

Figure 5: Data Selection Panel showing area selection options (states, HUC watersheds, or custom polygon), time period selection, and gridded dataset options. This panel allows users to configure data fetching operations and monitor download progress.

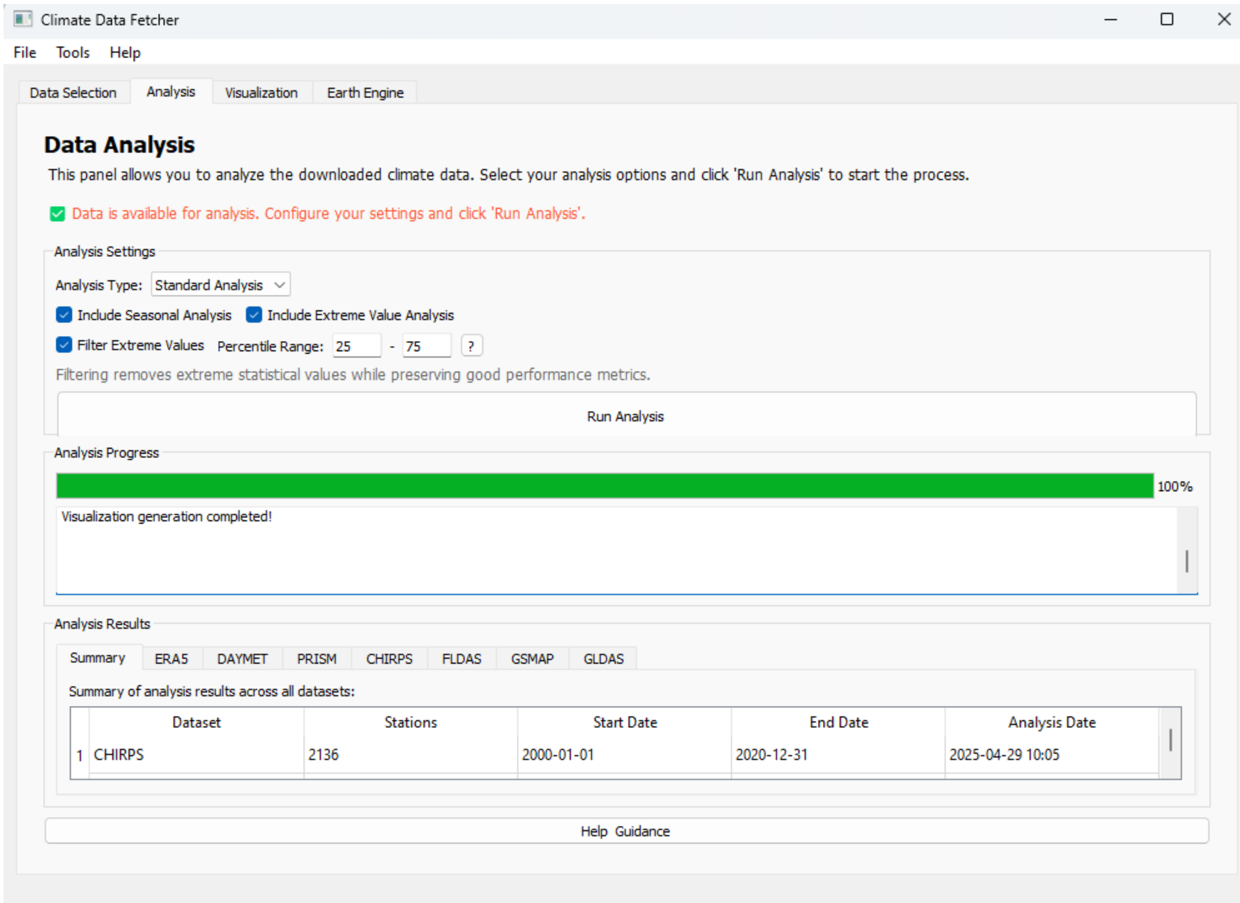

Figure 6: Analysis Panel displaying statistical results for multiple gridded datasets. The panel presents summary statistics and detailed metrics for each dataset, with tabs for different gridded products. Users can configure analysis settings and view results in tabular format.

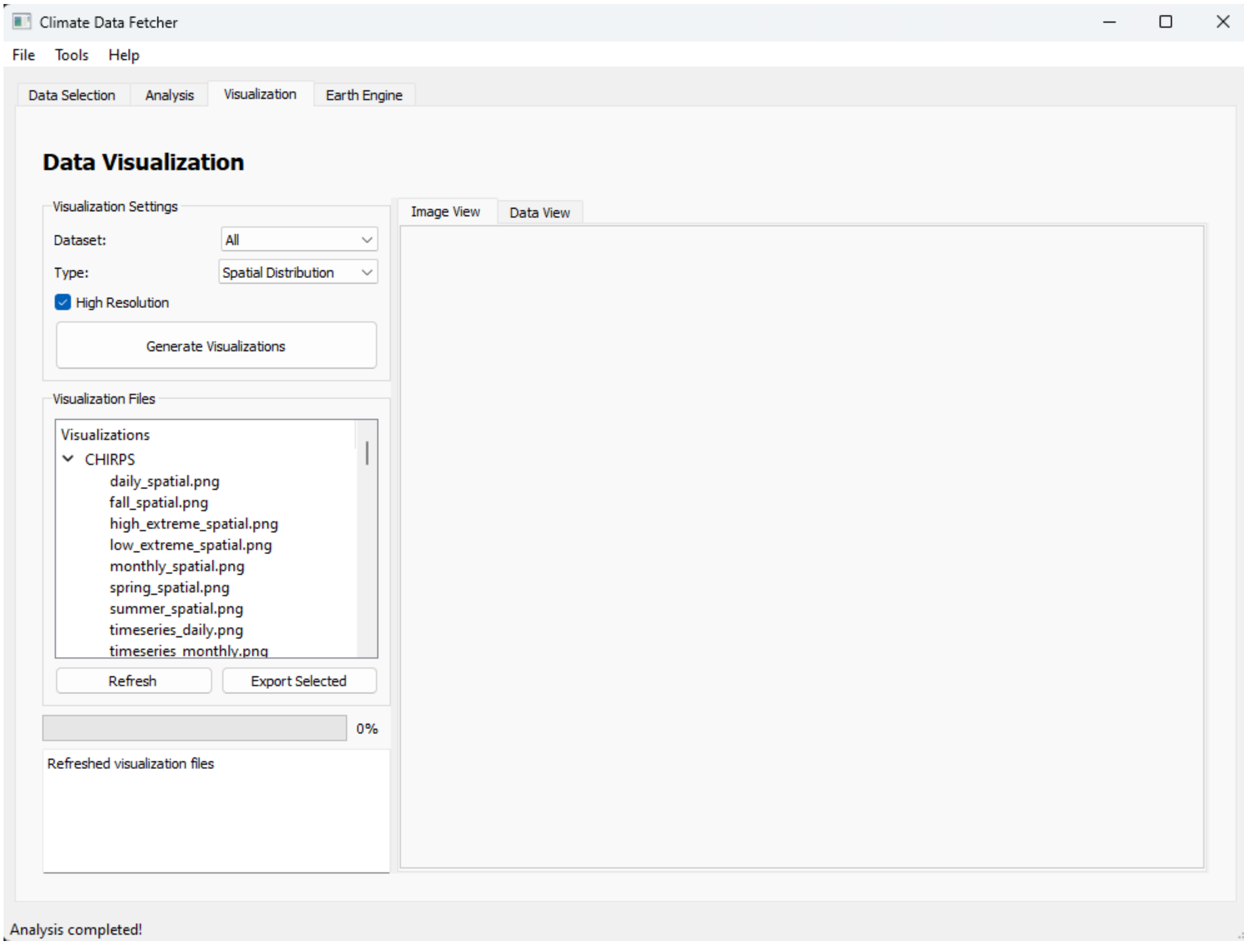

Figure 7: Visualization Panel with visualization type selection, dataset filtering options, and an image viewer displaying a spatial distribution map of RMSE values. The panel includes a file browser for navigating generated visualizations and tools for exporting images.

4 Case Study: Comparative Analysis of Precipitation Products in the Continental United States

To demonstrate the capabilities of the Gridded Data Validator GUI, we present a comprehensive case study analyzing precipitation products across the Continental United States (CONUS). This case study illustrates the workflow and analytical capabilities of the application, showcasing how users can perform large-scale comparative analyses.

4.1 Study Configuration

The case study employed the following configuration:

- **Study Area:** Continental United States (CONUS)
- **Time Period:** 2000-2020 (21 years)

- **Gridded Products:** ERA5, DAYMET, PRISM, CHIRPS
- **Analysis:** Daily, monthly, and yearly statistics; seasonal analysis
- **Filtering:** Extreme statistical values filtered using 25th-75th percentile range

This configuration enabled a comprehensive assessment of multiple precipitation products across diverse climatic regions, elevations, and land cover types present within CONUS. The relatively conservative filtering approach (25th-75th percentile) was selected to focus the analysis on the most representative performance patterns while excluding potential statistical anomalies.

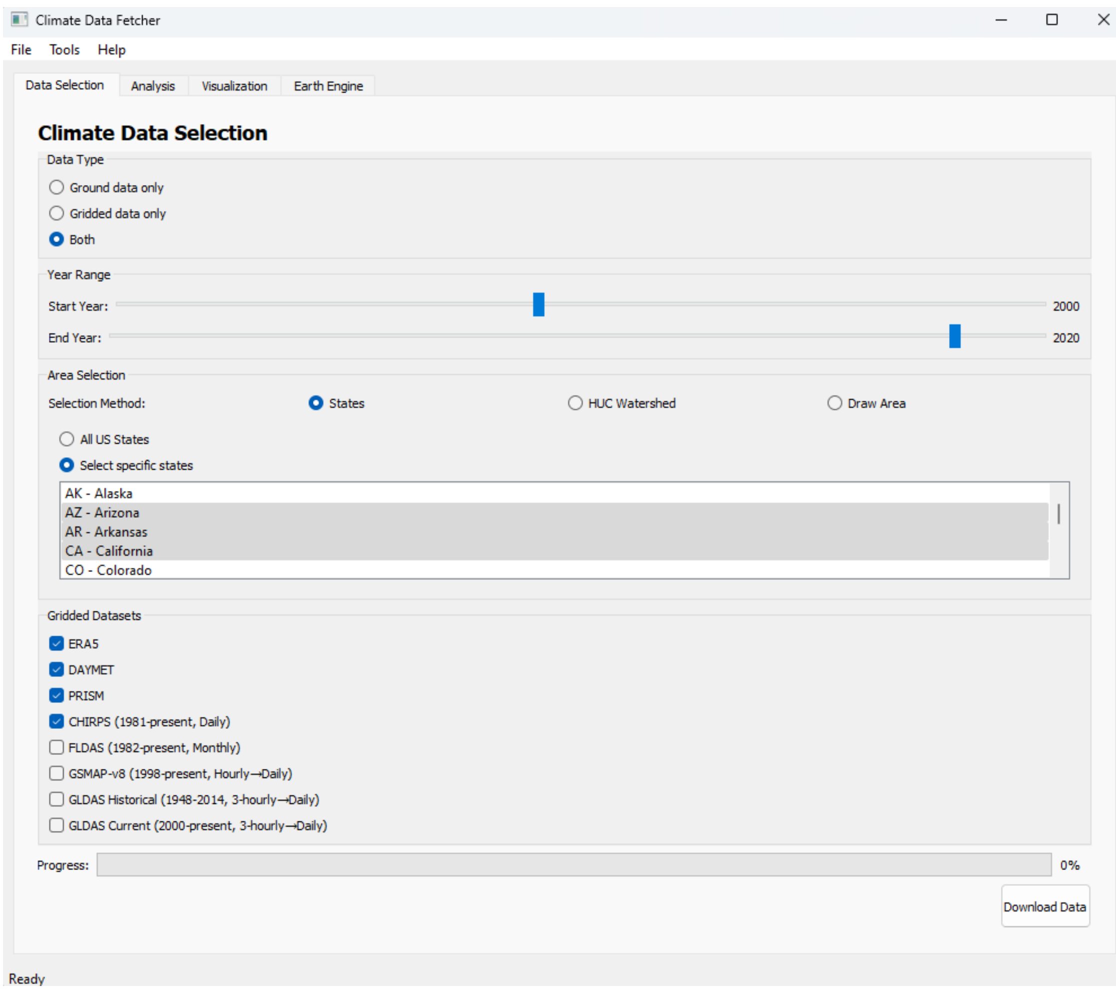

Figure 8: Data Selection Panel configured for CONUS-wide analysis. The interface shows: (A) selection of all US states, (B) time period configuration spanning 2000-2020, (C) selection of multiple precipitation products including ERA5, DAYMET, PRISM, and CHIRPS, and (D) progress indicators for the data download process.

Figure 8 shows the Data Selection Panel with the configuration used for this case study. The user interface provides an intuitive way to select the study area, time period, and precipitation products for analysis. In this case, all US states were selected to cover the entire CONUS region.

4.2 Analysis Process and Statistical Results

After data download completion, the Analysis Panel enables calculation of comprehensive statistical comparisons between ground observations and gridded precipitation products. For the CONUS analysis, the application processed data from over 3,000 ground stations, creating a robust evaluation framework.

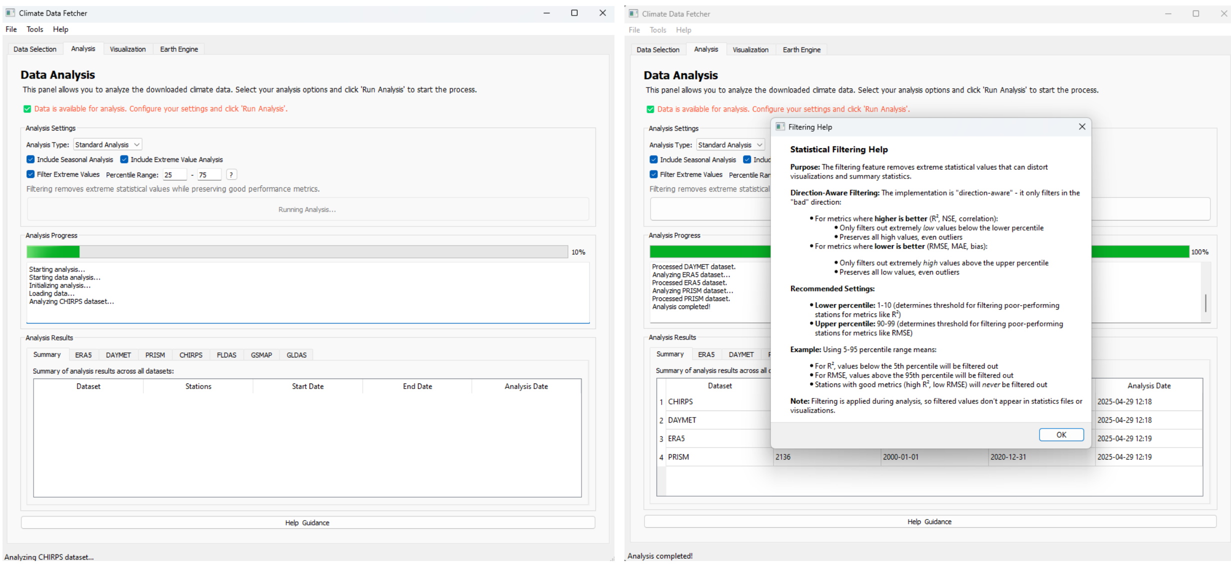

Figure 9: Left: Analysis Panel showing statistical results for CONUS analysis with performance metrics for different precipitation products. The panel displays daily, monthly, yearly, and seasonal statistics including R^2, RMSE, bias, and other metrics. Right: Visualization of the impact of 25th-75th percentile filtering on R^2 distribution, demonstrating how filtering removes statistical outliers while preserving valid data patterns.

As shown in Figure 9, the Analysis Panel presents statistical metrics across different temporal scales. The statistical analysis revealed significant differences in the performance of gridded products:

- Statistical comparisons showed varying performance patterns across products, with some products exhibiting stronger performance in specific regions
- The 25th-75th percentile filtering effectively removed extreme outliers that would otherwise distort the comparative analysis
- Performance metrics exhibited different patterns across temporal scales, with monthly aggregations generally showing higher agreement between ground and gridded data compared to daily values

4.3 Spatial Distribution Analysis

The spatial distribution of performance metrics provides crucial insights into the geographical patterns of product accuracy across CONUS.

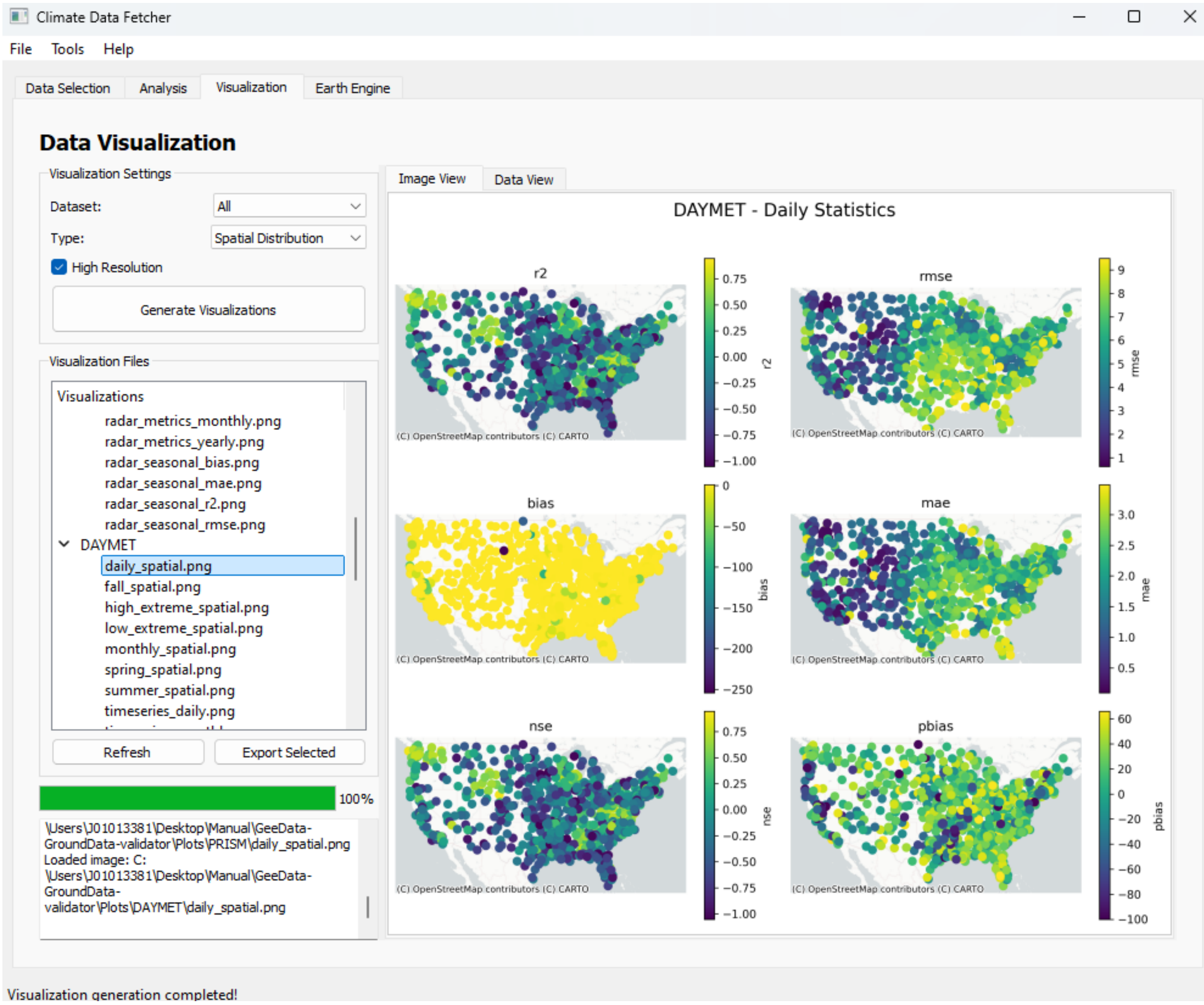

Figure 10: Spatial distribution of performance metrics across CONUS. Left: Distribution of R^2 values showing correlation between PRISM and ground observations. Right: Distribution of RMSE values highlighting regions with higher prediction errors. The maps reveal geographic patterns in product performance, including coastal effects, terrain influences, and regional variations.

The spatial visualizations in Figure 10 reveal important geographical patterns in product performance.

4.4 Temporal Analysis

The Gridded Data Validator GUI enables detailed temporal analysis, revealing how product performance varies across time and seasons.

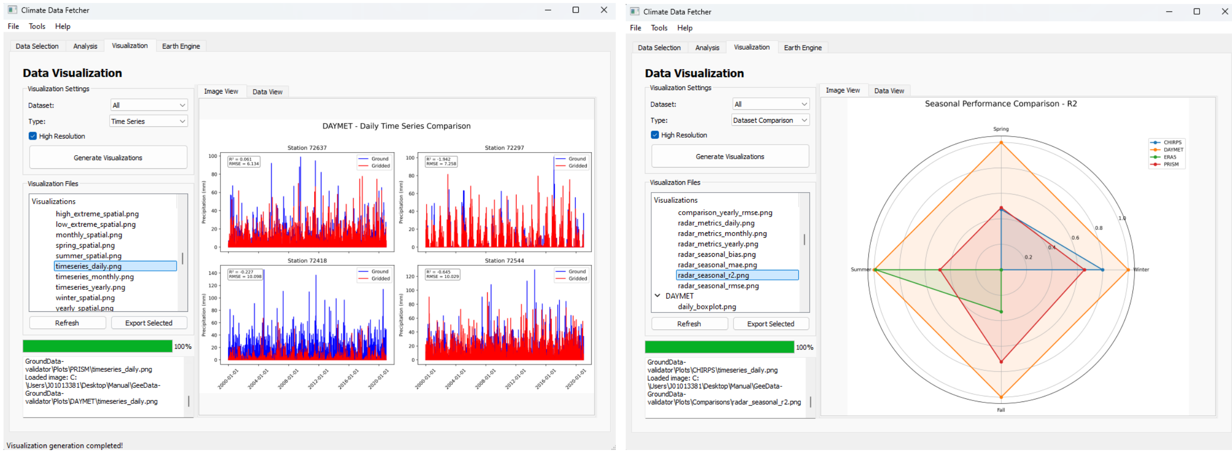

Figure 11: Temporal analysis visualizations. Left: Time series comparison showing precipitation patterns from ground observations and multiple gridded products for selected stations, with statistical metrics displayed for each comparison. Right: Seasonal performance patterns showing how different products perform across seasons, with metrics grouped by winter, spring, summer, and fall.

Figure 11 illustrates key temporal patterns identified through the analysis:

4.5 Cross-Product Comparative Visualization

A key strength of the Gridded Data Validator GUI is its ability to generate comprehensive cross-product comparisons through standardized visualizations.

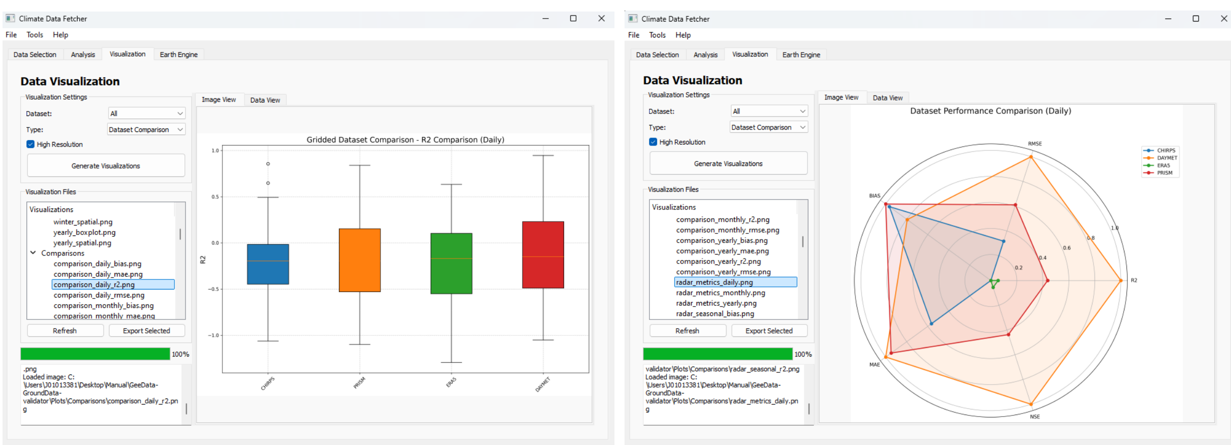

Figure 12: Cross-product comparison visualizations. Left: Box plot comparison of R^2, RMSE, bias, and MAE metrics across all precipitation products, showing the distribution of performance across CONUS. Right: Radar chart visualization displaying multi-metric performance for each dataset, enabling quick identification of relative strengths and weaknesses.

The comparative visualizations in Figure 12 facilitate comprehensive evaluation:

- Box plots reveal the full distribution of performance metrics, highlighting not just average performance but also consistency across stations
- The radar chart provides a multi-dimensional comparison that simultaneously visualizes performance across multiple metrics
- Visual patterns help identify each product's strengths and weaknesses, supporting informed selection for specific applications
- Statistical significance is visually apparent, helping users identify meaningful differences between products

4.6 User Workflow and Key Insights

The Gridded Data Validator GUI streamlines the entire workflow from data acquisition to visualization and insight generation.

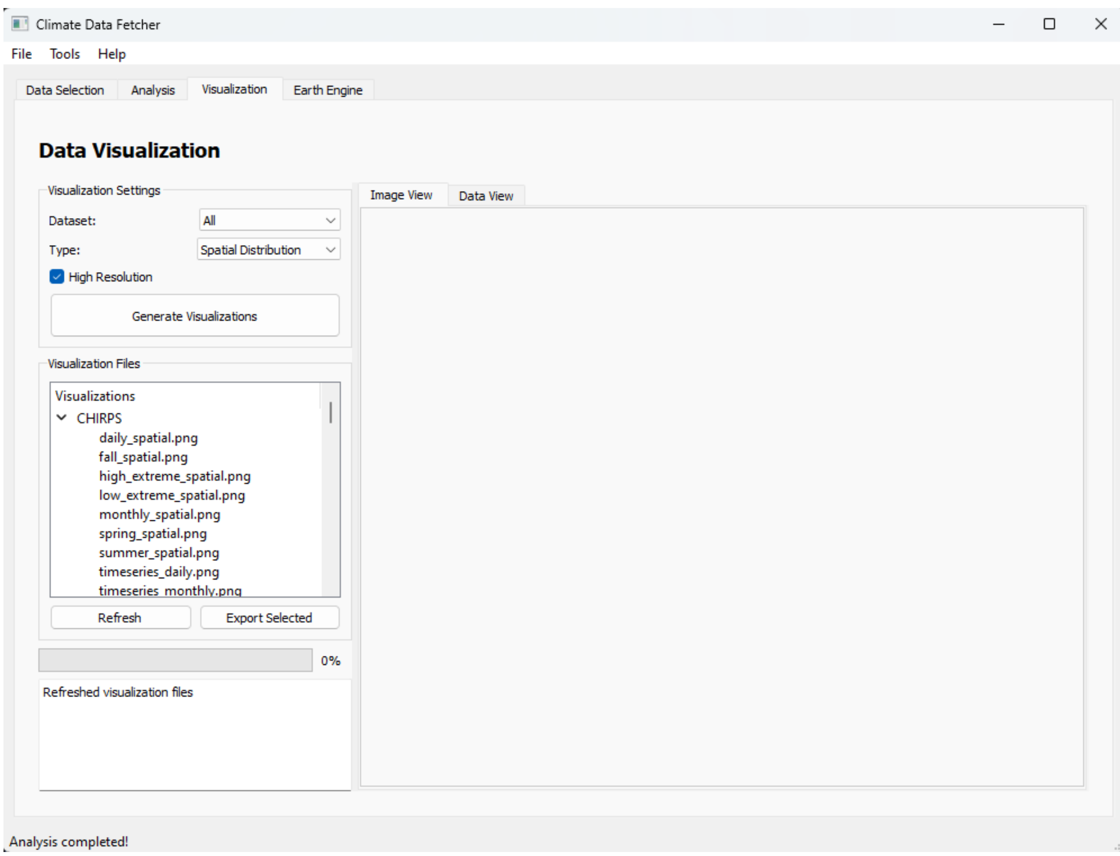

Figure 13: Visualization Panel showing: (A) visualization type selector for generating different analysis visualizations, (B) file browser for navigating generated visualization files, (C) preview area displaying selected visualization, and (D) export options for saving publication-ready figures. This integrated environment enables seamless transition from analysis to interpretation and communication of results.

As shown in Figure 13, the Visualization Panel enables users to:

- Generate multiple visualization types to explore different aspects of product performance
- Navigate through visualization files organized by product and analysis type
- Preview visualizations in high resolution
- Export publication-ready figures for reports, presentations, or scientific publications

This case study illustrates how the Gridded Data Validator GUI enables users to conduct sophisticated comparative analyses of precipitation products with minimal programming expertise. By streamlining the entire workflow from data acquisition to visualization, the tool facilitates the informed selection of appropriate precipitation products for diverse hydrological applications.

5 Summary and Recommendations

5.1 Summary

The Gridded Data Validator GUI provides a comprehensive solution for retrieving, analyzing, and visualizing precipitation data from multiple sources. The application streamlines what is typically a complex, multi-step process requiring expertise in multiple programming languages and APIs. Key features include:

- Unified access to ground station data and multiple gridded precipitation products
- Flexible area selection through states, HUC watersheds, or custom polygons
- Systematic statistical comparison at daily, monthly, yearly, and seasonal timescales
- Rich visualization capabilities to aid in interpretation of results
- User-friendly interface accessible to researchers without extensive programming expertise

The application's modular architecture facilitates maintenance and future extensions, while the implementation of background threading ensures responsiveness during long-running operations.

5.2 Limitations

Current limitations of the application include:

- Geographic scope is limited to the Continental United States due to the use of Meteostat for ground data
- Analysis is focused on precipitation only, not other meteorological variables

- Spatial analysis is limited to point-based comparison at station locations
- Limited support for direct export to hydrological models
- Requires local installation of Python and dependencies

5.3 Recommendations for Future Work

Based on the current implementation and identified limitations, we recommend the following areas for future development:

1. **Extended Geographic Coverage:** Incorporate additional ground data sources and develop UI options for users to import custom metadata and ground station data from countries beyond the US. This enhancement would leverage the existing Meteostat integration, which already contains global data for key stations (airports and other important weather monitoring locations), enabling the tool to have global relevance.
2. **Global Dataset Integration:** Extend the gridded dataset functionality to automatically access and compare global precipitation products when custom international ground data is imported. This would transform the application from a US-focused tool to one with worldwide applicability.
3. **Additional Variables:** Extend the application to support analysis of other meteorological variables such as temperature, relative humidity, and wind speed
4. **Advanced Spatial Analysis:** Implement gridded comparison methods that evaluate entire spatial fields rather than just point locations
5. **Bias Correction:** Add capabilities for applying and evaluating bias correction methods
6. **Model Integration:** Develop direct export options to common hydrological models such as GSSHA, HEC-HMS, SWAT, and VIC
7. **Web Application:** Develop a web-based version to eliminate installation requirements
8. **Additional Visualization Types:** Implement additional visualization types such as heat maps, animated time series, and interactive 3D visualizations
9. **API Integration:** Develop an API to allow programmatic access to the application's functionality

The transformation to a globally-relevant tool represents a significant enhancement to the Gridded Data Validator's utility. With international ground station data support and integration with worldwide precipitation products, the application could serve as a standard evaluation platform for precipitation data across diverse climatic regions and help researchers identify the most appropriate products for specific regions globally.

5.4 Implications for Practice

The Gridded Data Validator GUI has several practical implications for hydrological studies, climate research, and water resource management:

- **Improved Efficiency:** Streamlines a complex workflow that typically requires substantial time and expertise
- **Enhanced Understanding:** Provides detailed insights into the strengths and limitations of different precipitation products
- **Better Model Inputs:** Enables selection of optimal precipitation products for specific regions and applications
- **Enhanced Hydrological State:** The hydrological conditions of wetness and dryness are more accurately represented across both space and time
- **Increased Reproducibility:** Standardizes the evaluation process, enhancing reproducibility of results
- **Educational Value:** Serves as an educational tool for understanding the characteristics of different precipitation products

By addressing the limitations and implementing the recommended extensions, the Gridded Data Validator GUI can become an even more valuable tool for hydrometeorological research and water resource management.

6 References

References

[1] Bhattarai, S. (2025). Gridded Data Validator: A Python Tool for Precipitation Analysis. GitHub Repository. https://github.com/Saurav-JSU/GeeData-GroundData-validator

A Accessing the Source Code

The complete source code for the Gridded Data Validator GUI is available on GitHub at the following URL:
https://github.com/Saurav-JSU/GeeData-GroundData-validator

B Installation Instructions

To install and run the Gridded Data Validator GUI, follow these steps:

1. Clone the GitHub repository:
```
git clone https://Saurav-JSU/Saurav-JSU/GeeData-GroundData-validator.git
cd GeeData-GroundData-validator
```
2. Install the required dependencies:
```
pip install -r requirements.txt
```
3. Set up Earth Engine authentication:
```
earthengine authenticate
```
4. Run the application:
```
python main.py
```

C SF 298 Form Page

[To be inserted by editor]