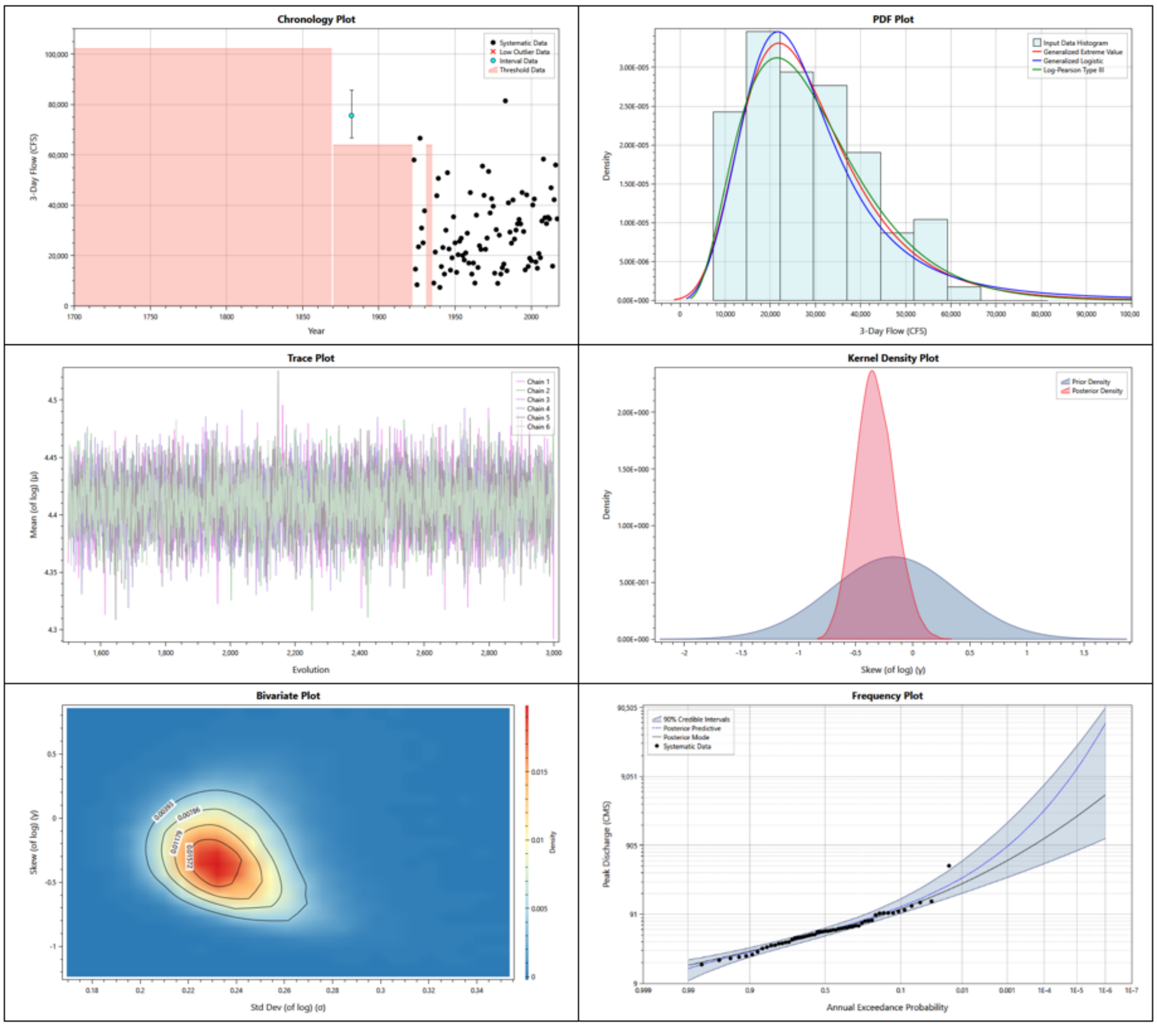
RMC-BestFit Quick Start Guide

RMC-TR-2020-03



US Army Corps of Engineers®
Institute for Water Resources
Risk Management Center

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
08-14-2020

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
RMC-BestFit Quick Start Guide

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Haden Smith, Risk Management Center, U.S. Army Corps of Engineers
Megan Doughty, Risk Management Center, U.S. Army Corps of Engineers

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Risk Management Center
12596 West Bayaud Ave. Suite 400
Lakewood, CO 80228

8. PERFORMING ORGANIZATION REPORT NUMBER
RMC-TR-2020-03

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
RMC-TR-2020-03

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
RMC-BestFit is designed to enhance and expedite flood hazard assessments within the Flood Risk Management, Planning, and Dam and Levee Safety communities of practice. RMC-BestFit is a menu-driven software package, which performs distribution fitting and Bayesian estimation from a choice of thirteen probability distributions. The software features a fully integrated modeling platform, including a modern graphical user interface, data entry capabilities, distribution fitting analysis, Bayesian estimation analysis, and report quality charts.

15. SUBJECT TERMS
Software, Bayesian estimation, distribution fitting, frequency analysis, probability distributions, uncertainty

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	a. REPORT	a. REPORT			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	99	

RMC-BestFit Quick Start Guide

August 2020

U.S. Army Corps of Engineers
Institute for Water Resources
Risk Management Center
12596 West Bayaud Ave. Suite 400
Lakewood, CO 80228

Authors

Haden Smith, USACE Risk Management Center

Megan Doughty, USACE Risk Management Center

Contents

Figures

Tables

This page blank intentionally

Acknowledgements

RMC-BestFit would not exist without support of RMC leadership, in particular the RMC Lead Engineers, David Margo and John England. RMC-BestFit was developed in collaboration with Brian Skahill (ERDC-CHL), who has made significant contributions within USACE toward the advancement of Bayesian estimation methods and tools. The development team would like to also recognize Ruben Jongejan (Risk Management Consulting B.V.), who performed an external peer review of the USACE methodology for assessing hydrologic uncertainty (Jongejan, 2018). His recommendation that USACE explore a Bayesian approach for handling distribution parameter uncertainties in flood frequency analyses served as a catalyst for the development of the RMC-BestFit software. The development team is very grateful to those who helped contribute to the software.

Welcome to RMC-BestFit

The U.S. Army Corps of Engineers (USACE) Risk Management Center (RMC), in collaboration with the Engineer Research and Development Center (ERDC) Coastal and Hydraulics Laboratory (CHL), developed the Bayesian estimation and fitting software (RMC-BestFit) to enhance and expedite flood hazard assessments within the Flood Risk Management, Planning, and Dam and Levee Safety communities of practice.

RMC-BestFit is a menu-driven software package, which performs distribution fitting and Bayesian estimation from a choice of thirteen probability distributions. The software features a fully integrated modeling platform, including a modern graphical user interface, data entry capabilities, distribution fitting analysis, Bayesian estimation analysis, and report quality charts.

Why you should use RMC-BestFit

In Bayesian estimation, the values of the probability distribution parameters converge to a distribution rather than to a single best value. Quantification of the parameter uncertainty is particularly important in risk analysis because it allows you to assess the value of reducing the uncertainty. Typically, the parameter uncertainty can be reduced with more and better information through means of additional measurement, data collection and quality control, filling gaps in missing gage data, and record extension. With Bayesian analysis, we can also specify our prior knowledge on the parameters through expert elicitation, causal modelling, or regional analysis.

The Bayesian approach is capable of incorporating all available sources of information into an analysis. For example, in a Bayesian flood frequency analysis, at-site gage data, paleoflood information, regional information, and causal rainfall-runoff results can be incorporated into the flood frequency curve, which can then be used to evaluate reservoir pool stage-frequency for use in Dam Safety risk assessments.

The Bayesian estimation approach is flexible and coherent. Its assumptions are made explicit, and the analysis is repeatable and revisable. The ability of the Bayesian approach to use all pieces of information in conjunction is a major advantage over other traditional statistical methods, such as those that are recommended in Bulletin 17B (U.S. Geological Survey, 1982) and Bulletin 17C (U.S. Geological Survey, 2018).

When should you use RMC-BestFit? Any time you perform a risk analysis that involves uncertainty, you can and should use RMC-BestFit. Typical applications in science and engineering include reliability analysis, and rainfall and flood frequency analysis.

System Requirements

The RMC-BestFit program and all dependent libraries were developed using the .NET Framework 4.6.1. As such, the program is currently only available for the Microsoft Windows operating system. The recommended system configuration for RMC-BestFit includes:

- 64-bit Windows Operating System (Windows 7 or newer)
- 512MB of RAM at a minimum
- 2GB of free disk space at a minimum
- Quad-Core CPU with 2.7GHz (or faster) clock speed
- 1280x1024 screen resolution or higher (and at least a 17" monitor)

Installing RMC-BestFit

RMC-BestFit version 1.0 is available as a portable package (.zip). Installing with a portable package does not require administrative rights. Simply extract all contents of the portable package to the desired computer location. After extraction, the package contents will look as shown in Figure 1. Simply double-click the executable **RMC-BestFit.exe** to get started. For easy access to the program, you can create a desktop shortcut and pin the program to your Windows taskbar.

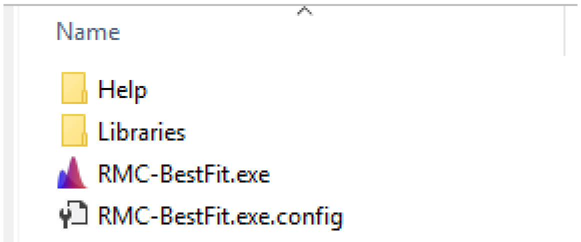


Figure 1 – Installing RMC-BestFit.

Setting the File Association in Windows

RMC-BestFit files have the "**.bestfit**" file extension. To have Windows automatically open **.bestfit** files with RMC-BestFit, first right-click a **.bestfit** project file. Select **Open with…** from the resulting menu.

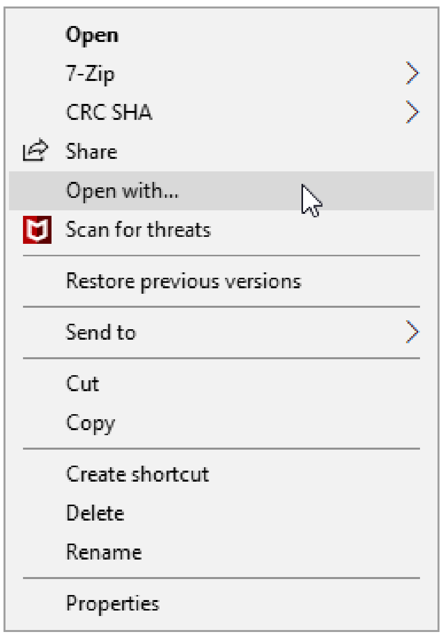


Figure 2 – Right-Click Menu for Project File in Windows.

Then, from the menu that appears when you select **Open with...**, click **Try an app on this PC** for an expanded list of already installed applications. If you do not see RMC-BestFit, scroll to the bottom and select **Look for another app on this PC**. This will open a Windows Explorer dialog. Navigate to and select the executable **RMC-BestFit.exe**, then click **Open.** When you've found RMC-BestFit and it has been selected, check the box labeled **Always use this app to open [.bestfit] files** before you click the **OK** button. Now **.bestfit** files will appear in Windows Explorer with the RMC-BestFit icon. You can now double-click a project file to automatically open it with RMC-BestFit.

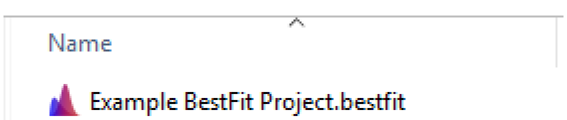


Figure 3 – RMC-BestFit Project with File Association in Windows.

Terms and Conditions for Use

The United States Government, US Army Corps of Engineers, Risk Management Center ("RMC") grants to the user the rights to install RMC-BestFit "the Software" (either from a copy obtained from RMC, a distributor or another user or by downloading it from a network) and to use, copy and/or distribute copies of the Software to other users, subject to the following Terms and Conditions of Use:

All copies of the Software received or reproduced by or for user pursuant to the authority of this Terms and Conditions of Use will be and remain the property of RMC. User may reproduce and distribute the Software provided that the recipient agrees to the Terms and Conditions for Use noted herein.

RMC is solely responsible for the content of the Software. The Software may not be modified, abridged, decompiled, disassembled, unobfuscated or reverse engineered. The user is solely responsible for the content, interactions, and effects of any and all amendments, if present, whether they be extension modules, language resource bundles, scripts or any other amendment.

The name "RMC-BestFit" must not be used to endorse or promote products derived from the Software. Products derived from the Software may not be called "RMC-BestFit" nor may any part of the "RMC-BestFit" name appear within the name of derived products. No part of this Terms and Conditions for Use may be modified, deleted or obliterated from the Software. No part of the Software may be exported or re-exported in contravention of U.S. export laws or regulations.

Waiver of Warranty

THE UNITED STATES GOVERNMENT AND ITS AGENCIES, OFFICIALS, REPRESENTATIVES, AND EMPLOYEES, INCLUDING ITS CONTRACTORS AND SUPPLIERS PROVIDE RMC-BESTFIT "AS IS," WITHOUT ANY WARRANTY OR CONDITION, EXPRESS, IMPLIED OR STATUTORY, AND SPECIFICALLY DISCLAIM ANY IMPLIED WARRANTIES OF TITLE, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT. Depending on state law, the foregoing disclaimer may not apply to you, and you may also have other legal rights that vary from state to state.

Limitation of Liability

IN NO EVENT SHALL THE UNITED STATES GOVERNMENT AND ITS AGENCIES, OFFICIALS, REPRESENTATIVES, AND EMPLOYEES, INCLUDING ITS CONTRACTORS AND SUPPLIERS, BE LIABLE FOR LOST PROFITS OR ANY SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF OR IN CONNECTION WITH USE OF RMC-BESTFIT REGARDLESS OF CAUSE, INCLUDING NEGLIGENCE.

THE UNITED STATES GOVERNMENT’S LIABILITY, AND THE LIABILITY OF ITS AGENCIES, OFFICIALS, REPRESENTATIVES, AND EMPLOYEES, INCLUDING ITS CONTRACTORS AND SUPPLIERS, TO YOU OR ANY THIRD PARTIES IN ANY CIRCUMSTANCE IS LIMITED TO THE REPLACEMENT OF CERTIFIED COPIES OF RMC-BESTFIT WITH IDENTIFIED ERRORS CORRECTED. Depending on state law, the above limitation or exclusion may not apply to you.

Indemnity

As a voluntary user of RMC-BestFit you agree to indemnify and hold the United States Government, and its agencies, officials, representatives, and employees, including its contractors and suppliers, harmless from any claim or demand, including reasonable attorneys' fees, made by any third party due to or arising out of your use of RMC-BestFit or breach of this Agreement or your violation of any law or the rights of a third party.

Assent

By using this program you voluntarily accept these terms and conditions. If you do not agree to these terms and conditions, uninstall the program and return any program materials to RMC (If you downloaded the program and do not have disk media, please delete all copies, and cease using the program).

Graphical User Interface

RMC-BestFit is a menu-driven software package that performs distribution fitting and Bayesian estimation from a choice of thirteen probability distributions. The software features a fully integrated modeling platform, including a modern graphical user interface, data entry capabilities, distribution fitting analysis, Bayesian estimation analysis, and report quality charts.

The graphical user interface consists of a **Menu Bar**, **Tool Bar**, and four window panes. Starting from the left in Figure 4 and moving clockwise, the panes will be referred to as the **Project Explorer**, the **Tabbed Documents**, the **Properties Window**, and the **Message Window**. You may move, dock, hide, or close the window panes as desired.

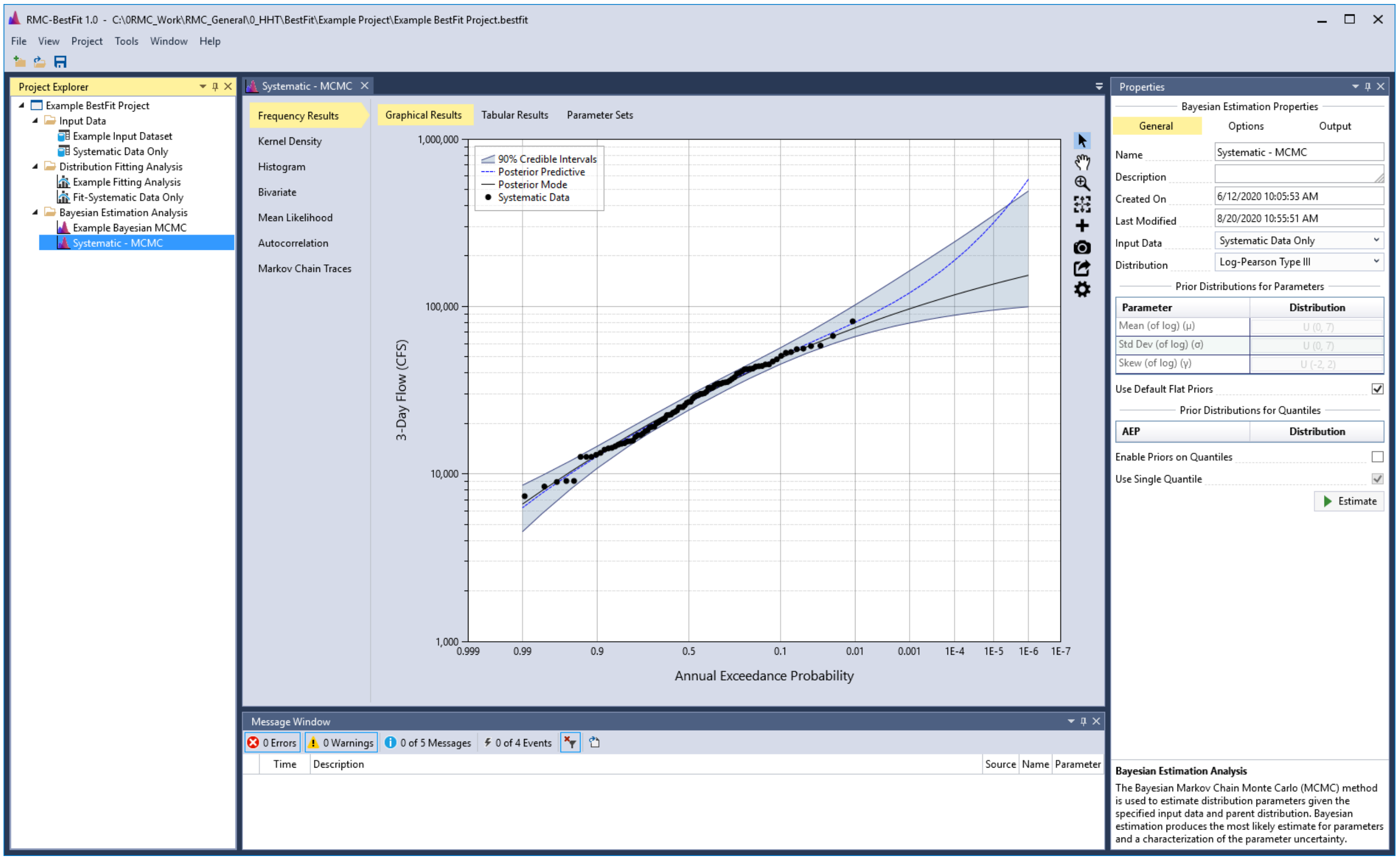


Figure 4 – RMC-BestFit User Interface.

Menu Bar

The menu bar located at the top of main window, as shown in Figure 5, contains important commands needed for working with RMC-BestFit. For example, the **Project** menu contains commands related to the project you are working in. On the **Tools** menu, you can customize how RMC-BestFit behaves by selecting **Options…**, or restore the project from a backup file by selecting **Restore from Backup**.

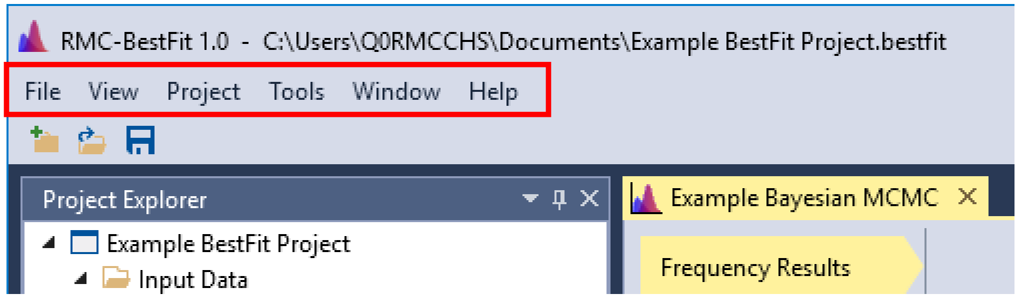


Figure 5 – RMC-BestFit Menu Strip.

File

The **File** menu provides essential file management functionality. From this menu, you may create a new project, open an existing project, save or save as, open recent projects, or exit the application.

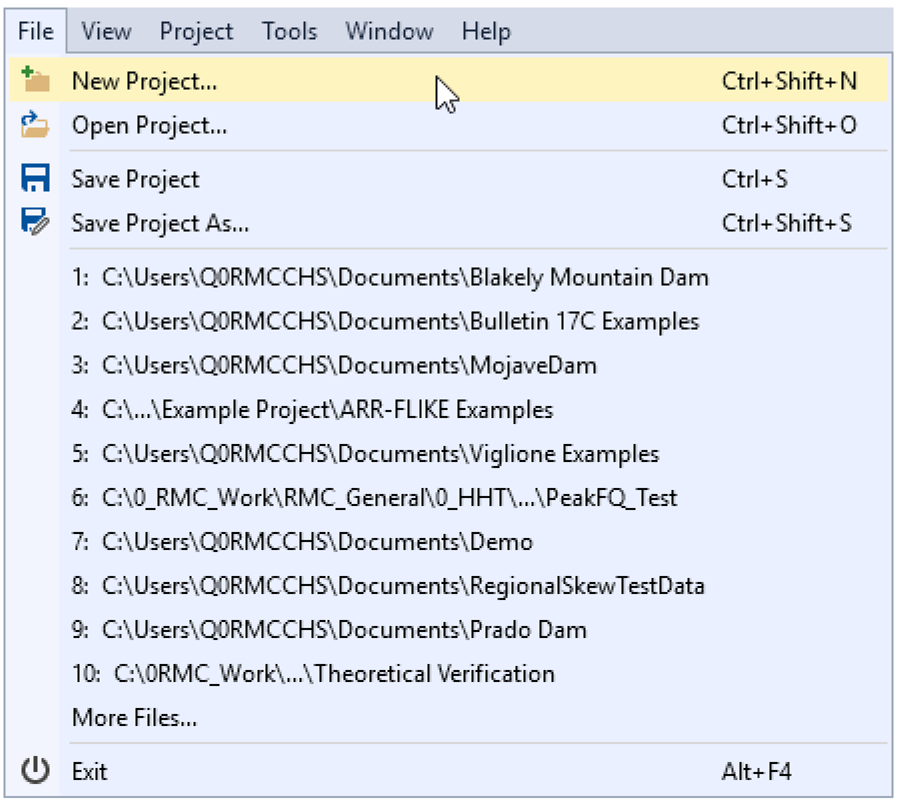


Figure 6 – File Menu.

If this is the first time you're using RMC-BestFit, your recent projects list will be empty. Once you have exceeded the number of projects shown in the recent project list (the default is 10), a menu item labeled as **More Files…** will be available. After clicking this item, a **Recent Files** dialog will open showing all available recent project files (Figure 7). From here, you may open any recently used project file, or clear the recent project file list.

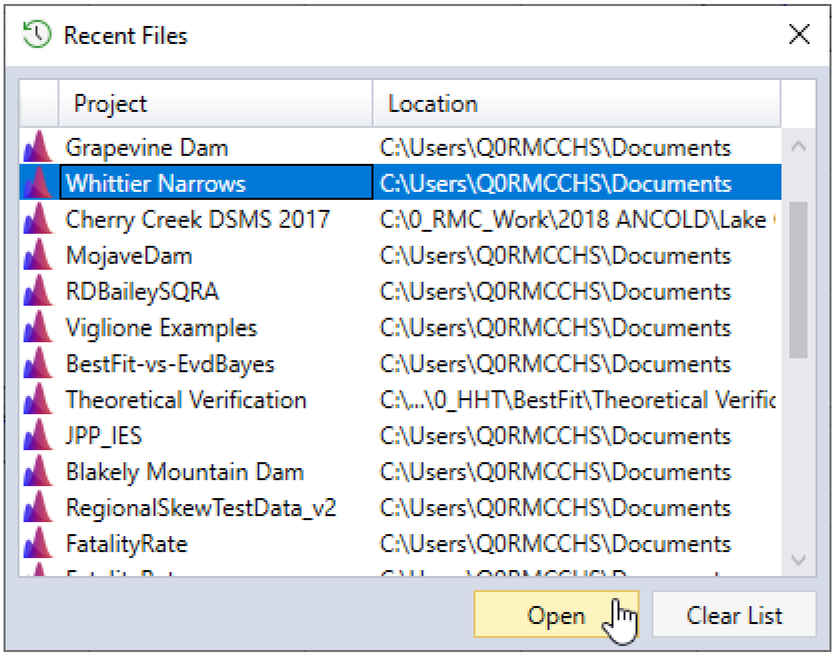


Figure 7 – Recent File Dialog.

View

You may move, dock, hide, or close the **Project Explorer**, **Properties Window**, and **Message Window** as desired. The **View** menu allows you to unhide or open these windows. In addition, you can restore the default layout of the window panes.

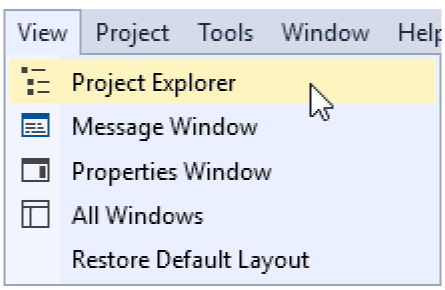


Figure 8 – View Menu.

Project

The **Project** menu contains commands related to the project you are working in. From this menu, you can create a new input data, distribution fitting analysis, or Bayesian estimation analysis. In addition, you can edit the project properties.

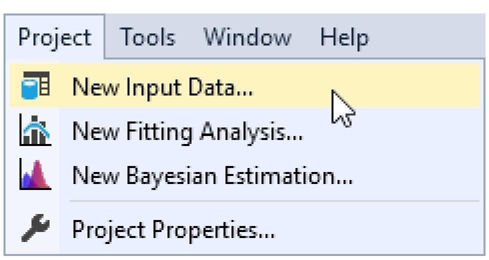


Figure 9 – Project Menu.

Tools

The **Tools** menu provides important tools for managing your project file. On the **Tools** menu, you can customize how RMC-BestFit behaves by selecting **Options…**, or restore the project from a backup file by selecting **Restore from Backup**. See the Personalize RMC-BestFit section for more details.

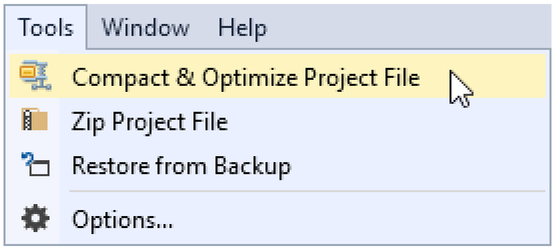


Figure 10 – Tools Menu.

Window

The **Window** menu allows you to close or activate the document windows. The active document will have a check mark next to it as shown in Figure 11. You can see all open windows by clicking **Windows…**, which will open a **Windows** dialog. From here, you can activate or save specific documents. In addition, you can select and close multiple documents at once.

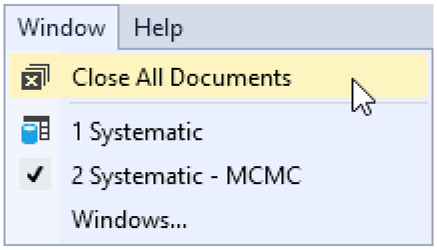


Figure 11 – Windows Menu.

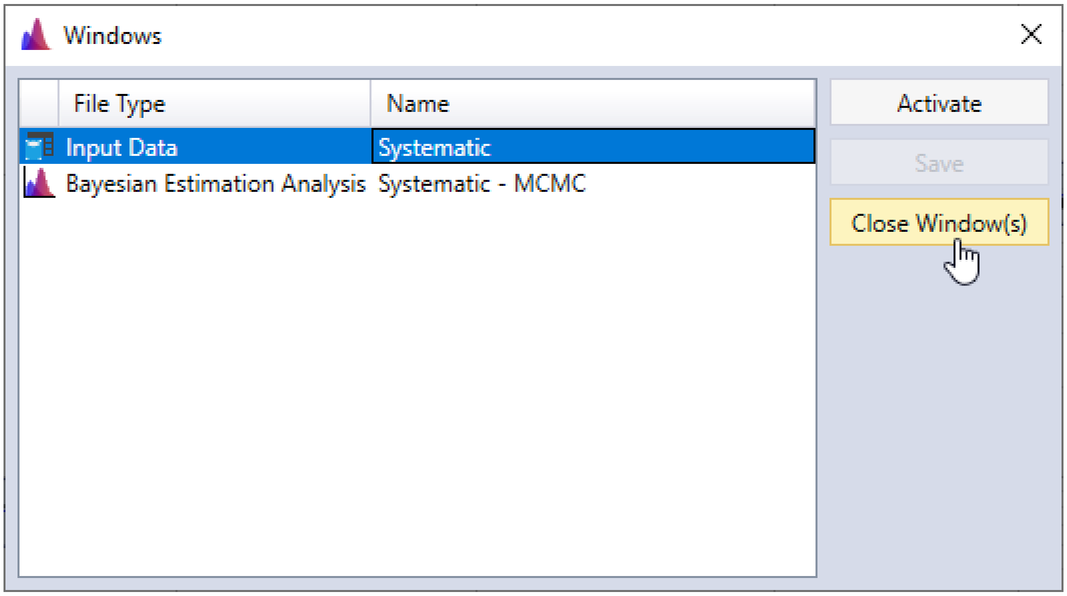


Figure 12 – Windows Dialog.

Help

From the **Help** menu, you can access this **Quick Start Guide**, view the **Terms & Conditions for Use**, or view the **About RMC-BestFit** splash screen.

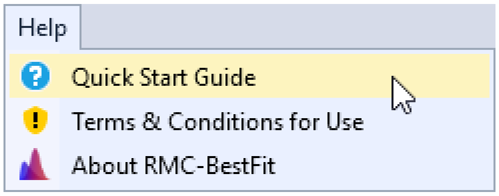


Figure 13 – Help Menu.

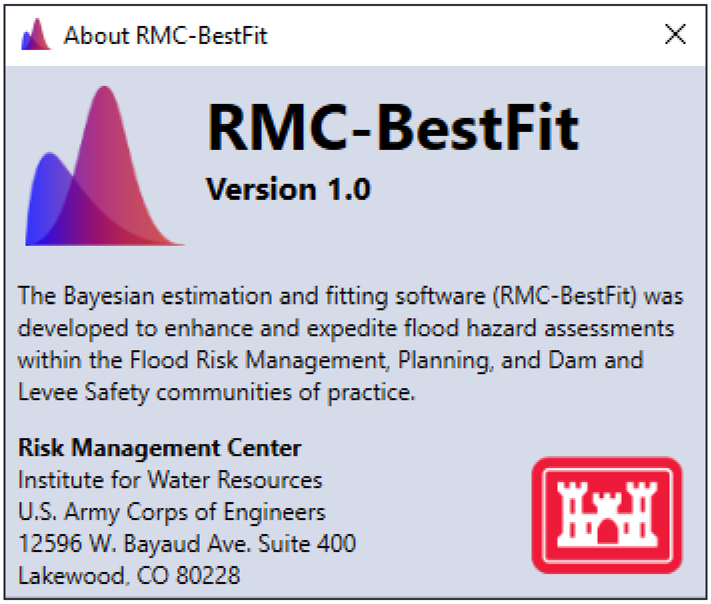


Figure 14 – About RMC-BestFit Splash Screen.

Tool Bar

The **Tool Bar** is located on the main window, below the **Menu Bar**. The buttons on the tool bar provide the most frequently used options under the **File** menu:

- New Project
- Open Project
- Save Project

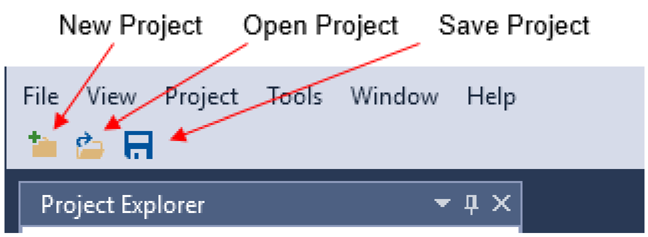


Figure 15 – RMC-BestFit Tool Bar.

Window Layout

In RMC-BestFit, you can customize the position, size, and behavior of windows to create window layouts that work best for you. When you customize the layout, RMC-BestFit will remember the configuration. For example, if you change the docking location of the **Project Explorer** and then close RMC-BestFit, the next time that you open the software, the **Project Explorer** will be docked in that same location.

Types of Windows

RMC-BestFit has two basic window types, *tool windows* and *document windows*. Tool windows include the **Project Explorer**, **Properties Window**, and **Message Window**. Document windows contain the project element files, such as input data, distribution fitting analyses, and Bayesian estimation analyses. The tool windows can be resized and dragged by their title bar, whereas the document windows can be dragged by their tab. On the tool windows title bar, there is a drop-down with other window options. Likewise, you can right-click on the document tab to set other options on the window. These options include docking, floating, and hiding windows.

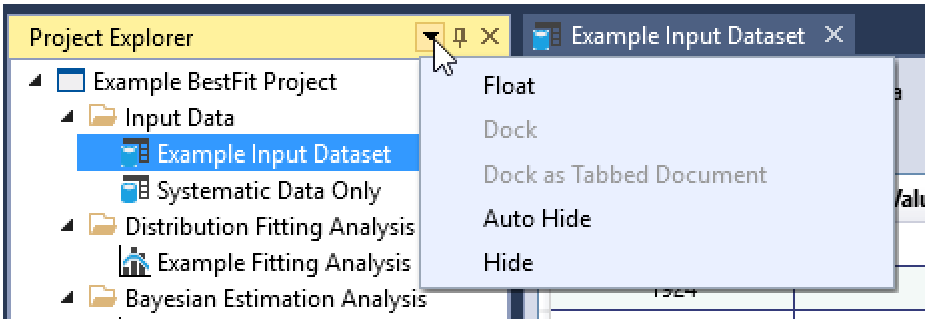


Figure 16 – Tools Window Options.

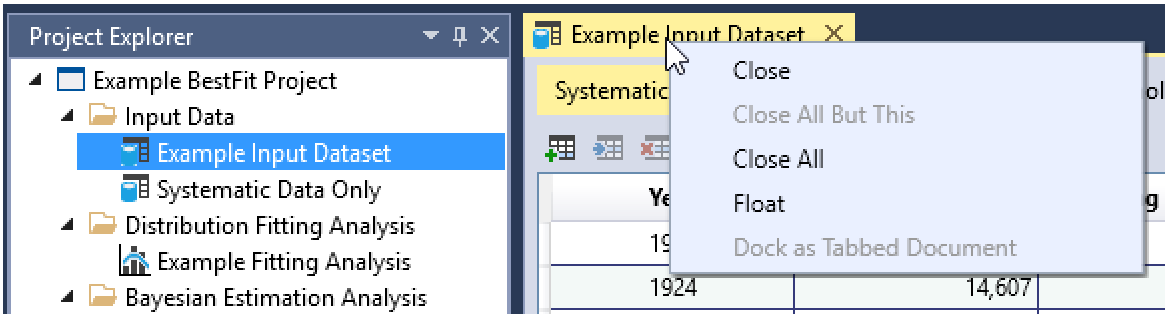


Figure 17 – Document Window Options.

Tab Groups

The windows can be grouped together to enhance your ability to manage a limited workspace while you are working with two or more open documents in RMC-BestFit. You can organize multiple document windows and tool windows into either vertical or horizontal tab groups, or move documents from one group to another. The windows and tab groups can also be floated and moved to different monitors. When you need to view or edit two documents at once, you can split windows by creating two horizontal tab groups as shown below in Figure 18.

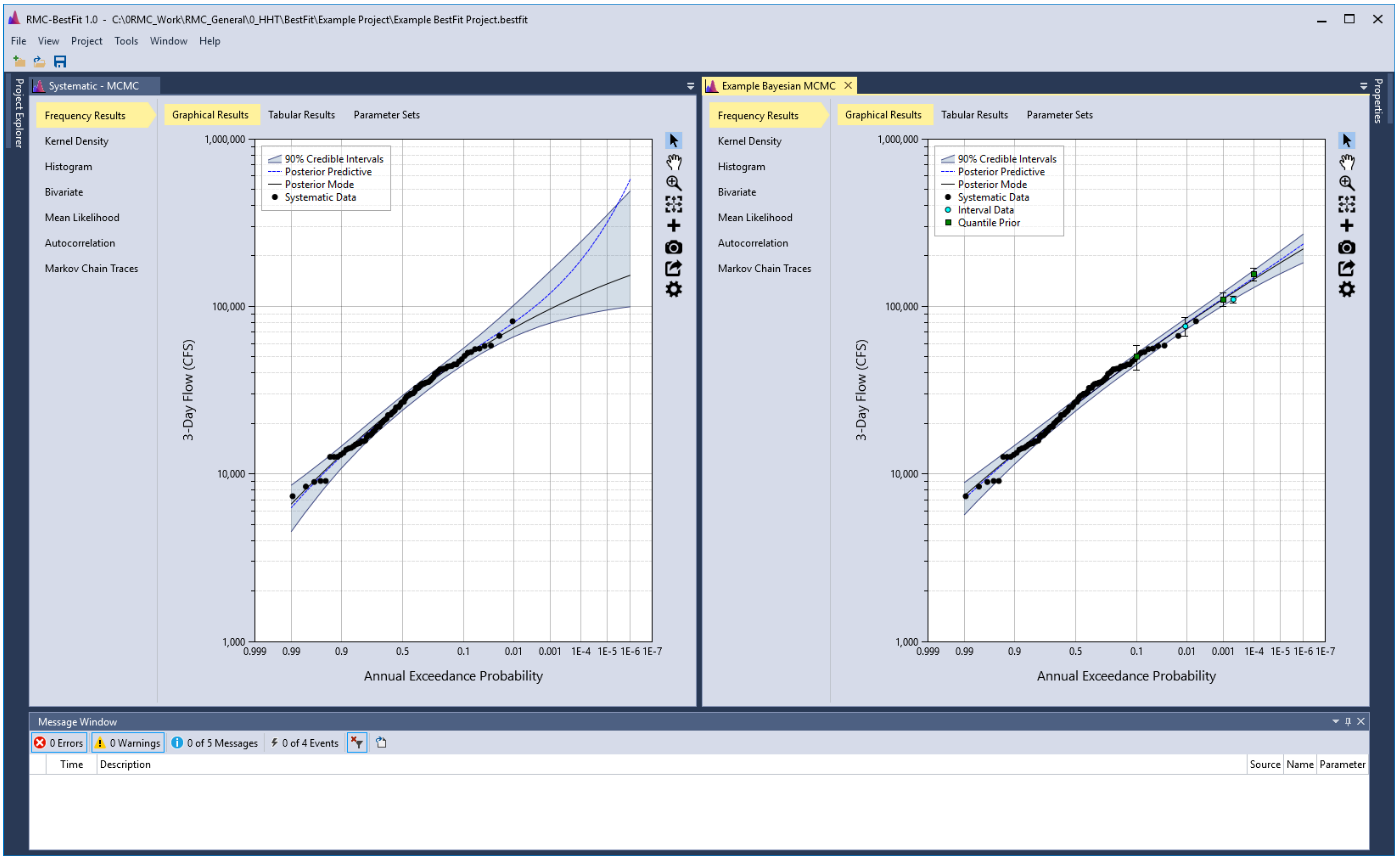


Figure 18 – Split Document Windows.

Arrange and Dock Windows

A document window can be *docked*, so that it will be located in the **Tabbed Documents** area, or the document can be *floated* as a separate window independent of the main window. You can arrange windows in the following ways:

- Dock tool windows to the edge of the main window frame.
- Float tool and document windows over or outside the main window.
- Hide tool windows along the edge of the main window.
- Display windows on different monitors.
- Reset window placement to the default layout by choosing **View** > **Reset Default Layout**.

Arrange windows by dragging or right-clicking the title bar or tab of the window to be arranged.

Dock Windows

When you click and drag the title bar of a tool window, or the tab of a document window, a cross shaped window placement guide will appear. During the drag operation, when the mouse cursor is over one of the arrows in the guide, a shaded area will appear that shows you where the window will be docked if you release the mouse button. An example of the window placement guide is shown below in Figure 19 and Figure 20.

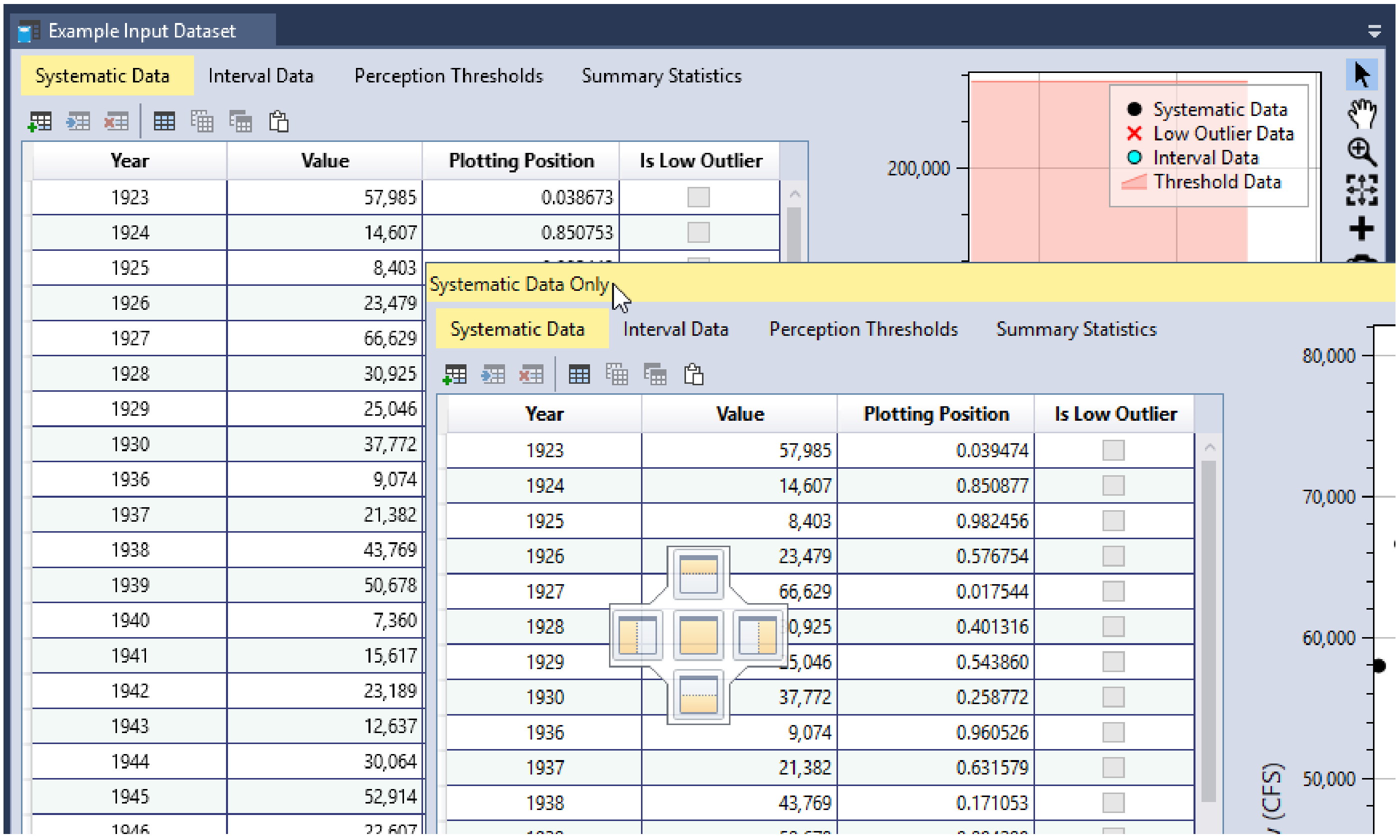


Figure 19 – Moving and Docking Document Windows.

Figure 20 shows the **Properties Windows** being docked below the **Project Explorer**. The new location is demarcated by the light blue shared area.

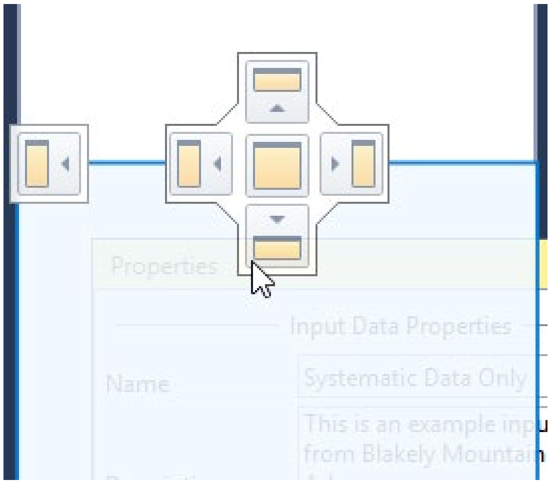


Figure 20 – Example of Docking the Properties Window Below the Project Explorer.

Close and Auto-Hide Tool Windows

You can close a tool window by clicking on the **X** in the upper right of the title bar. To reopen the tool window, navigate to the **View** menu and select the desired tool window to show. Tool windows have an *auto-hide* feature, which causes a window to slide out of the way when you use a different window. When a window is auto-hidden, the window name appears on the tab at the edge of the main window as shown in Figure 21. To show the window again, move your mouse cursor over the tab so that the window slides back into view.

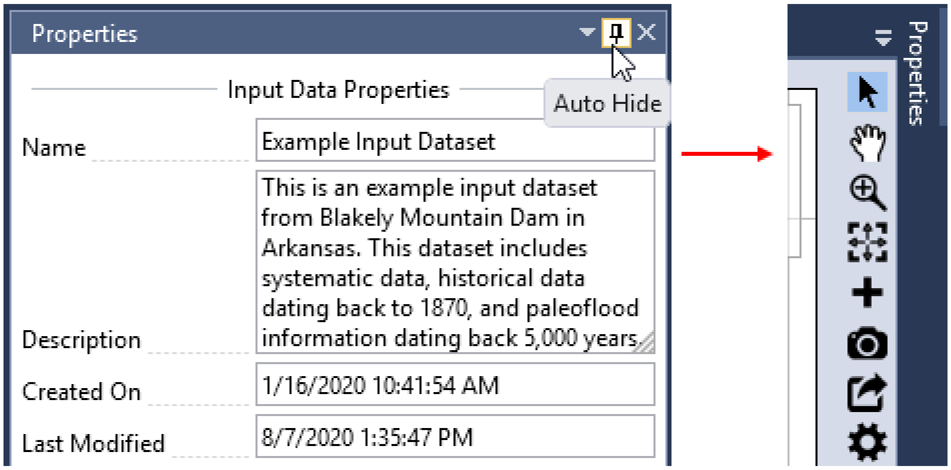


Figure 21 – Auto-Hide Tool Window.

Project Explorer

The **Project Explorer**, which is typically on the left-hand side of the main window, shows a graphical representation of the hierarchy of elements within your project. After you create a new project, you can use the **Project Explorer** to view, navigate, and manage the project elements.

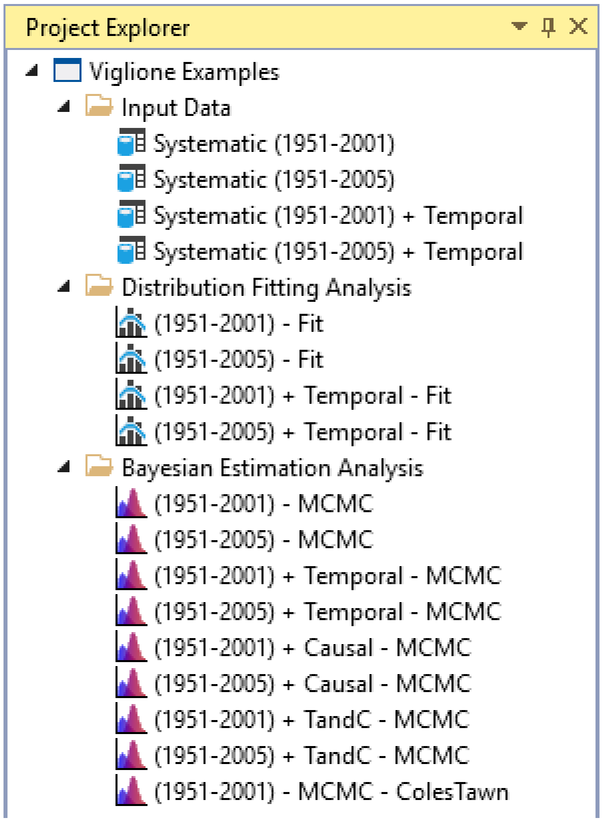


Figure 22 – Project Explorer.

Project elements are organized under the **Input Data**, **Distribution Fitting Analysis**, and **Bayesian Estimation Analysis** folder headers. Many menu commands are available from the right-click menu on various items in the Project Explorer. You can create new elements, or sort the element collections, by right-clicking the header.

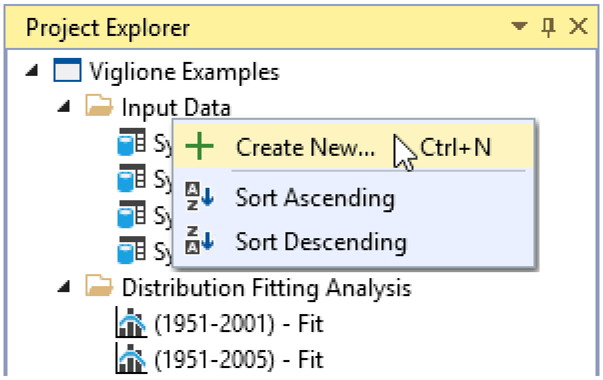


Figure 23 – Project Explorer Element Collection Header Right-Click Menu.

When right-clicking an individual project element, the following commands are made available: edit, copy, rename, delete, move up, and move down. When multiple project elements are selected, only two right-click menu commands are available: edit and delete. Double-clicking a project element will open it for editing.

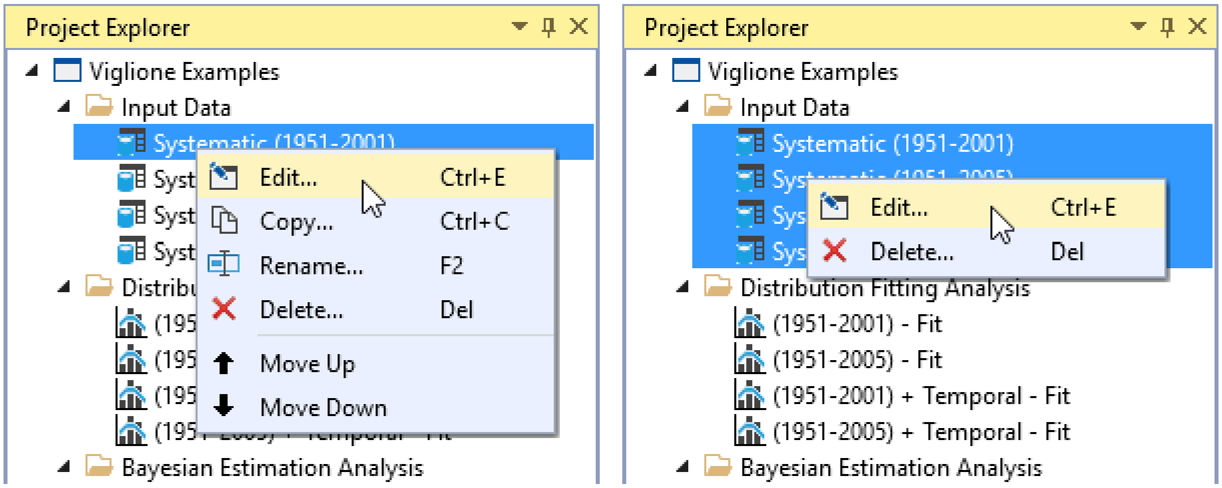


Figure 24 – Project Explorer Element Right-Click Menu.

You can move individual project elements within a collection by left-clicking and dragging the element. A mouse cursor adornment will appear indicating where the element will be moved, as shown in Figure 25. You can also drag and drop Input Data from one RMC-BestFit project to another.

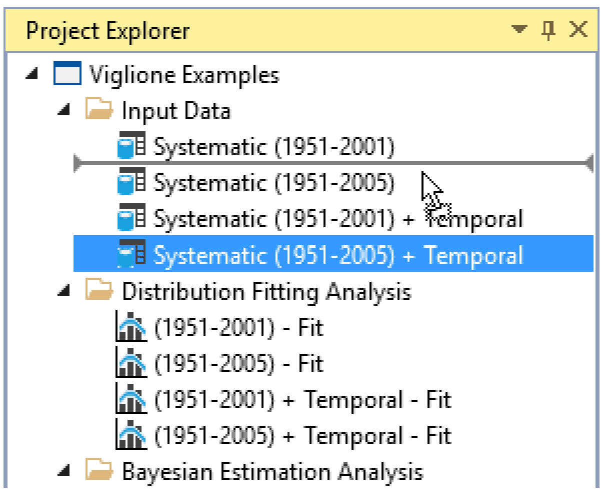


Figure 25 – Project Explorer Element Drag and Drop.

Tabbed Documents

In RMC-BestFit, *document windows* contain the project element files, such as input data, distribution fitting analyses, and Bayesian estimation analyses. When you open a project element for editing, it will automatically open into the tabbed document group, which is typically located in the center of the main window (see Figure 26). You can reorder the tabbed documents, drag, or float them outside of the main window. When you click on, or activate, a document window, the associated project element properties will be shown in the **Properties Window**.

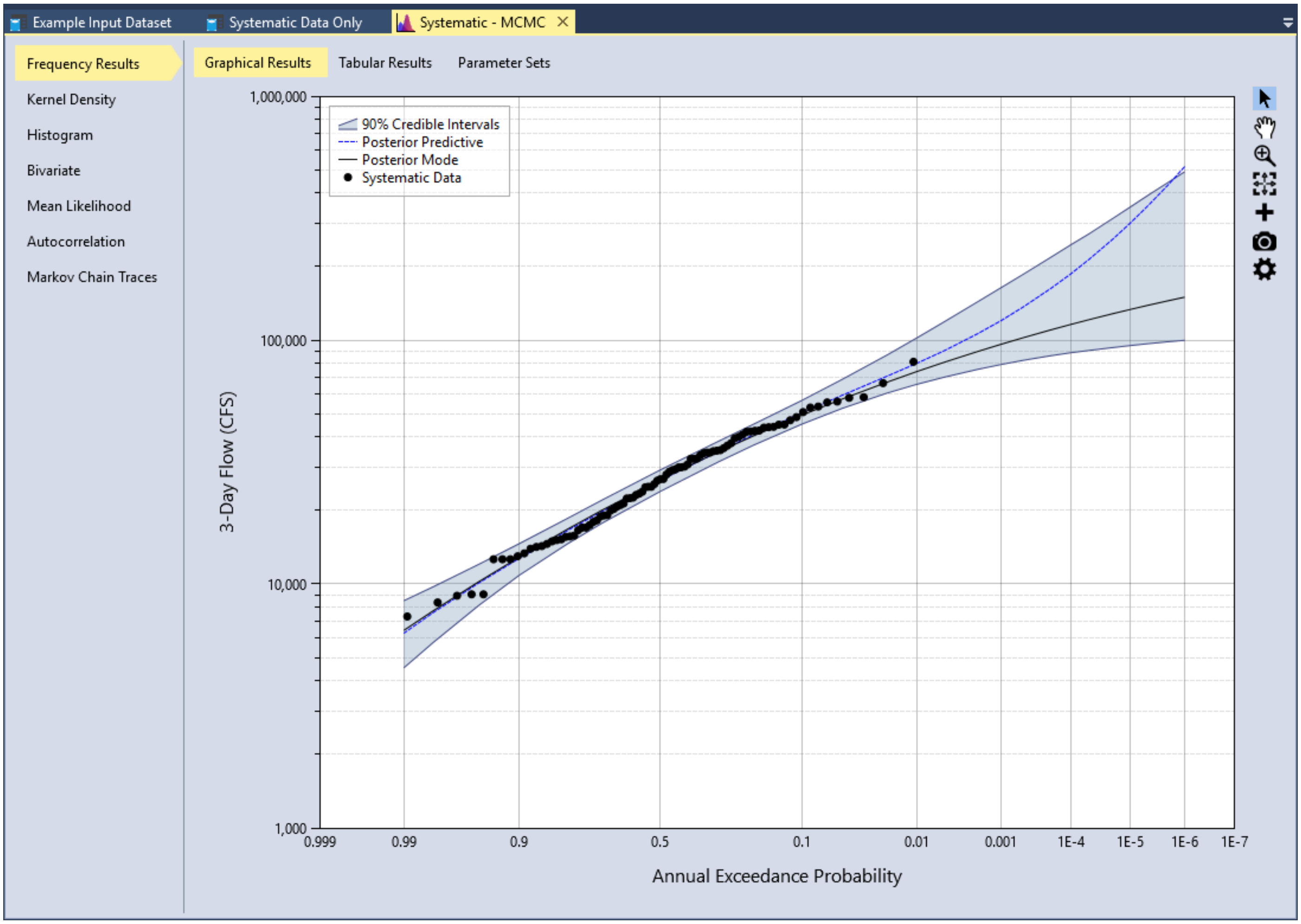


Figure 26 – Tabbed Document Group.

Properties Window

The **Properties Window**, which is typically on the right-hand side of the main window, displays the properties for the project and project elements. To access the element properties, first open the element for edit, then click on, or activate, the associated tabbed document. Some properties are common among all project elements, such as **Name** and **Description**, while others are unique to the specific elements. Properties are organized into groups for easier navigation. When you click on a property, the property description will be placed at the bottom of the properties window as shown in Figure 27. All analyses will be run from the properties window.

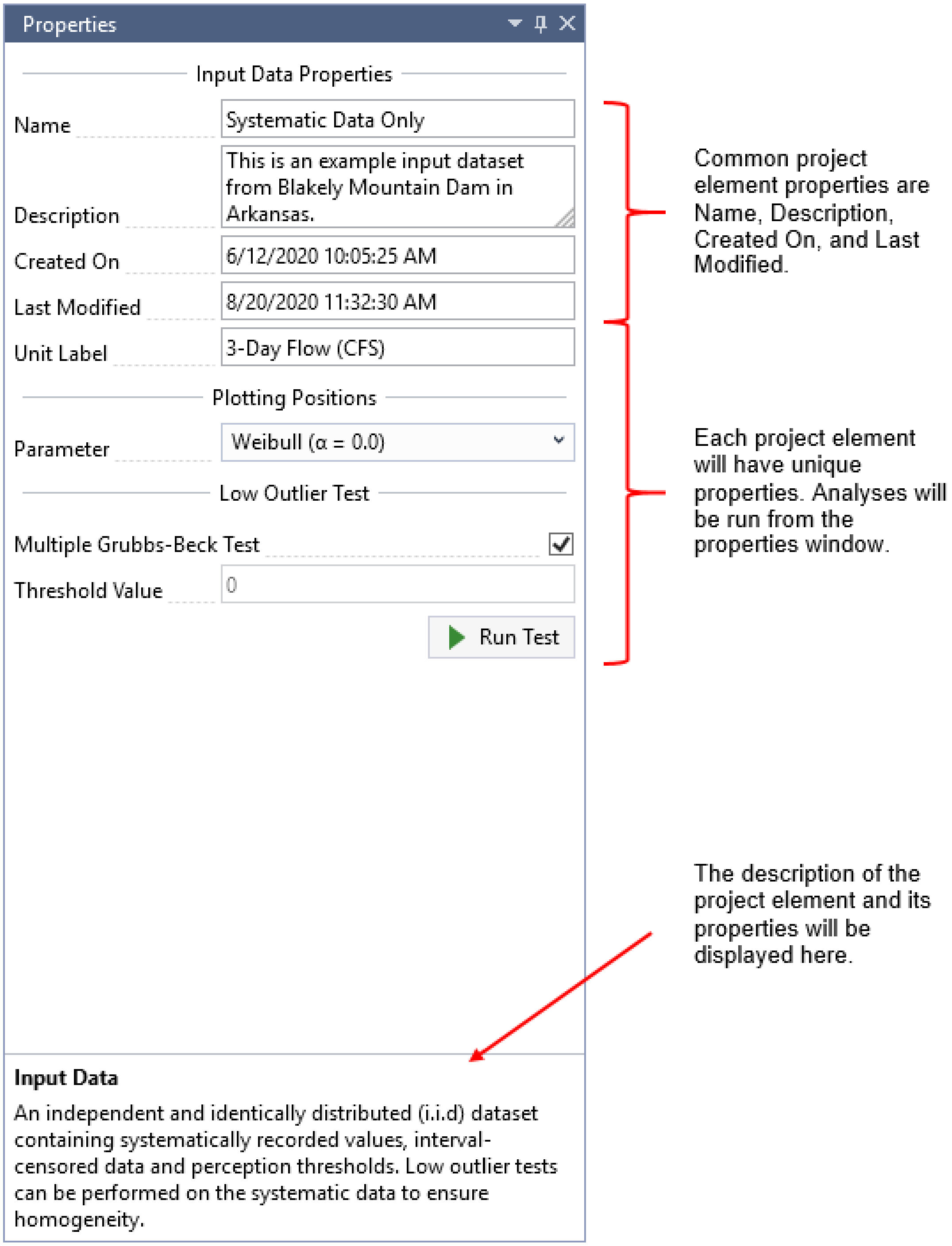


Figure 27 – Properties Window.

Message Window

The **Message Window** shows you errors, warnings, messages and event logs regarding the current state of your project. If there are any errors in your project file, they are listed here. The **Message Window** lets you perform the following tasks:

- Display the errors, warnings, messages, and events produced while you work in RMC-BestFit.
- Double-click any error message entry to open the project element where the problem occurs.
- Filter the type of entries that are displayed in the Message Window. The default is to only display errors and warnings.
- Export all entries to a text file.

Once you resolve an error, warning, or message, the entry will be removed from the Message Window. You may customize the font color of the various message types by navigating to **Tools** > **Options**. See the Personalize RMC-BestFit section for more details.

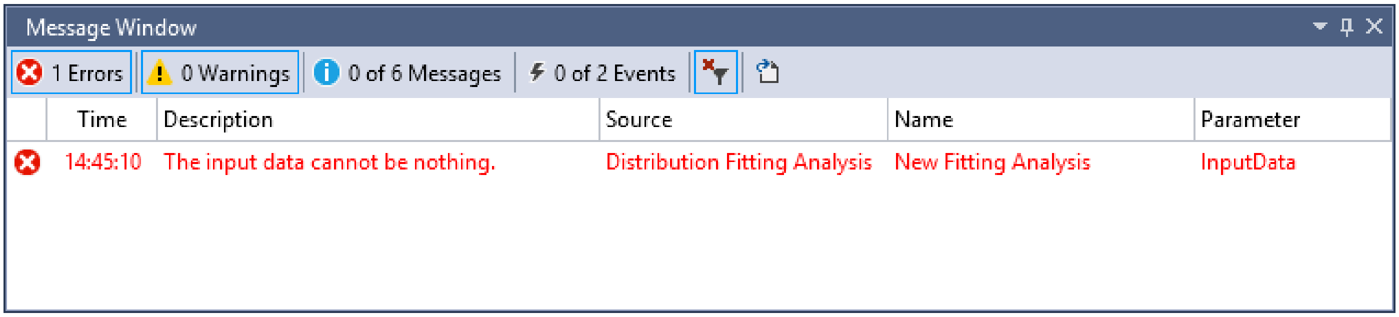


Figure 28 – Message Window.

Personalize RMC-BestFit

You can personalize RMC-BestFit in various ways to enhance the user experience by navigating to **Tools** > **Options**. The **Options** dialog allows the user to customize the Application, File Management, Message Window, and Default settings.

Application Options

You can set the application color theme to be the Light theme or Blue theme as shown in Figure 29 and Figure 30. The default is set to the Light theme. You can choose whether or not to **Save the window docking layout on close** so that RMC-BestFit will remember it when you reopen the application. You can also set the number of window items shown in the **Window** menu and the number of recent project items shown in the Recent Project list under the **File** menu.

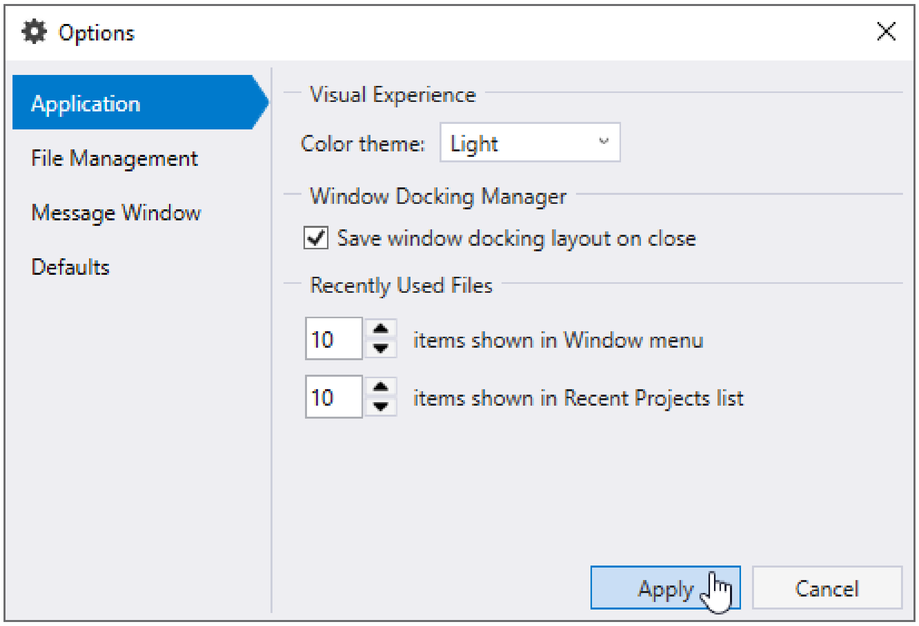


Figure 29 – Options Dialog with Light Color Theme.

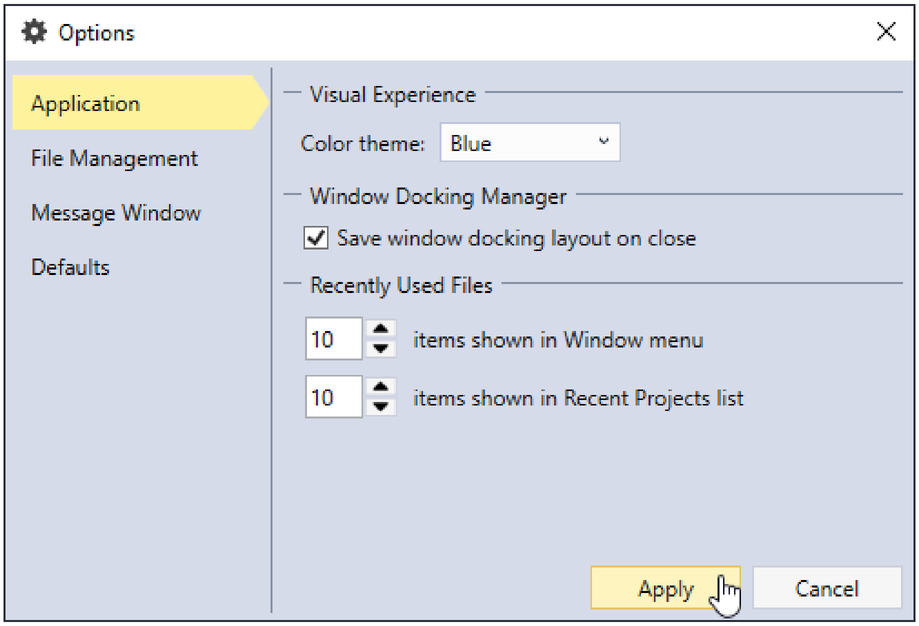


Figure 30 – Options Dialog with Blue Color Theme.

File Management Options

RMC-BestFit project files are saved as SQLite databases. SQLite is a self-contained, high-reliability, SQL database engine, which is also the most used database engine in the world (https://www.sqlite.org/). Database files can grow quickly as you use them, sometimes hindering performance. They can also occasionally become corrupt or damaged. You can use the **Tools** > **Compact & Optimize Project File** command to prevent or fix these problems. You can also set RMC-BestFit to **Compact & optimize project file on close**.

You can set the time interval in which a project backup file is created. A file with the extension .bak is automatically created when a project is opened. If the project closes successfully, then the .bak file is deleted. The database file can be damaged or corrupted, or the system could close unexpectedly causing you to lose important data, so on occasion, you might need to restore the project from a backup file by selecting **Tools** > **Restore from Backup**.

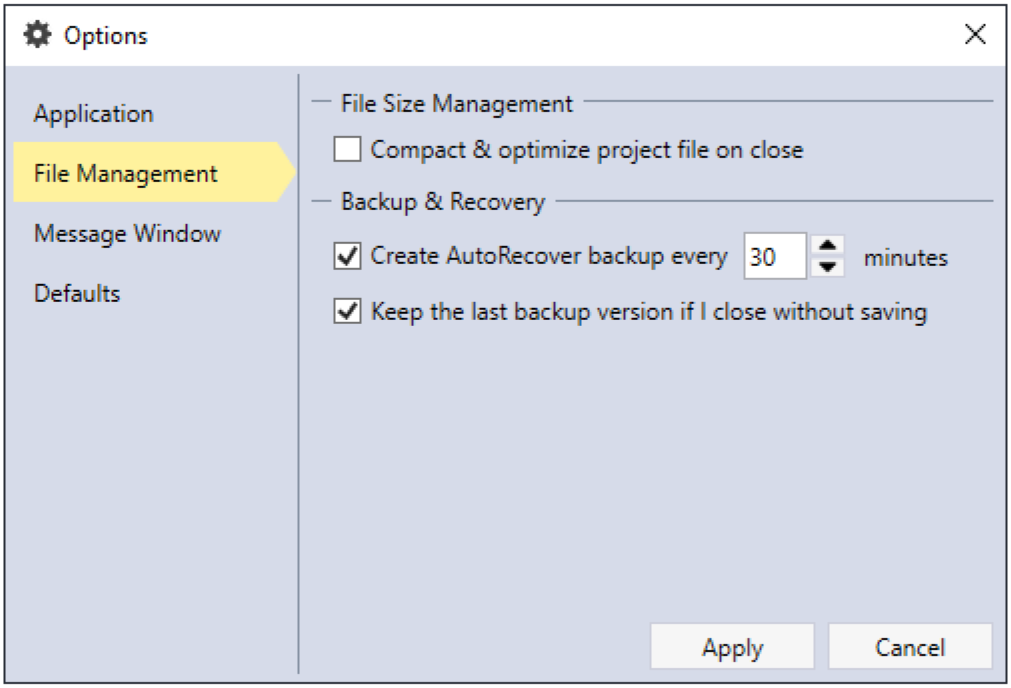


Figure 31 – File Management Options.

Message Window Options

You can adjust the auditory and visual setting for the **Message Window**. You can choose to turn on or off the beep sound effect and select the font color for the different message types, as shown below in Figure 32.

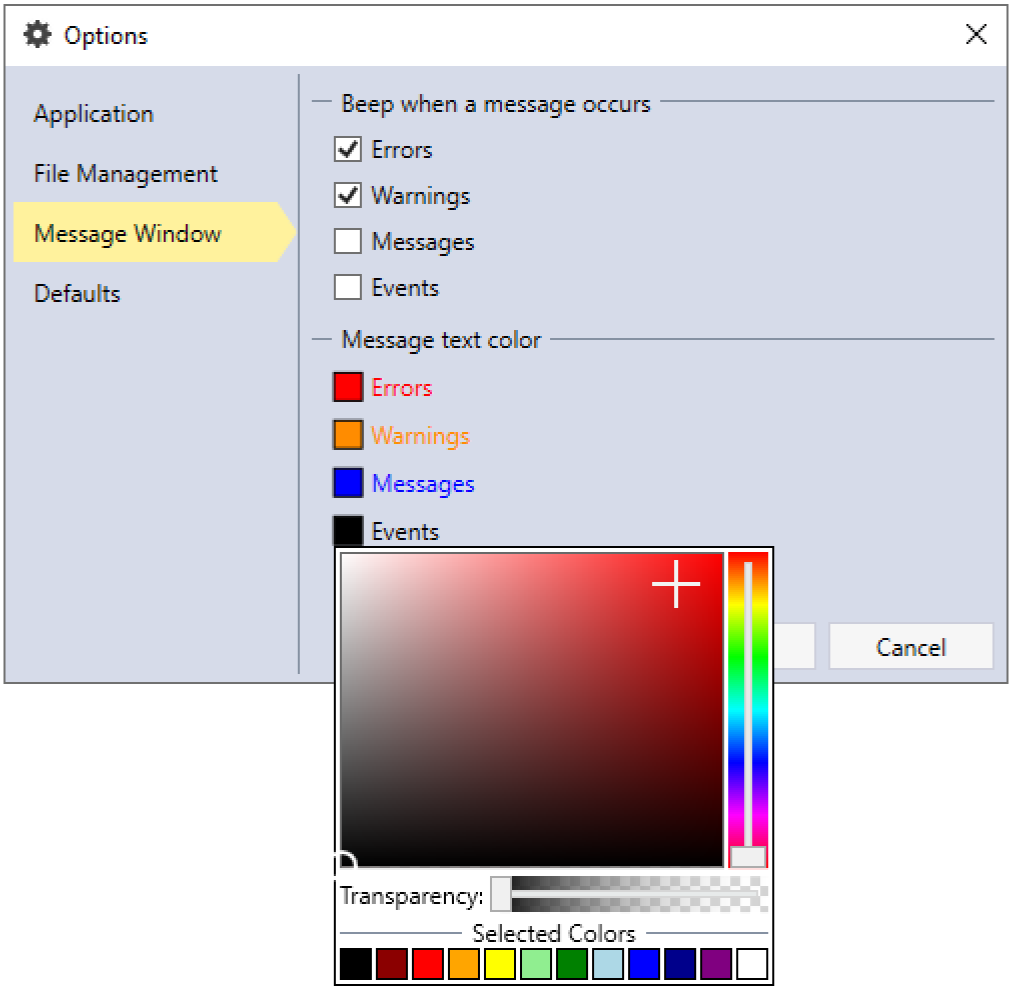


Figure 32 – Message Window Options.

Default Options

You can set the default project location directory, and the default decimal digits for the project inputs and outputs.

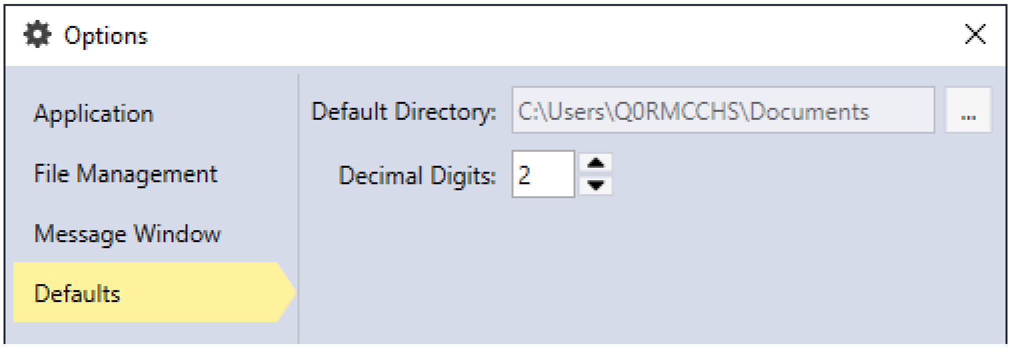


Figure 33 – Default Options.

Plot Features

RMC-BestFit has plot features which allow you to examine the plot data and customize the plots for reporting purposes. There is a tool bar located to the right of each plot, which allows you to interact with the plot in the following ways:

- Track data to get the X-Y point values.
- Pan the plot up and down or side to side.
- Zoom in to the plot data.
- Zoom out to the plot extents.
- Add annotations to the plot.
- Save the plot image as a PNG, PDF, or SVG file.
- Export plot series data to a Comma Separate (.csv), Excel (.xlsx), or SQLite (.sqlite) file.
- Change the plot properties to create a customized appearance. The changes to the plot properties are saved and persist when you open and close the project and project element.

Figure 34 – RMC-RFA Plot Features.

Track Data

The **Track Data** mouse pointer disables all other plot features when selected. Click on any plotted point or line in the graph to show a tooltip displaying the plot series name and X and Y data points.

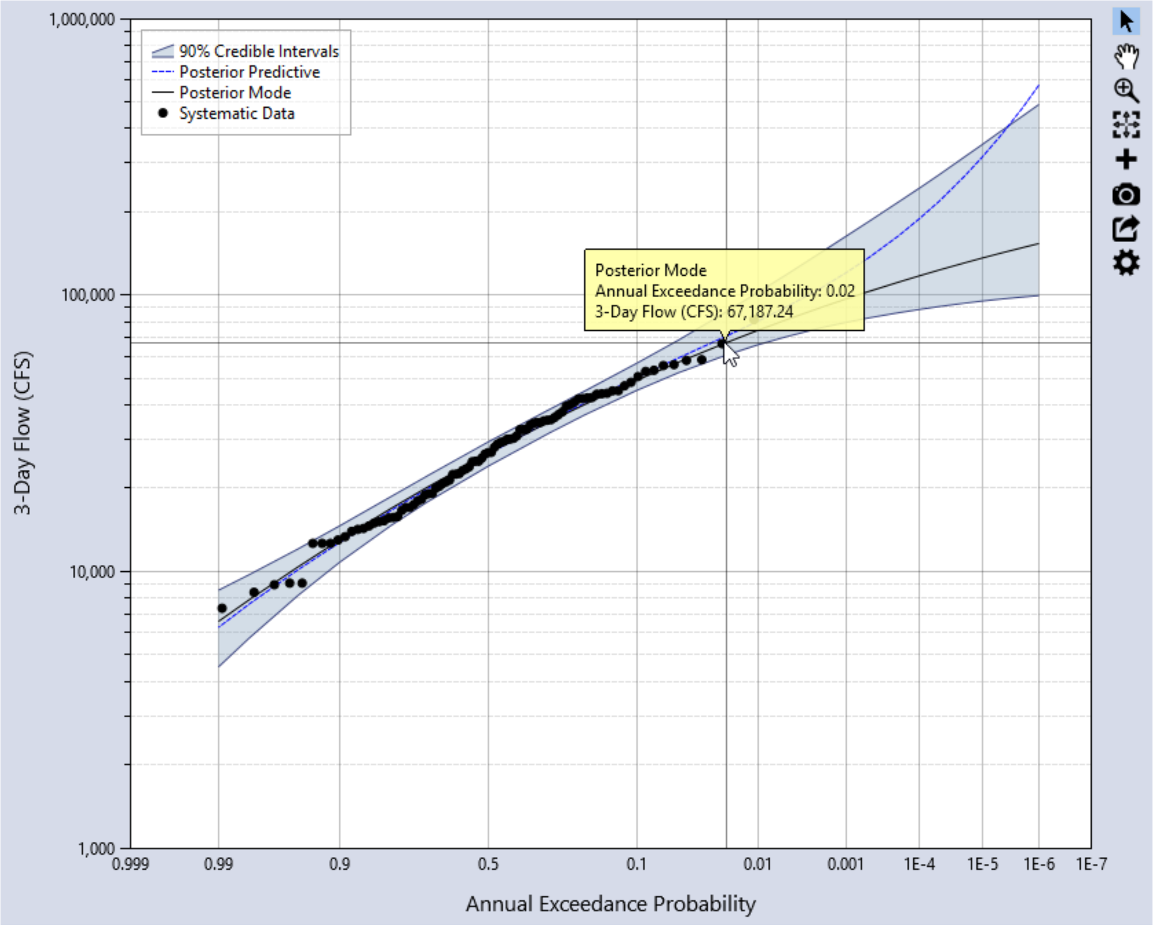


Figure 35 – Track Data on Plot.

Pan

The **Pan** feature allows you to navigate through the plot by clicking and dragging on the graph. The plot can be panned in any direction, up, down, and side to side.

Zoom

You can **Zoom In** on the plot to view data more precisely at smaller intervals. Zooming works by highlighting the area of interest with the magnifying glass cursor (see Figure 36), or by zooming in and out using the mouse wheel. When you zoom in on the plot you may also **Pan** the plot area. To zoom all the way back out, click the **Zoom to Extents** button on the plot tool bar.

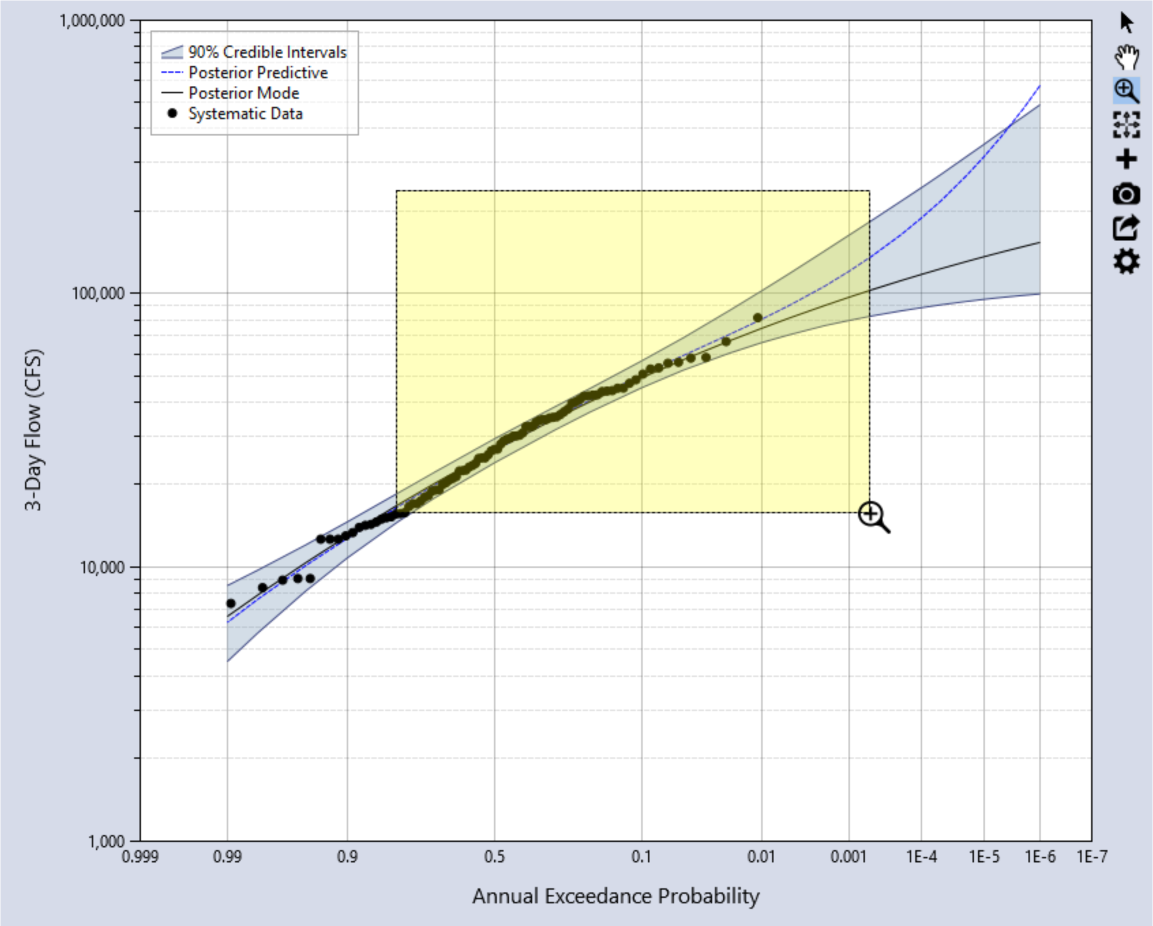


Figure 36 – Zoom In on Plot.

Add Annotation

Annotations are elements of the plot that show information that is not part of a plot series. Annotations are not included in the legend and not used by the tracker. You may add the following annotation types to the plot:

- Arrow
- Text
- Vertical Line
- Horizontal Line
- Rectangle
- Ellipse
- Point
- Polygon
- Polyline

The annotations can be moved by clicking and dragging them to a new spot. An example of adding an arrow annotation is shown in Figure 37 and Figure 38.

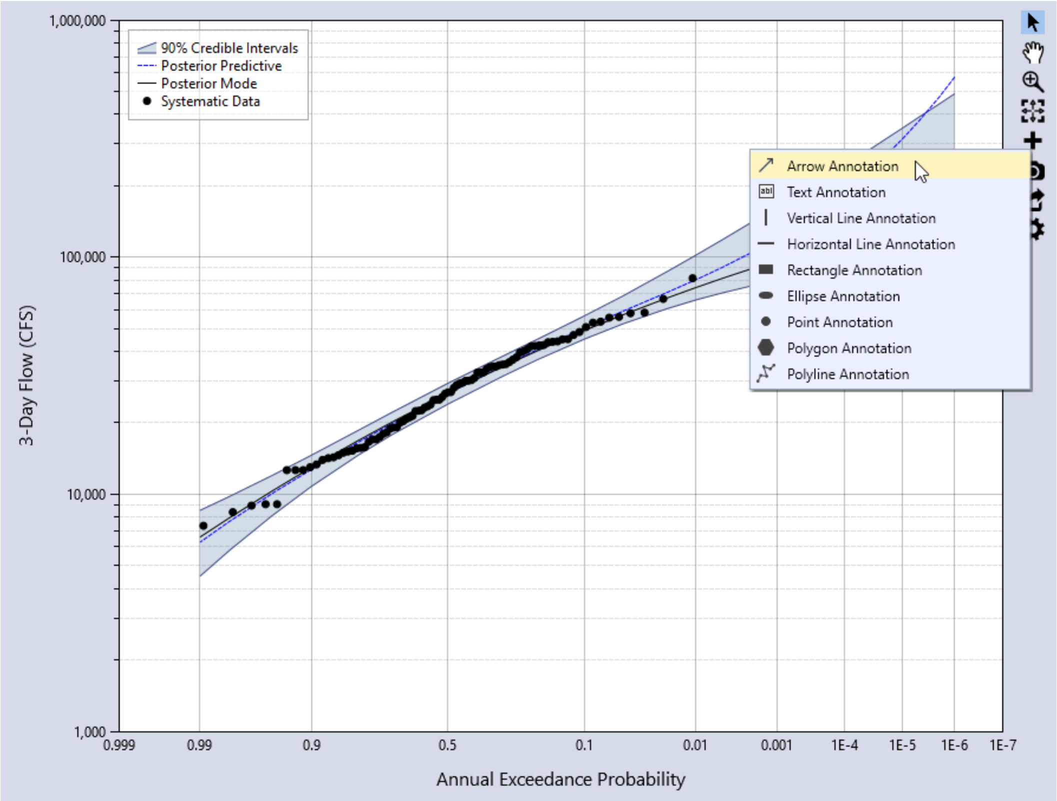


Figure 37 – Plot Annotations.

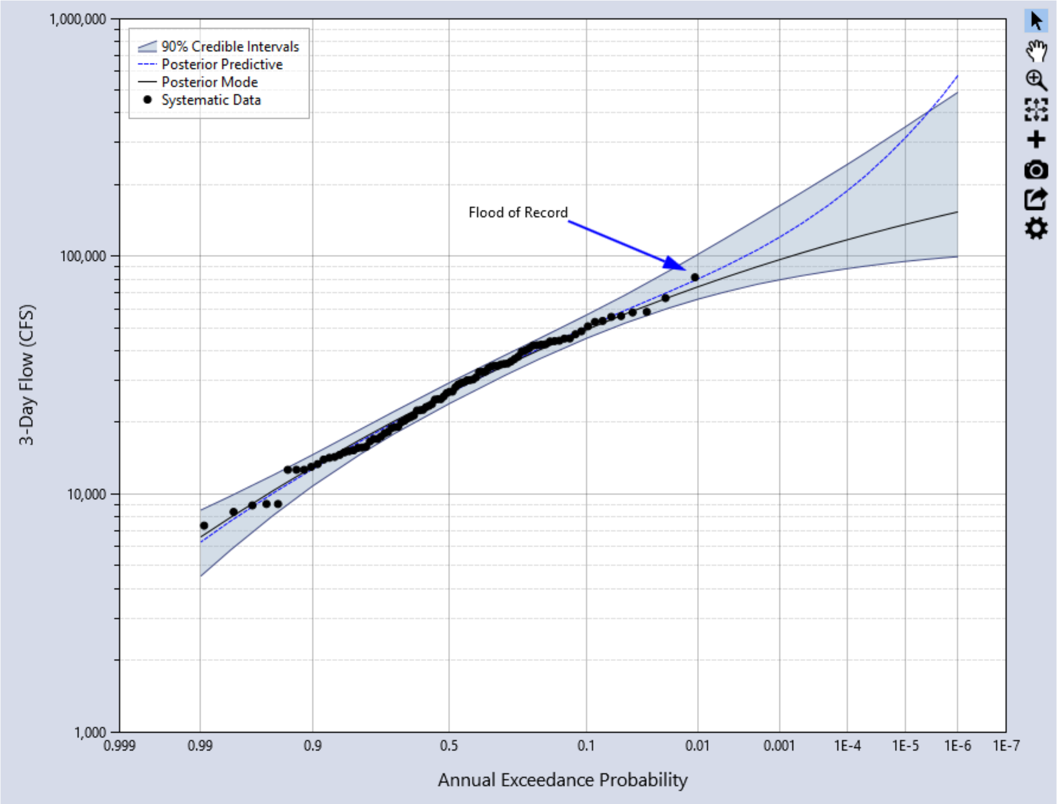


Figure 38 – Example of Arrow Annotation.

Export Series Data

The plot series data can be exported to a Comma Separated (.csv), Excel (.xlsx), or SQLite (.sqlite) file. After clicking the Export Series Data button, a Save As dialog will open. Select the desired file type and give the file a name, then click save. If the data is exported to a Comma Separated file, each plot series will be stored in separate columns. Whereas, if the data is exported to Excel, the series data will be stored on separate worksheet tabs. Likewise, if the data is exported to SQLite, the series data will be stored in separate tables.

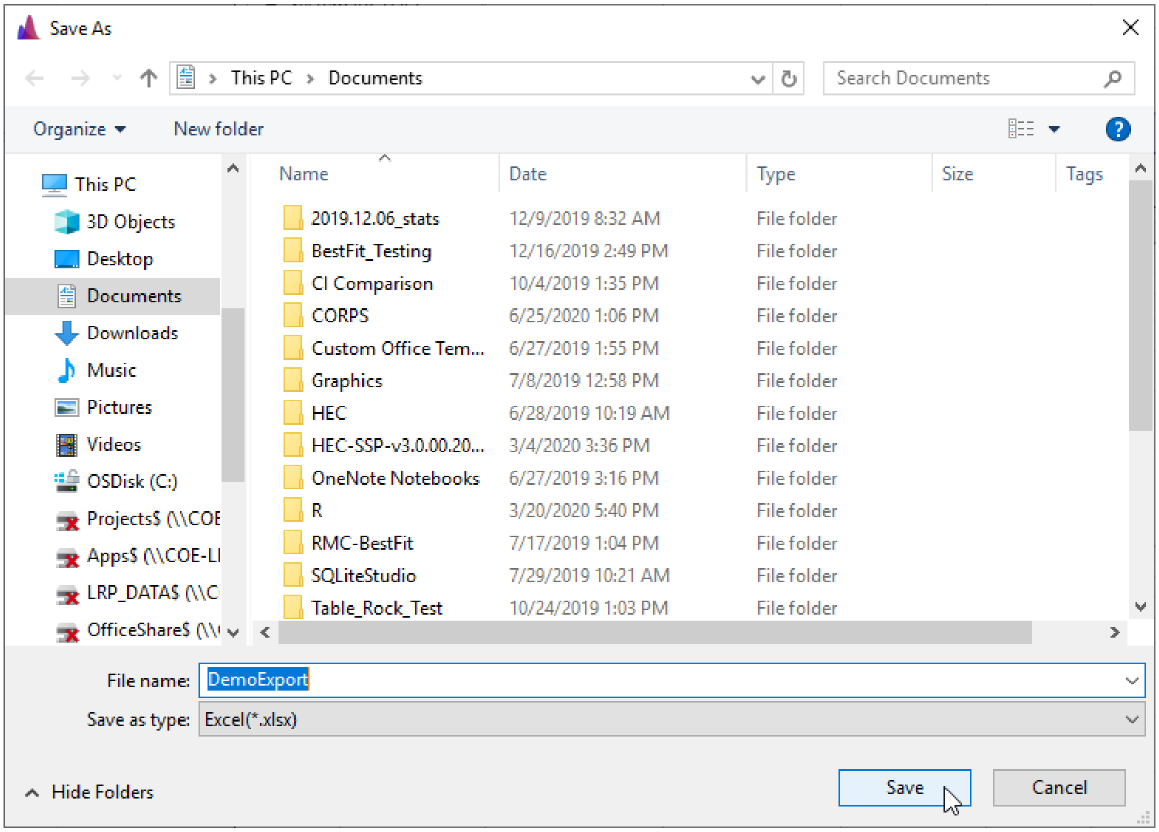


Figure 39 – Save As Dialog for Exporting Series Data.

Save Plot Image

The plot image can be saved as a PNG, PDF, or SVG file. Before saving, you can set the width and height of the image. After the **Save As** dialog opens, give the plot image a name and click save.

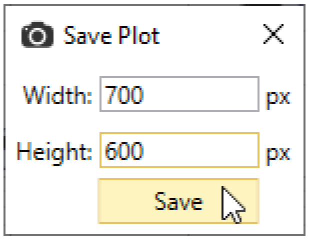


Figure 40 – Set the Width and Height of the Plot Image.

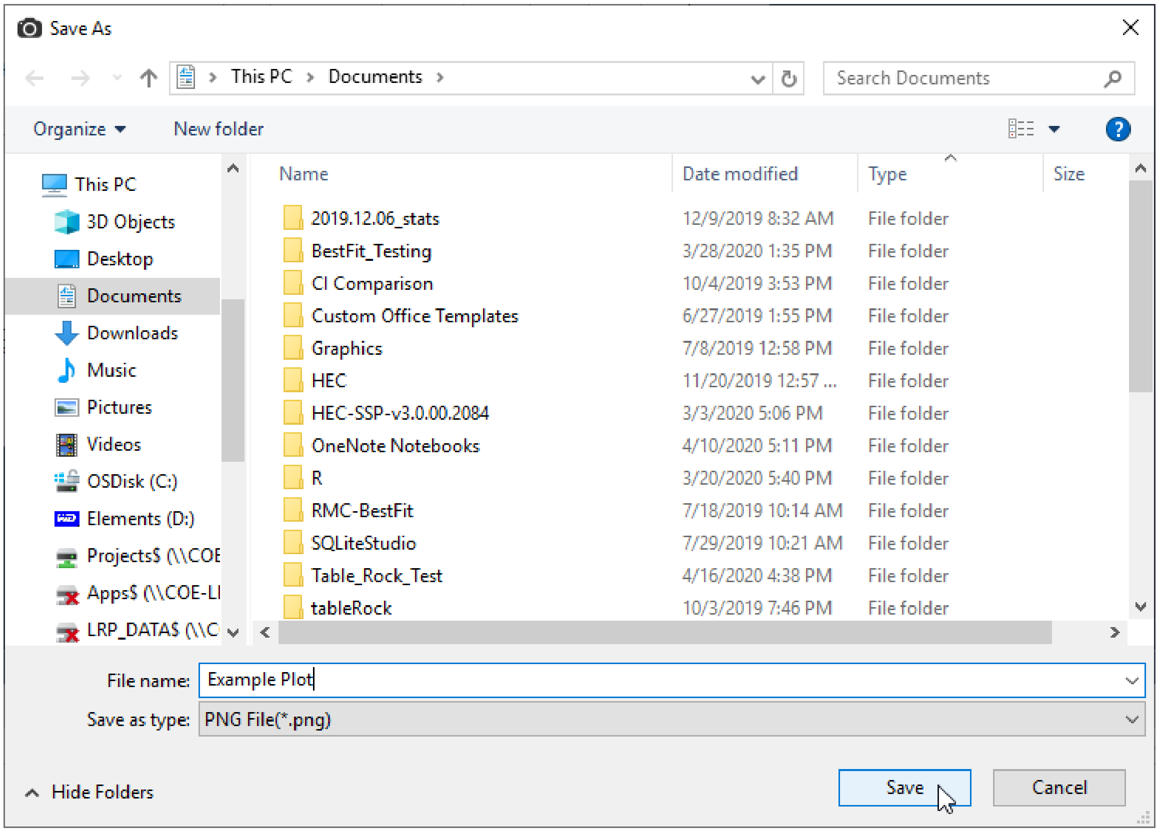


Figure 41 – Save Plot Dialog.

Plot Properties

You can edit the **Plot Properties** by clicking the gear wheel button in the plot tool bar. The plot properties will open in the **Properties Window**. From here, you can edit the general plot settings, legend, axes, series, and annotations. The plot properties will open to the **General Plot Settings** by default. Click the chevron in the upper right to open a drop down showing the other plot elements to edit.

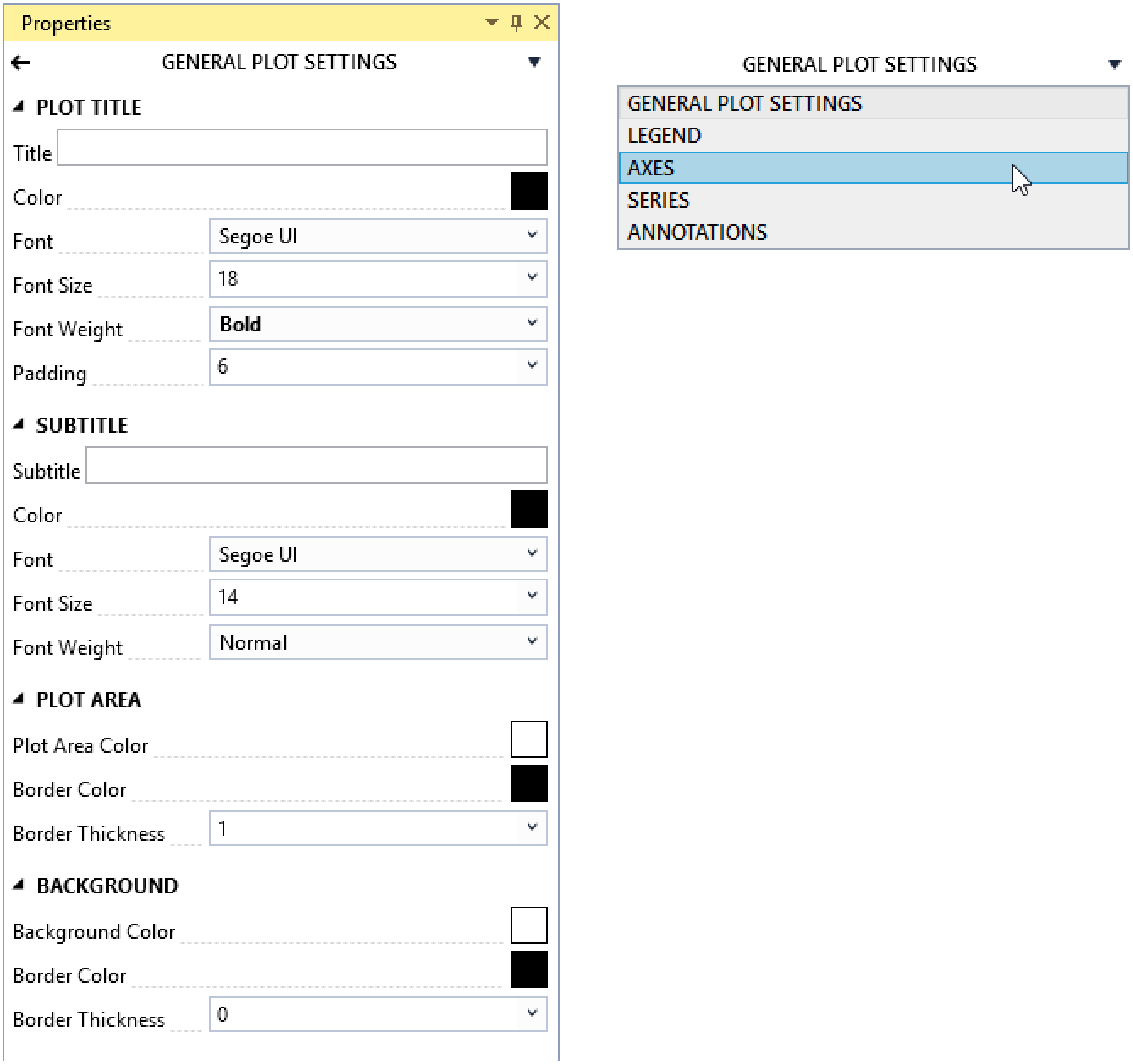


Figure 42 – Plot Properties.

Working with RMC-BestFit

In this quick start guide, we will demonstrate how to create a project, enter input data, select a model using the distribution fitting analysis, and perform a Bayesian estimation analysis.

Create a Project

To begin, let's create a new project. When you open RMC-BestFit, a **Blank Project** file is automatically created, as shown in Figure 43. The blank project is stored in your local temp directory. You may begin working with the blank project file immediately.

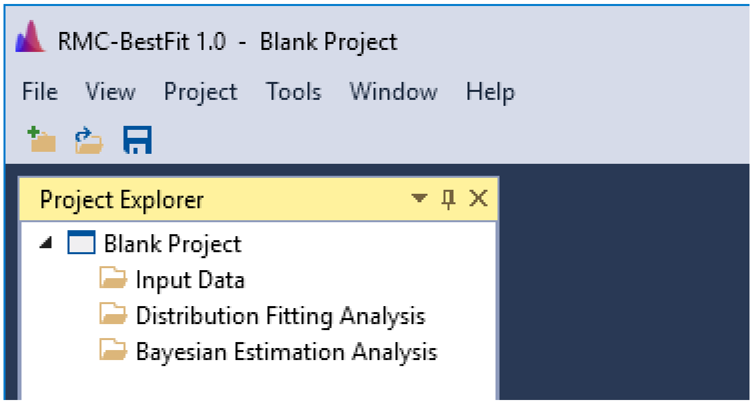


Figure 43 – RMC-BestFit Blank Project.

To save changes to the blank project, click the **Save** button on the tool bar or under the File menu. This will open the **Save Project As…** prompt. Enter the desired file name and click the save button in the bottom right. Now you are ready to continue working with RMC-BestFit.

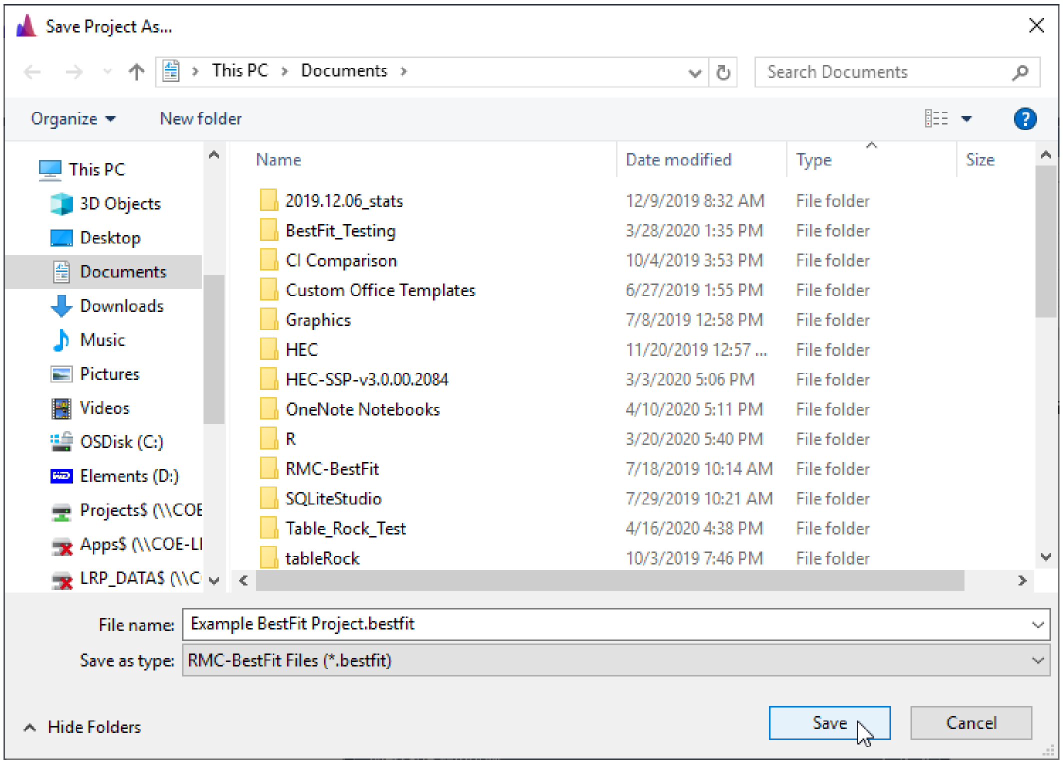


Figure 44 – Save Project As….

A new project can also be created by clicking **New Project…** under the File menu, or by clicking the New Project button located on the tool bar as shown in Figure 45 and Figure 46. If this is the first time you're using RMC-BestFit, your recent projects list will be empty.

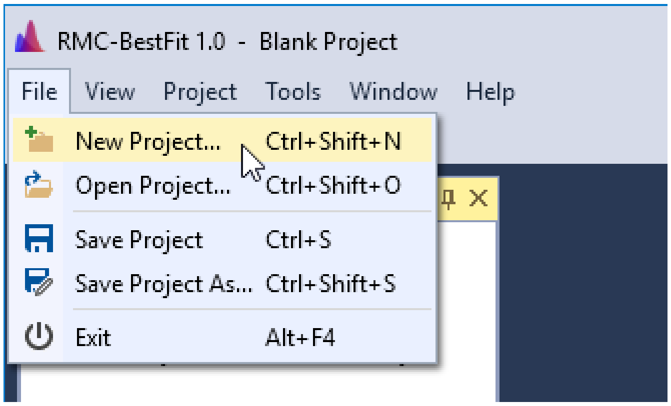


Figure 45 – Create New Project from the File Menu.

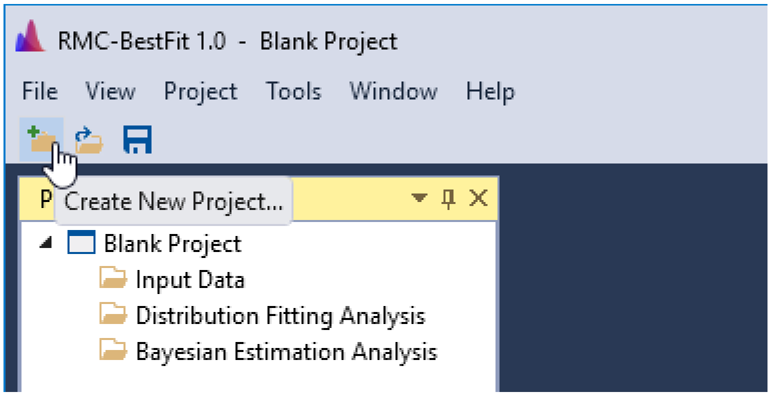


Figure 46 – Create New Project from the Tool Bar.

The project properties will be shown in the **Properties Window**, which is typically located on the right-hand side of the main window. You may edit the project name and description.

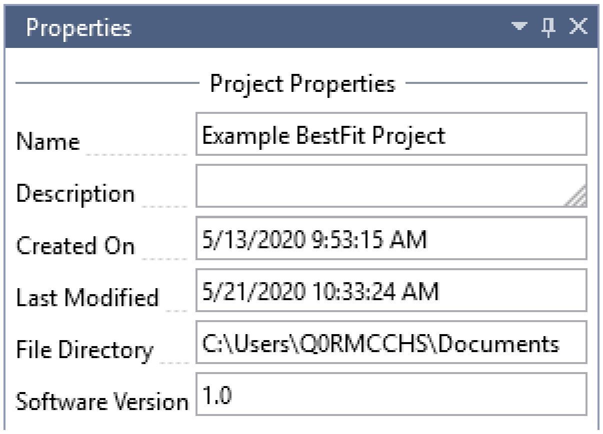


Figure 47 – Project Properties.

Input Data

In RMC-BestFit, the input data must be entered as block annual maxima, which are assumed to be independent and identically distributed (iid). RMC-BestFit supports three different data types:

1. Systematic Data: Data that are collected at regular, prescribed intervals under a defined protocol. In a maximum likelihood context, these values are treated as exact measurements. Low outlier tests can be performed on the systematic data to ensure homogeneity.

2. Interval Data: Data whose magnitudes are not known exactly, but are known to fall within a range or interval. In a maximum likelihood context, these values are treated as interval-censored.

3. Perception Thresholds: Data points that occurred during a period of years and have magnitudes that are below a threshold value, but unknown by how much. In a maximum likelihood context, these values are treated as left-censored.

The Distribution Fitting Analysis chapter provides greater detail on how these data types are treated in a likelihood context.

Create New Input Data

Let's begin by creating a new input dataset. Right-click on the **Input Data** folder header and click **Create New…** as shown in Figure 48. Next, give the **Input Data** a name and click **OK**.

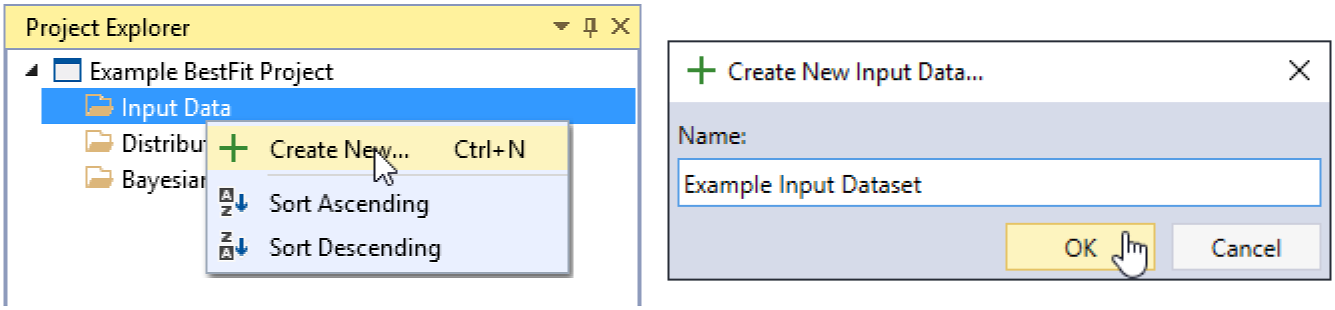


Figure 48 – Create New Input Data.

Once the new **Input Data** is created, it will be automatically opened into the **Tabbed Documents** area, and the input data properties will be displayed in the **Properties Window**. From here, you can set the **Description**, **Unit Label**, **Plotting Position Parameter**, and perform a **Low Outlier Test**.

For this example, we will be using 3-day inflow volumes for Blakely Mountain Dam, which is located near Hot Springs, Arkansas. This dataset includes systematic data, historical data dating back to 1870, and paleoflood information dating back 5,000 years. As shown in Figure 49, set the **Unit Label** to be "3-Day Flow (CFS)" and you will notice that this automatically updates the Y-axis labels on the **Chronology** and **Frequency** plots.

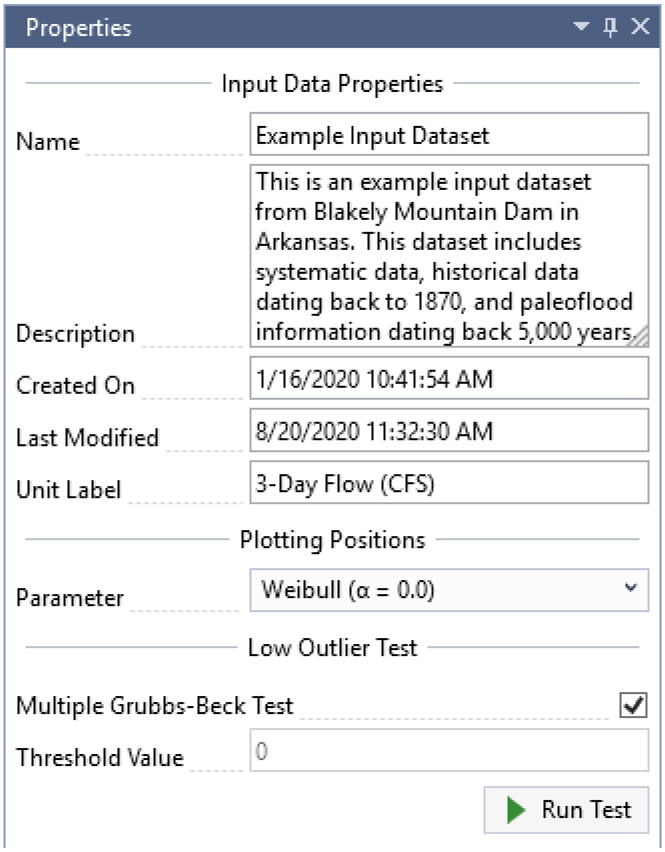


Figure 49 – Input Data Properties.

Systematic Data

Systematic data are collected at regular, prescribed intervals under a defined protocol. In a maximum likelihood context, these values are treated as exact measurements. The systematic dataset for Blakely Mountain Dam is provided in Table 1. Note that the systematic data does not need to be continuous; e.g., the Blakely Mountain dataset has missing data from 1931 to 1935. Gaps in data can be accounted for using thresholds, which will be demonstrated later in the Perception Thresholds section.

To enter data, first click the **Add Row(s)** button located on the left side of the table tool bar. This will add a blank row to the bottom of the table. Next, you can either manually enter your dataset, or copy and paste the dataset into the table as shown in Figure 51. Once you have entered all of the data, you will see that the plotting positions are automatically calculated and the data is plotted in the **Chronology** and **Frequency** plot as illustrated in Figure 52.

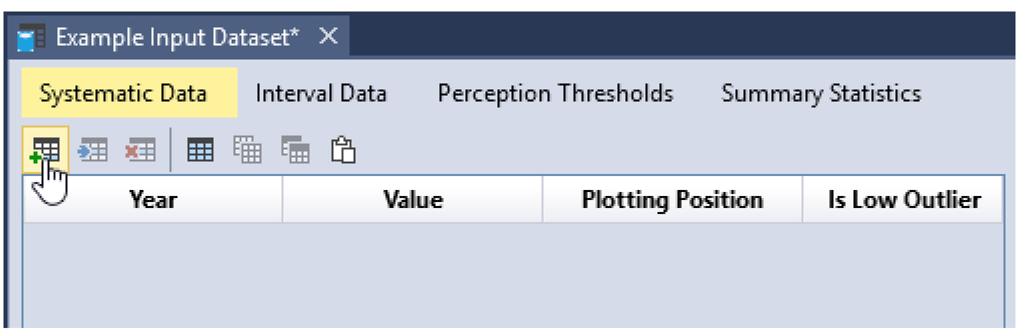


Figure 50 – Add Systematic Data.

Table 1 – Systematic Dataset of 3-Day Inflows at Blakely Mountain Dam near Hot Springs, Arkansas.

Year	Flow (CFS)	Year	Flow (CFS)	Year	Flow (CFS)
1923	57,985	1959	17,032	1990	30,122
1924	14,607	1960	45,014	1991	32,586
1925	8,403	1961	12,637	1992	34,357
1926	23,479	1962	17,037	1993	32,586
1927	66,629	1963	9,074	1994	45,065
1928	30,925	1964	36,066	1995	29,547
1929	25,046	1965	15,261	1996	14,310
1930	37,772	1966	23,886	1997	44,107
1936	9,074	1967	22,432	1998	15,676
1937	21,382	1968	55,536	1999	18,922
1938	43,769	1969	43,938	2000	17,981
1939	50,678	1970	22,490	2001	40,097
1940	7,360	1971	26,955	2002	42,471
1941	15,617	1972	53,417	2003	17,451
1942	23,189	1973	36,920	2004	14,964
1943	12,637	1974	42,584	2005	20,798
1944	30,064	1975	39,587	2006	19,157
1945	52,914	1976	12,996	2007	33,729
1946	22,607	1977	30,294	2008	58,319
1947	14,191	1978	8,952	2009	35,041
1948	19,098	1979	28,109	2010	32,700
1949	35,383	1980	12,637	2011	35,212
1950	25,046	1981	15,142	2012	34,585
1951	13,355	1982	16,624	2013	46,921
1952	20,330	1983	81,464	2014	15,795
1953	25,701	1984	13,952	2015	42,189
1954	26,897	1985	40,946	2016	55,982
1955	20,019	1986	29,317	2017	34,585
1956	18,240	1987	24,930	2018	48,324
1957	21,084	1988	42,076		
1958	28,886	1989	26,551		

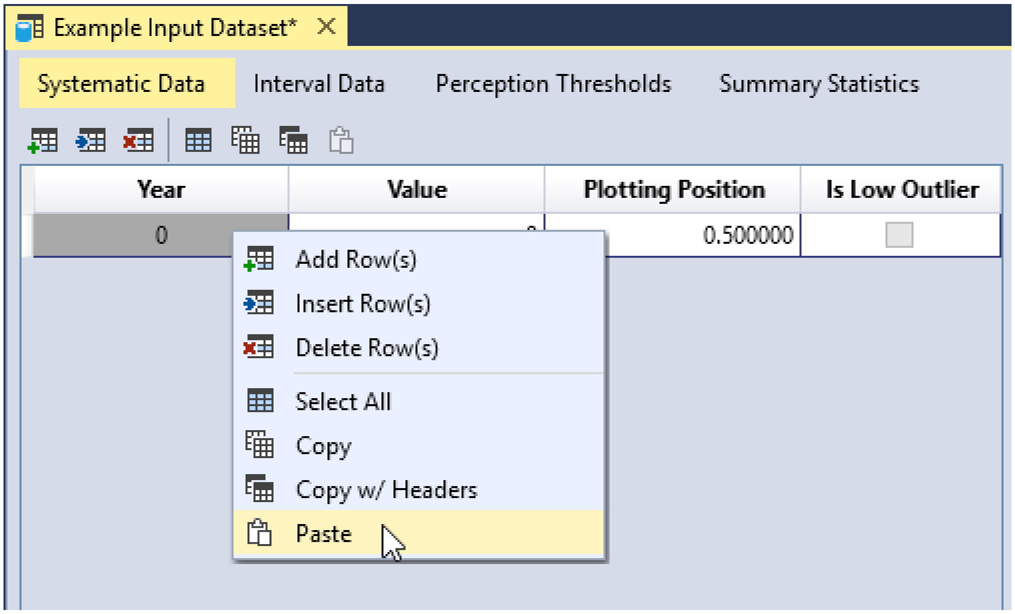


Figure 51 – Paste Data into Table.

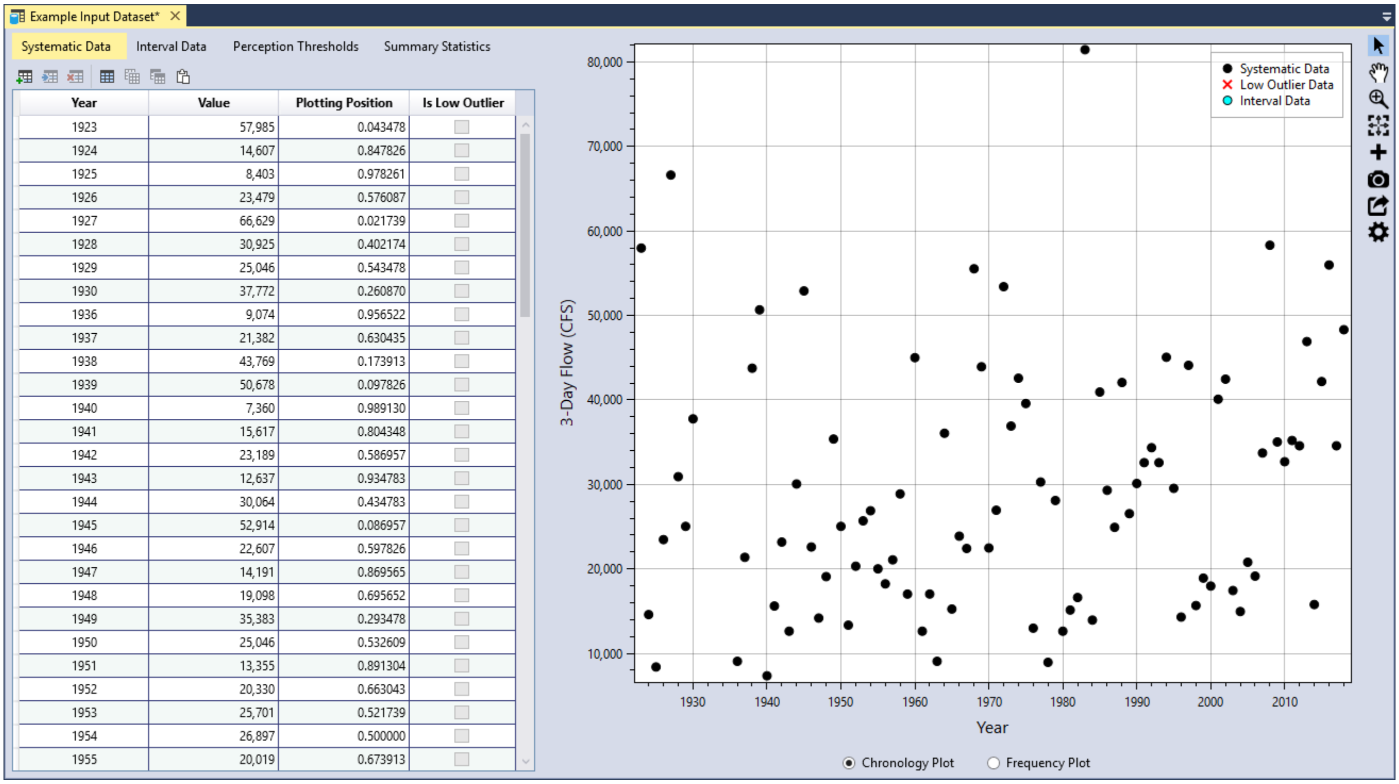


Figure 52 – Systematic Data.

You can view the data as a chronology plot or as a frequency plot by toggling the radio buttons located underneath the plot as shown in Figure 53. Clicking the **Chronology Plot** radio button will display the data in chronological order, with the years on the X-axis. The **Frequency Plot** radio button will display the data as a nonparametric frequency plot based on the Hirsch-Stedinger plotting positions.

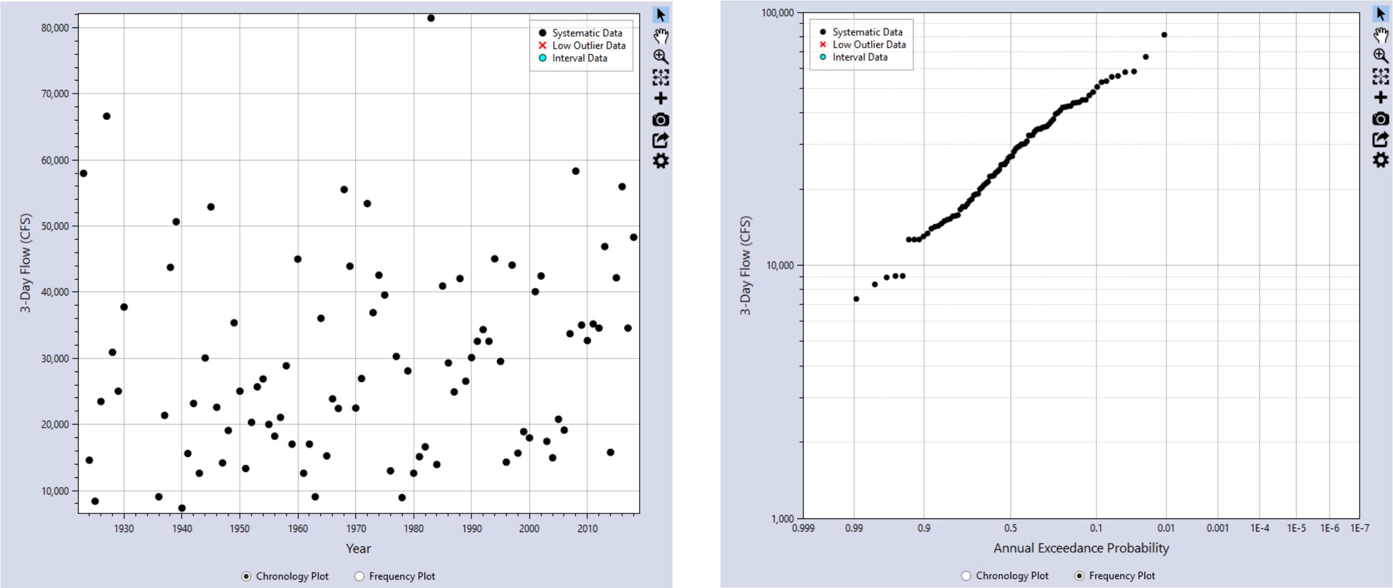


Figure 53 – View Chronology or Frequency Plot.

Data Table Features
You can edit the data table by using the tool bar located above the table, or by right-clicking within the table. You can interact with the table in the following ways:

- Add rows(s) to the bottom of the table.
- Insert row(s) into the table.
- Delete row(s) from the table.
- Select all table cells.
- Copy the selected cells.
- Copy the selected cells with the table headers.
- Paste from the clipboard into the table.
- Sort a column in ascending or descending order.
- Clear all table sorting.

You can sort a column by right-clicking on the column header as shown below.

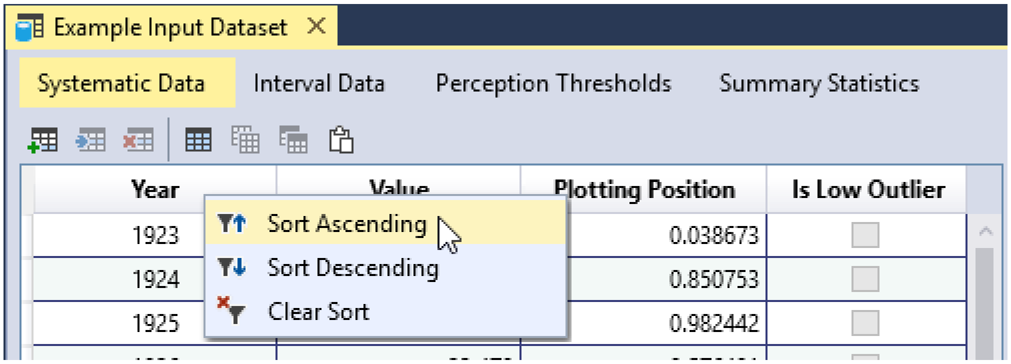


Figure 54 – Sort Input Data Tables.

Data Validation
The input data tables have built-in validation. The **Systematic Data** have the following requirements:

- The year must be unique.
- The year must be between -100,000 and +100,000.
- The value must be non-negative, or greater than or equal to zero.

When you enter invalid data, the table cell will turn red, and provide a tooltip indicating the source of the error. In addition, an error message will appear in the **Message Window** indicating that you must resolve all errors in the data table.

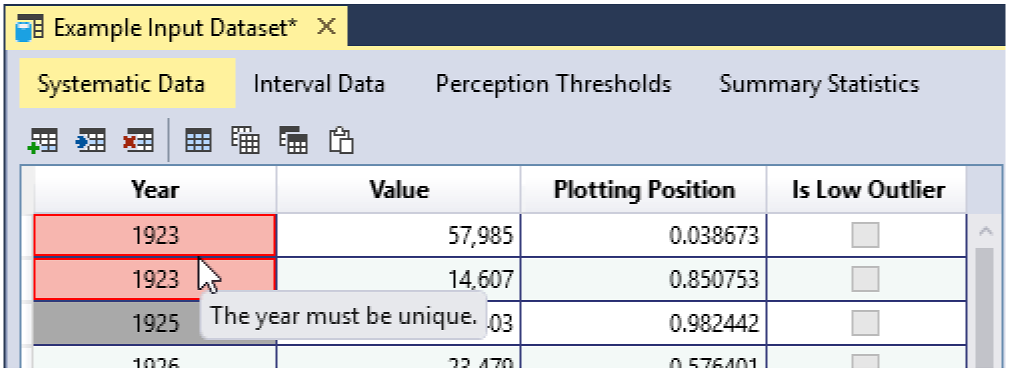


Figure 55 – Input Data Validation.

Plotting Positions

The input data can be plotted as a **Chronology Plot** or as a nonparametric **Frequency Plot** as shown in Figure 53. A *frequency plot* or *probability plot* is a plot of magnitude versus a probability. The probability assigned to each data point is commonly determined using a *plotting position* formula. Plotting positions are a method for creating an *empirical frequency*. The formula computes the exceedance probability of a data point based on the rank of the data point in a sample of a given size. The plotting positions typically have significant uncertainty due to sampling error resulting from small sample sizes.

A rank-order method is used to plot the annual maxima data. This involves ordering the data from the largest event to the smallest event, assigning a rank of 1 to the largest event and a rank of *n* to the smallest event, and using the rank (*i*) of the event to obtain a probability plotting position. Many plotting position formulae are special cases of the general formula:

$$P_i = \frac{i - \alpha}{n + 1 - 2\alpha}$$ *Equation 1*

where *i* is the rank of the event, n is the sample size, α is a constant greater than or equal to 0 and less than 1, and P_i is the exceedance probability for an event with rank *i*. The value of α determines how well the calculated plotting positions will fit a given theoretical probability distribution.

RMC-BestFit uses the Hirsch-Stedinger (H-S) plotting position formula (Hirsch & Stedinger, 1987) (U.S. Geological Survey, 2018), which is an extension of the general formula above that is also capable of incorporating threshold-censored data. The H-S plotting positions are used to visually and quantitatively assess the goodness-of-fit of the fitted distributions (see the Goodness-of-Fit Measures section for more detail).

You can choose from the following α parameter options:

- Weibull ($\alpha = 0.0$): Recommended as the default value because it is unbiased for all distributions.
- Median ($\alpha = 0.3175$): Provides median exceedance probabilities for all distributions.
- Blom ($\alpha = 0.375$): Recommended for Normal, Gamma, 2-parameter Log-Normal, 3-parameter Log-Normal, and Log-Pearson Type III distributions.
- Cunnane ($\alpha = 0.40$): Recommended for Generalized Extreme Value and Log-Gumbel distributions, approximately quantile unbiased.
- Gringorten ($\alpha = 0.44$): Recommended for Exponential, Gumbel and Weibull distributions.
- Hazen ($\alpha = 0.50$): Recommended when the parameters of the parent distribution are unknown.

As you can see, each plotting position parameter has a different motivation. Some attempt to achieve unbiasedness in quantile estimates across multiple distributions, while other formulas are optimized for use with a particular theoretical probability distribution. Choosing a plotting position parameter is similar to choosing a probability distribution to represent a particular set of data. It is often better to select a plotting position parameter that is flexible and makes the fewest assumptions. For this reason, the Weibull parameter ($\alpha = 0.0$) is set as the default value in RMC-BestFit, which is consistent with current practice.

When you hover over the plotting position drop-down, you will see a tooltip providing the recommended use for the given parameter as shown below.

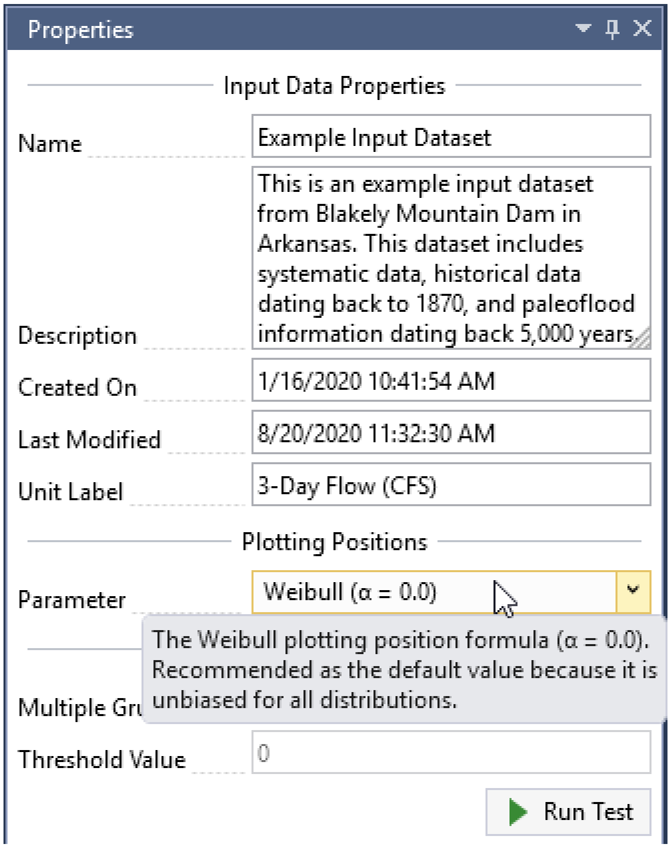


Figure 56 – Set Plotting Position Parameter.

Low Outlier Test

For the distribution fitting or Bayesian estimation analyses to be theoretically valid, the input data must be independent and identically distributed. As a means to ensure homogeneity, RMC-BestFit provides the Multiple Grubbs-Beck test (MGBT) (Cohn, et al., 2013) for low outliers, which is consistent with the Bulletin 17C guidelines (U.S. Geological Survey, 2018). In RMC-BestFit, the MGBT is only applied to systematic data, which are considered exact measurements. Interval- and threshold-censored data are not included in the test.

To run the MGBT, make sure the **Multiple Grubbs-Beck Test** checkbox is checked, and click the **Run Test** command button. When the test is complete, a message box will appear that reports how many low outliers were identified. In addition, the MGBT Critical Value will be displayed in the **Threshold Value** textbox. There should be zero low outliers identified for this dataset, so the threshold value should also be set to zero as shown in Figure 57.

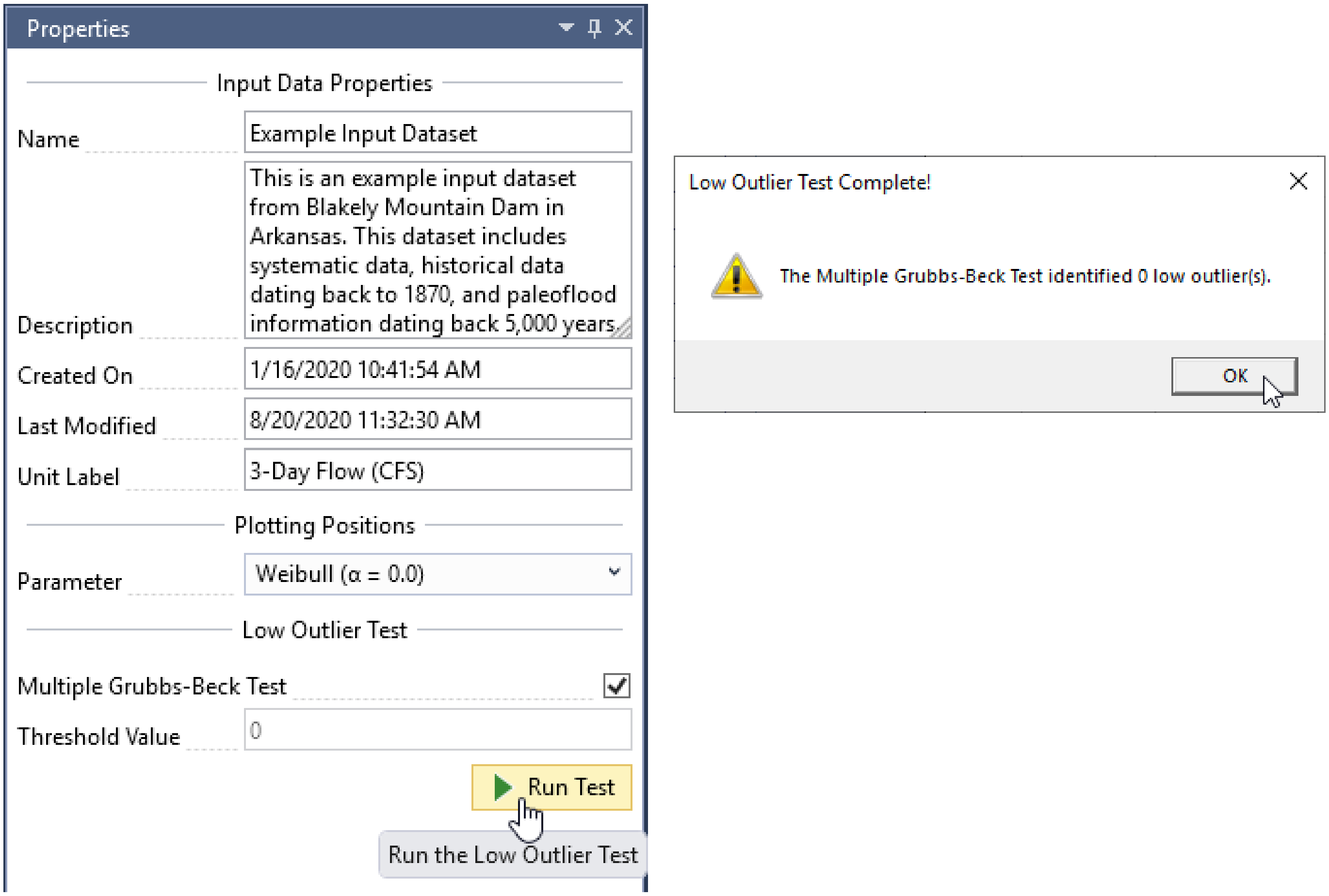


Figure 57 – Run Multiple Grubbs-Beck Low Outlier Test.

If desired, you also have the option to enter a value for the low outlier threshold. When you run the test, any data value below this threshold will be identified as a low outlier. The **Threshold Value** cannot be set to a value that would censor more than 50 percent of the values within the data set. To use this option, uncheck the **Multiple Grubbs-Beck Test** checkbox and enter the preferred value.

Values that are identified as low outliers will be checked in the Is Low Outlier column of the systematic data table. The low outliers will be displayed as a red **X** in the chronology and frequency plots.

The low outlier threshold value identified by the MGBT or manual threshold value is automatically treated as a left-censored threshold in the distribution fitting and Bayesian estimation analyses. For example, if the low outlier threshold value is 8,000 and there are eight data points below the threshold identified as low outliers, then this is treated equivalent to a left-censored threshold with eight values below and zero above. However, RMC-BestFit does not include the low outlier threshold in the H-S plotting position routine. Conceptually, the low outlier test removes exact data points and replaces them with a threshold-censored value. This represents a loss in information. However, if this low outlier threshold is included in the H-S routine, then it will make the plotting positions rarer, signaling an increase in information. This is counterintuitive, and for this reason RMC-BestFit does not include the low outlier threshold in the H-S plotting position routine.

Interval data

Interval data have magnitudes that are not known exactly, but are known to fall within a range or interval. A paleoflood study was performed for the Blakely Mountain watershed, where two major historical floods were identified: one occurred in 1882 and another occurred sometime around the year 1020.

Table 2 – Interval Data for 3-Day Inflows at Blakely Mountain Dam near Hot Springs, Arkansas.

Year	Lower	Most Likely	Upper
1020	105,000	110,000	115,000
1882	66,000	76,000	86,000

You can add the interval data in the same manner as was done for the systematic data. First click the **Add Row(s)** button located on the left side of the table tool bar. This will add a blank row to the bottom of the interval data table. Next, you can either manually enter the data, or copy and paste the interval data into the table. Once you have entered the data, you will see that the plotting positions are automatically calculated and the intervals are plotted in the **Chronology** and **Frequency** plot as vertical bars (see Figure 58).

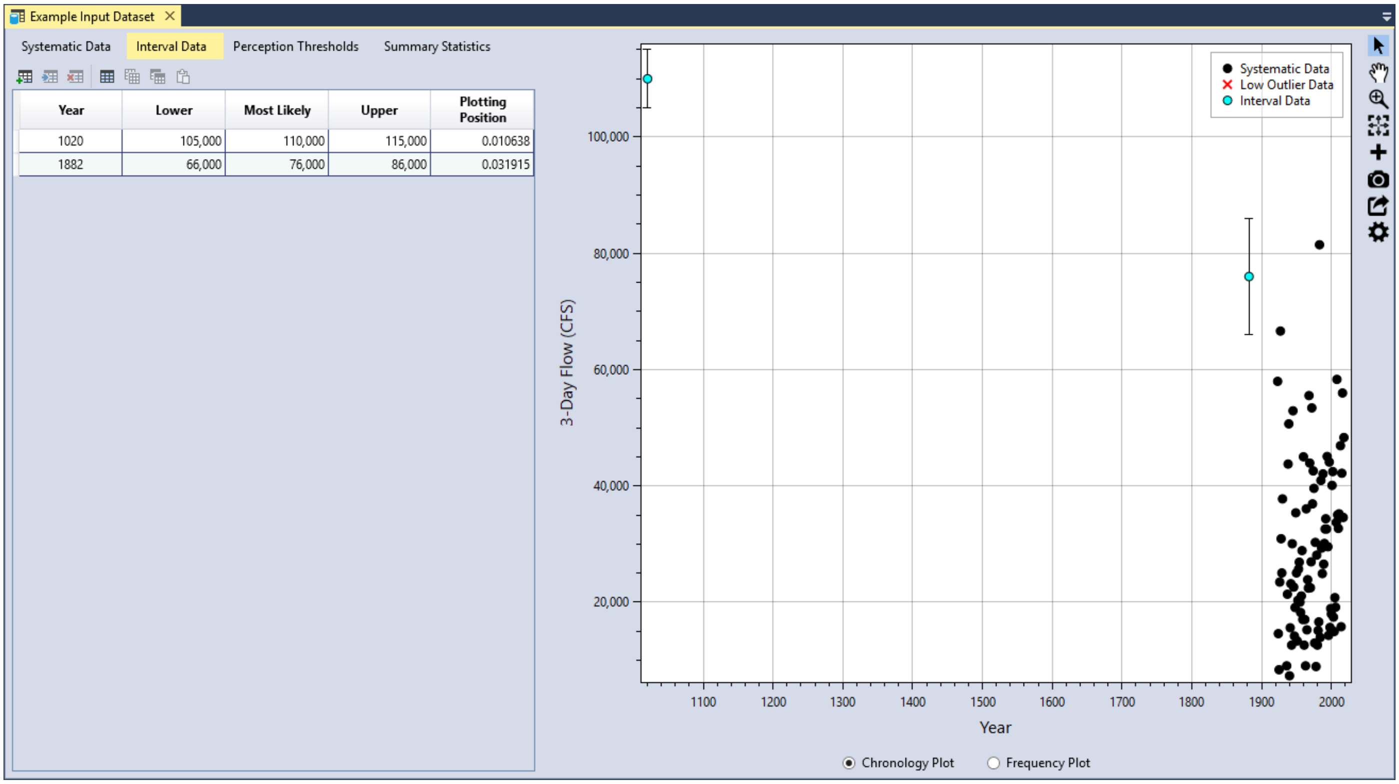


Figure 58 – Interval Data.

The **Interval Data** have the following requirements:

- The year must be unique.
- The year must be between -100,000 and +100,000.
- The year cannot overlap with any data point in the **Systematic Data** table.
- The lower, most likely, and upper values must be non-negative, or greater than or equal to zero.
- The lower must be less than the most likely value, and the upper must be greater than the most likely value.

Perception Thresholds

The term *perception threshold* originates from Bulletin 17C (U.S. Geological Survey, 2018). For the purposes of RMC-BestFit, a **Perception Threshold** defines a threshold level over a period of years. Data points that occurred during the threshold period have magnitudes that are below the threshold value, but it is unknown by how much. Conversely, we can also say that if an event occurred during the threshold period, and it had a magnitude larger than the threshold value, then we would have evidence it occurred; i.e., data points larger than the threshold would have been *perceived* by us.

There were four perception thresholds identified for the Blakely Mountain Dam 3-day inflow volume dataset (see Table 3). The paleoflood study determined that 3-day inflows have not exceeded 220,000 cfs in the last ~5,000 years. Flood volumes did not exceed 104,000 cfs during the years between the paleoflood in 1019 and 1869. From 1870 to the 1922, which is the beginning of the systematic dataset, 3-day flood volumes did not exceed 65,000 cfs, except for the large 1882 event. Finally, 3-day flood volumes during the missing years of systematic data from 1931 to 1935, also did not exceed 65,000 cfs.

Table 3 – Threshold Data for 3-Day Inflows at Blakely Mountain Dam near Hot Springs, Arkansas.

Start Year	End Year	Value
-2980	1018	220,000
1019	1869	104,000
1870	1922	65,000
1931	1935	65,000

Threshold data is entered in the same manner as was done for the systematic and interval data. Click the **Add Row(s)** button located on the left side of the table tool bar. Next, you can either manually enter the data, or copy and paste the threshold data into the table. Once you have entered the data, you will see that the plotting positions are automatically calculated and the thresholds are plotted in the **Chronology** plot as a shaded area (see Figure 59).

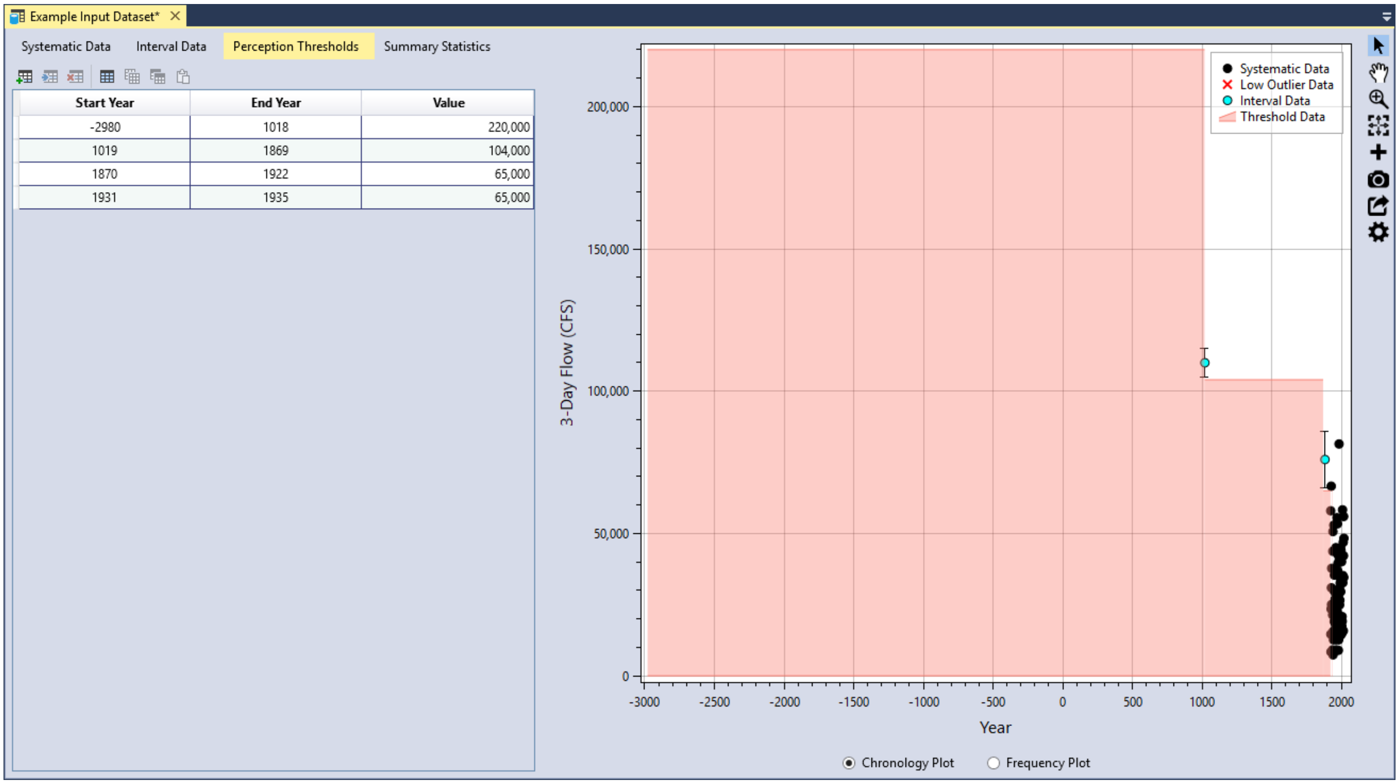


Figure 59 – Perception Thresholds.

The **Perception Threshold Data** have the following requirements:

- The start and end years must be between -100,000 and +100,000.
- The start year must be less than or equal to the end year.
- The start and end years of the thresholds must be entered in ascending order.
- The threshold periods cannot overlap with each other.
- The threshold values must be non-negative, or greater than or equal to zero.

Summary Statistics

RMC-BestFit provides summary statistics for the systematic data and for all of the data, including low outliers, intervals, and perception thresholds (see Figure 60). Summary statistics for the systematic data are based on the sample moments and percentile estimates, while summary statistics for all data are based on the nonparametric H-S plotting positions. The central moments of the nonparametric distribution are estimated using numerical integration. The nonparametric distribution functions are provided in Equation 2 to Equation 4. Percentiles are estimated using the inverse cumulative distribution function as shown in Equation 4.

$$f(x) = \frac{p_{i+1} - p_i}{x_{i+1} - x_i}, \qquad \textit{Equation 2}$$

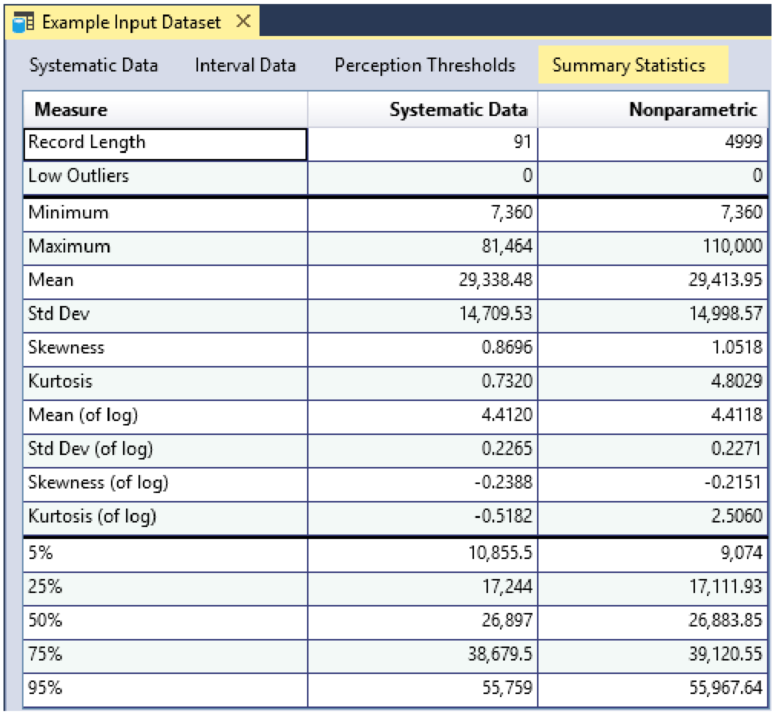
where $f(x)$ is the probability density function (PDF) of the variable X; there is an array of continuous values $\{x\} = \{x_1, x_2, \dots, x_n\}$ for $x_i \leq x < x_{i+1}$ with non-exceedance probabilities $\{p\} = \{p_1, p_2, \dots, p_n\}$ with $0 \leq p_i \leq 1$.

$$F(x) = p_i + (p_{i+1} - p_i)\left(\frac{x - x_i}{x_{i+1} - x_i}\right) \qquad \textit{Equation 3}$$

$$F^{-1}(p) = x_i + (x_{i+1} - x_i)\left(\frac{p - p_i}{p_{i+1} - p_i}\right) \qquad \textit{Equation 4}$$

where $F(x)$ is the cumulative distribution function (CDF) of the variable X; $F^{-1}(p)$ is the inverse CDF; and there is an array of continuous values $\{x\} = \{x_1, x_2, \dots, x_n\}$ for $x_i \leq x \leq x_{i+1}$ with non-exceedance probabilities $\{p\} = \{p_1, p_2, \dots, p_n\}$ with $0 \leq p_i \leq 1$ and $p_i \leq p \leq p_{i+1}$.

The summary statistics provide a preview for what to expect when performing distribution fitting or Bayesian estimation. For example, in the case of Blakely Mountain Dam, we can see that the inclusion of historical and paleoflood data slightly increased the skewness of the data. The systematic data has a skewness (of log) of -0.2388; whereas the nonparametric analysis, which includes all of the data, has a skewness (of log) of -0.2151. We should expect to see a similar behavior when fitting the Log-Pearson Type III distribution.



Measure	Systematic Data	Nonparametric
Record Length	91	4999
Low Outliers	0	0
Minimum	7,360	7,360
Maximum	81,464	110,000
Mean	29,338.48	29,413.95
Std Dev	14,709.53	14,998.57
Skewness	0.8696	1.0518
Kurtosis	0.7320	4.8029
Mean (of log)	4.4120	4.4118
Std Dev (of log)	0.2265	0.2271
Skewness (of log)	-0.2388	-0.2151
Kurtosis (of log)	-0.5182	2.5060
5%	10,855.5	9,074
25%	17,244	17,111.93
50%	26,897	26,883.85
75%	38,679.5	39,120.55
95%	55,759	55,967.64

Figure 60 – Input Data Summary Statistics.

Distribution Fitting Analysis

The **Distribution Fitting Analysis** in RMC-BestFit uses the method of Maximum Likelihood Estimation (MLE) to fit several univariate probability distributions to the user-specified **Input Data**. You can use the distribution fitting analysis results to inform model selection for use in the Bayesian estimation analysis. For each fitted distribution, RMC-BestFit provides three goodness-of-fit measures: the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC), and Root-Mean Squared Error (RMSE). These measures indicate how well the distribution fits the input data, with a smaller value representing a better fit.

To fit distribution with RMC-BestFit, there are four steps required:

- Define **Input Data**.
- Run the fitting analysis.
- Interpret the results.
- Select a distribution to use in the **Bayesian Estimation Analysis**.

Further details of these steps are discussed in the following sections.

Create New Distribution Fitting Analysis

Let's create a new **Distribution Fitting Analysis**. Right-click on the **Distribution Fitting Analysis** folder header and click **Create New…** as shown in Figure 61. Next, give the fitting analysis a name and click **OK**.

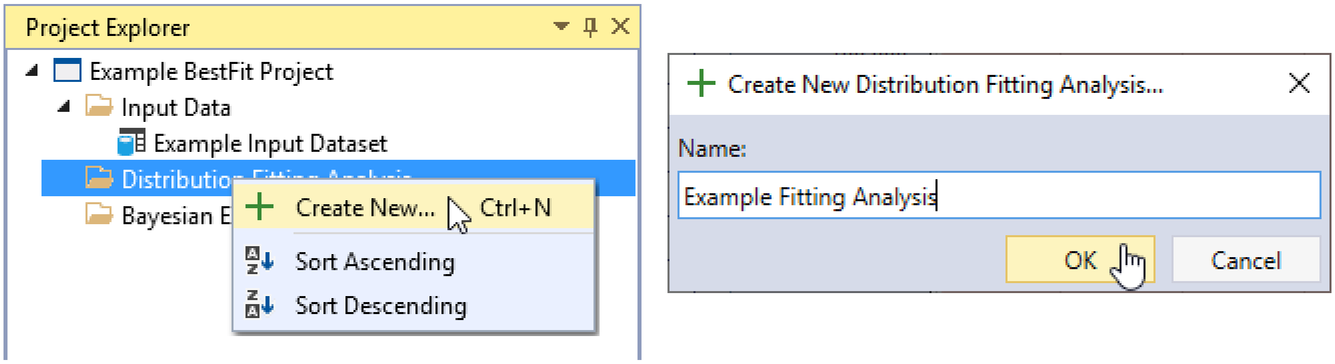


Figure 61 – Create New Distribution Fitting Analysis.

Once, the new **Distribution Fitting Analysis** is created, it will be automatically opened into the **Tabbed Documents** area, and the fitting analysis properties will be displayed in the **Properties Window**. From here, you can set the **Description**, **Input Data, Output Frequency Ordinates**, and **Fit Distributions**.

Define Input Data

Click the **Input Data** drop-down and select the desired data for the fitting analysis as shown in Figure 62. You can set the **Output Frequency Ordinates** by clicking on the **Options** tab at the top of the **Properties Window** as shown in Figure 63. The output frequency ordinates are the annual exceedance probabilities (AEP) used for plotting the fitted distributions on the frequency plot. The default frequency ordinates range from 0.99 to 1E-6 AEP.

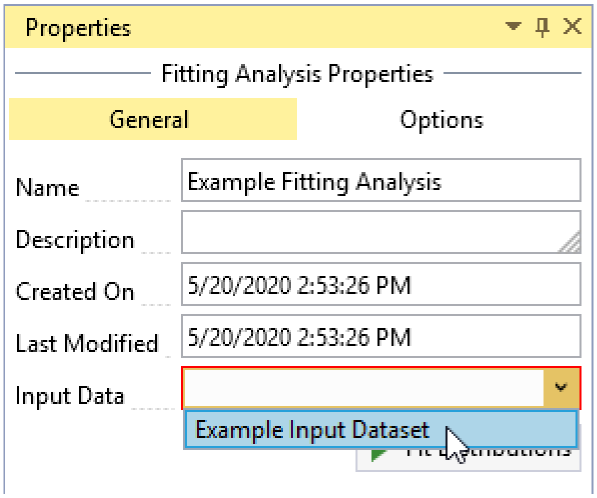


Figure 62 – Distribution Fitting Analysis Properties.

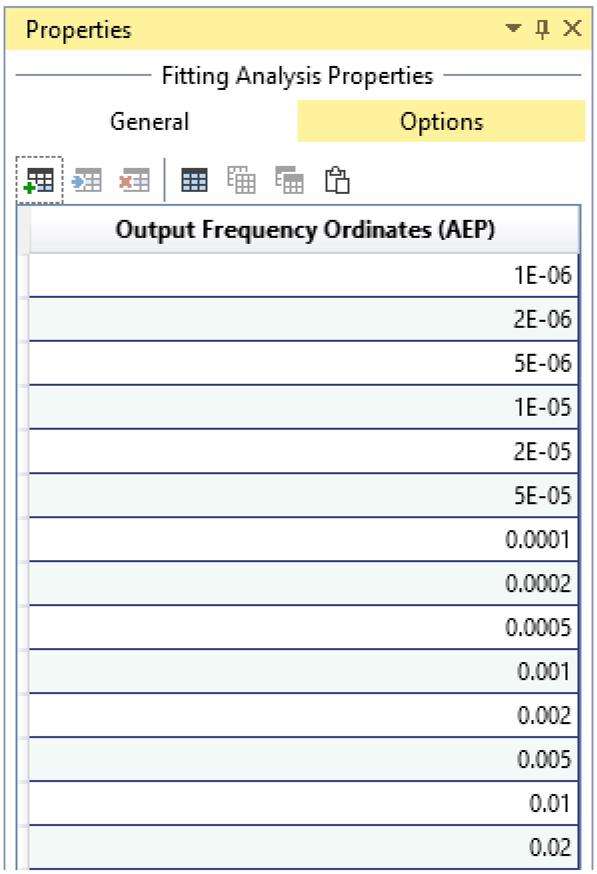


Figure 63 – Distribution Fitting Analysis Output Frequency Ordinates.

Run the Fitting Analysis

After you have selected the **Input Data**, click the **Fit Distributions** command button to run the distribution fitting analysis. The runtime typically takes less than a second. When the analysis is complete, you will see a table of goodness-of-fit measures and all of the distributions plotted on the **Frequency Plot** as shown in Figure 65.

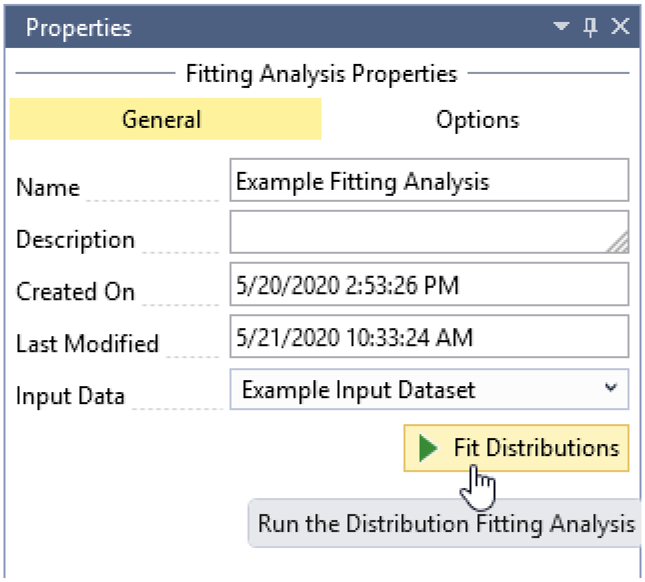


Figure 64 – Run the Distribution Fitting Analysis.

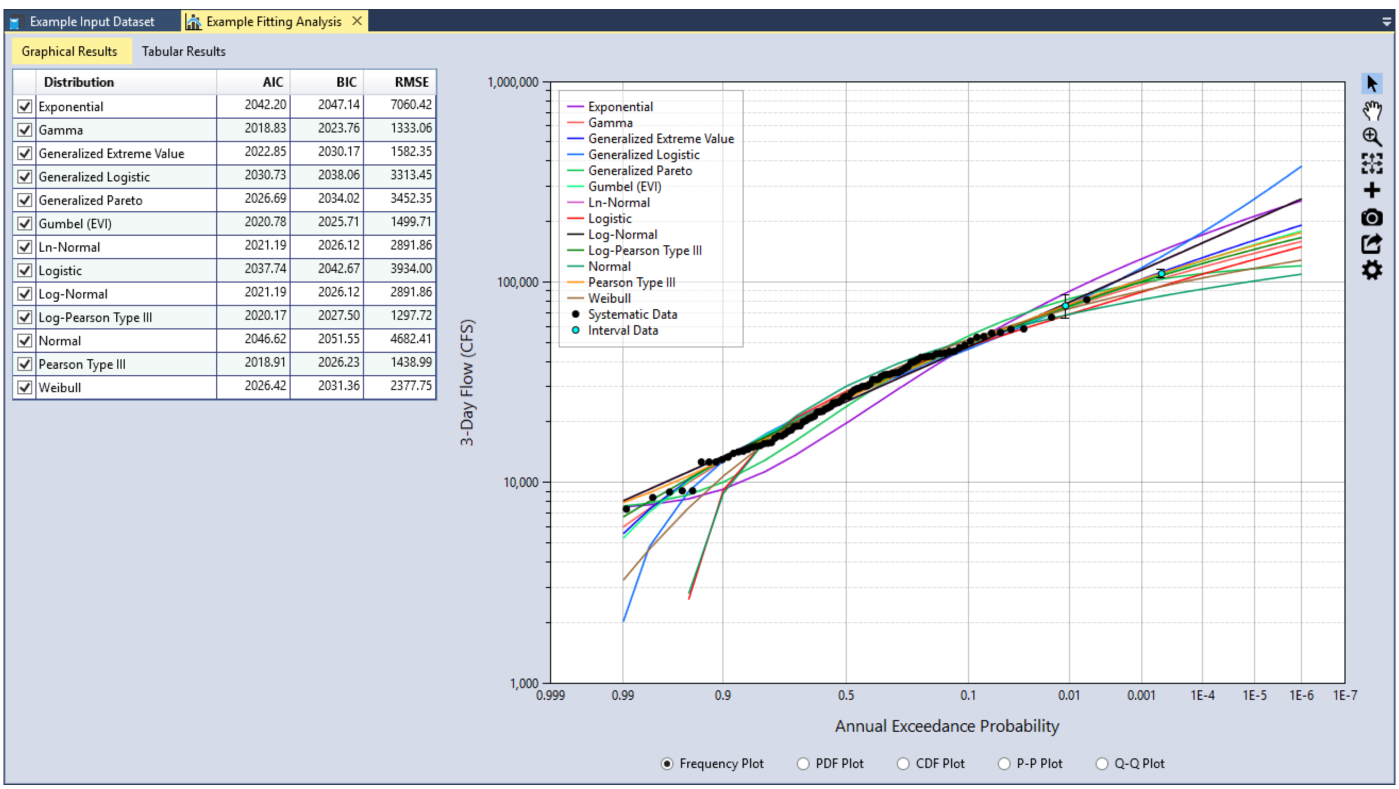


Figure 65 – Distribution Fitting Analysis Graphical Results.

Maximum Likelihood Estimation (MLE)

In the distribution fitting analysis, parameters are estimated using the MLE method. The MLE method formulates a likelihood function using sample data $D = (X_1, \ldots, X_n)$ and the parameters θ of the probability distribution, and solves for the value of the parameters that maximize the likelihood function (Rao & Hamed, 2000) (Jongejan, 2018). The likelihood function gives the probability of the data conditional on the distribution parameters (Equation 5).

$$L_S(D|\theta) = \prod_{i=1}^{n_S} f(X_i|\theta) \qquad \textit{Equation 5}$$

where D is the sample of systematically recorded annual discharge maxima $(X_1, \ldots, X_{n_S})$; and $f(\cdot)$ is the probability density function (PDF) of the variable X. Censored data can be incorporated into the MLE method by augmenting the likelihood function. Left-censored threshold data has the following likelihood function:

$$L_L(D|\theta) = \prod_{i=1}^{n_L} \binom{h}{k} F(X_0|\theta)^{(h-k)} \qquad \textit{Equation 6}$$

where X_0 is the threshold; h is the threshold period; k is the number of observations that exceeded the threshold during the period; $\binom{h}{k}$ is the binomial coefficient; and $F(\cdot)$ is the cumulative distribution function (CDF) of the variable X_0. The binomial coefficient can be dropped from Equation 6 because it will be held constant as θ is varied. Interval-censored data has the following likelihood function:

$$L_I(D|\theta) = \prod_{i=1}^{n_I} \left[F\left(X_{U_i}|\theta\right) - F\left(X_{L_i}|\theta\right)\right] \qquad \textit{Equation 7}$$

where there are n_I observations known to lie between upper and lower bounds, X_U and X_L. The overall likelihood function is then constructed by multiplying the components:

$$L(D|\theta) = L_S(D|\theta) \cdot L_L(D|\theta) \cdot L_I(D|\theta) \qquad \textit{Equation 8}$$

These likelihood formulations for censored data are consistent with those presented in (Stedinger & Cohn, 1986), (Kuczera G. , 1999), and (O'Connell, Ostenaa, Levish, & Klinger, 2002).

From the perspective of Bayesian estimation, MLE is a special case of *maximum a posteriori* (MAP) that assumes a uniform prior distribution for each model parameter. Therefore, if we assume uniform priors in the **Bayesian Estimation Analysis**, we will get the same posterior mode as the MLE method used in the **Distribution Fitting Analysis** (any differences in results would be attributed to convergence errors).

RMC-BestFit uses the Nelder-Mead method (also commonly called the downhill simplex method or amoeba method) to perform MLE for every distribution. The Nelder-Mead method finds the parameter set that maximizes the likelihood function using a direct search method.

Interpret the Results

Once the **Distribution Fitting Analysis** is complete, you should evaluate the results. RMC-BestFit provides goodness-of-fit measures and comparison graphs to help you evaluate the fits and select the best probability distribution to use in the **Bayesian Estimation Analysis**.

Goodness-of-Fit Measures

RMC-BestFit provides three goodness-of-fit measures: AIC, BIC, and RMSE. These measures indicate how well the distribution fits the input data, with a smaller value representing a better fit. The goodness-of-fit statistics are used for two purposes: 1) *Model selection* is the process of picking one fitted distribution over another; 2) whereas, *fit validation* is the process of determining whether a fitted distribution agrees well with the data.

AIC and BIC are used for model selection among a finite set of models (the term *model* is synonymous with *probability distribution*). The model with the lowest AIC or BIC is preferred. When comparing multiple models, additional parameters often yield larger, optimized log-likelihood values. AIC and BIC penalize for more complex models, i.e., models with additional parameters. However, for BIC, the penalty is a function of the sample size, and so it is typically more severe than that of AIC. The formulas for AIC and BIC are shown in Equation 9 and Equation 10, respectively. To address potential over-fitting, RMC-BestFit implements a correction for small sample sizes for AIC.

$$AIC = 2k - 2\ln(\hat{L}) + \frac{2k^2 + 2k}{n - k - 1}$$ *Equation 9*

$$BIC = \ln(n)\, k - 2\ln(\hat{L})$$ *Equation 10*

where k is the number of parameters; n is the sample size; and $\hat{L}$ is the maximum value of the likelihood function for the model.

RSME provides a measure for fit validation, with smaller values indicating a better fit. RMSE is computed based on how well the probability distribution agrees with the plotting positions of the input data. You can set the plotting position parameter based on preference or theoretical motives, so this measure has the potential to be biased. To minimize this issue, the default plotting position coefficient in the input data interface is set to Weibull ($\alpha = 0$), which is unbiased. The formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$ *Equation 11*

where n is the sample size; $\hat{y}_i$ is the predicted value for item i; and y_i is the observed value for step i.

You can sort each column of the goodness-of-fit table in ascending or descending order by right-clicking the column header as shown in Figure 66. As you move your cursor over the table, a tooltip will appear showing the fitted parameters of the distribution as shown in Figure 67. You may check or uncheck distributions in this table to add or remove the fitted distributions from the comparison plots.

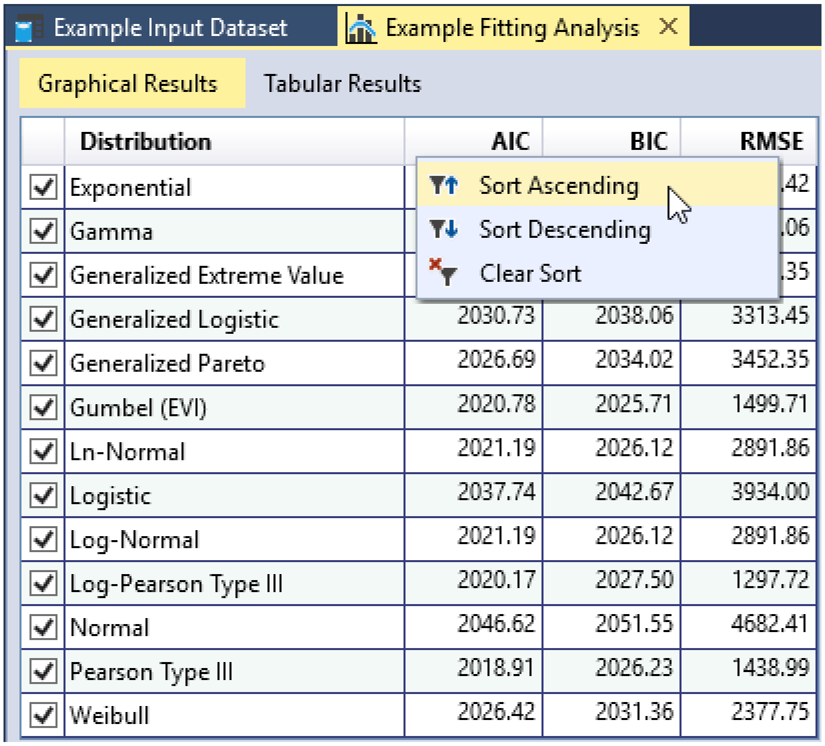


Figure 66 – Distribution Fitting Analysis Goodness-of-Fit Table.

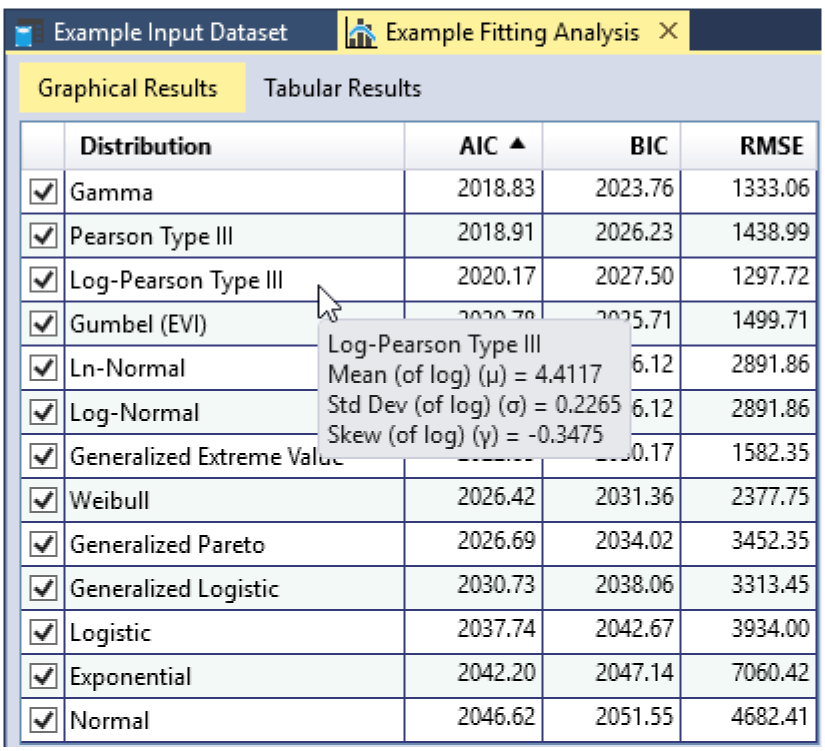


Figure 67 – Check or Uncheck Distributions from the Goodness-of-Fit Table.

Comparison Plots

RMC-BestFit provides five types of graphs to help you visually assess the quality of the distribution fits. A comparison graph plots the input data and fitted distributions on the same graph, allowing you to visually compare them. The graphs allow you to determine whether the fitted distribution matches the input data in critical areas. For example, for flood frequency analyses, it is important to have good agreement in the extreme, right-hand tail of the distribution.

Frequency Plot

A **Frequency Plot** is a plot of magnitude versus annual exceedance probability (AEP). AEP is typically plotted on the X-axis using a Normal or Gumbel probability scale to linearize the plot and exaggerate the extreme right-hand tail of the data. The frequency plot compares the fitted distributions to the plotting positions of the input data as shown in Figure 68.

For demonstration purposes, let's only evaluate the three-parameter distributions. Uncheck all of the two-parameter distributions in the goodness-of-fit table. Then sort the RMSE column in ascending order as shown in Figure 68. We see that the Log-Pearson Type III (LPIII) distribution provides the smallest RMSE. In addition, we can also see that it provides a very good fit through the input data plotted in the **Frequency Plot**. The Pearson Type III (PIII) and Generalized Extreme Value (GEV) distributions also fit through the data well and have a relatively small RSME.

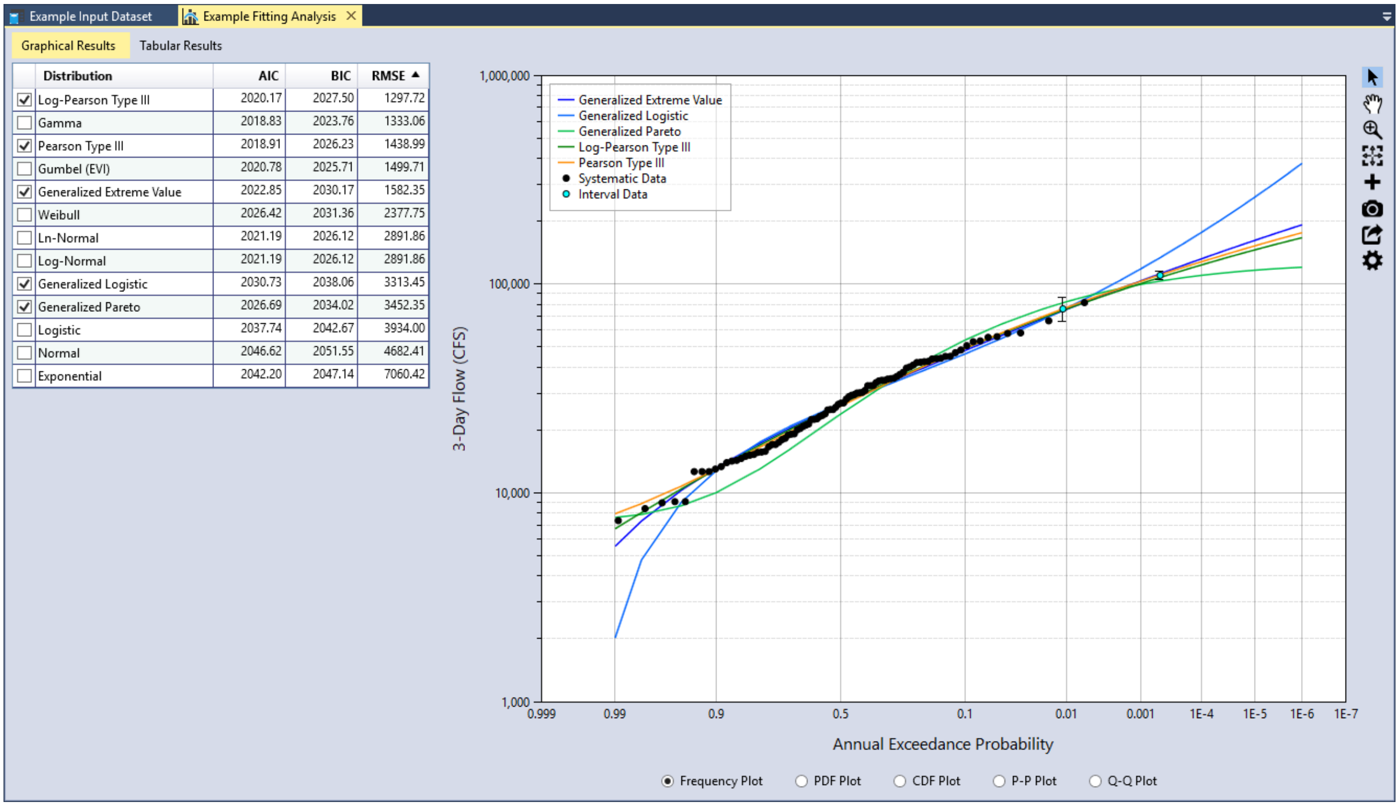


Figure 68 – Distribution Fitting Analysis Frequency Plot.

PDF Plot

A **PDF Plot** compares the probability density function (PDF) of the fitted distribution to a histogram of the input data as shown in Figure 69. The **Frequency Plot** and **PDF Plot** are usually the most informative comparisons. With the PDF plot, it is easy to see the where the highest discrepancies are and whether the general shape of the data and fitted distributions agree well. From the PDF plot, we can see that Generalized Pareto (GPA) distribution does not fit the data as well as the others.

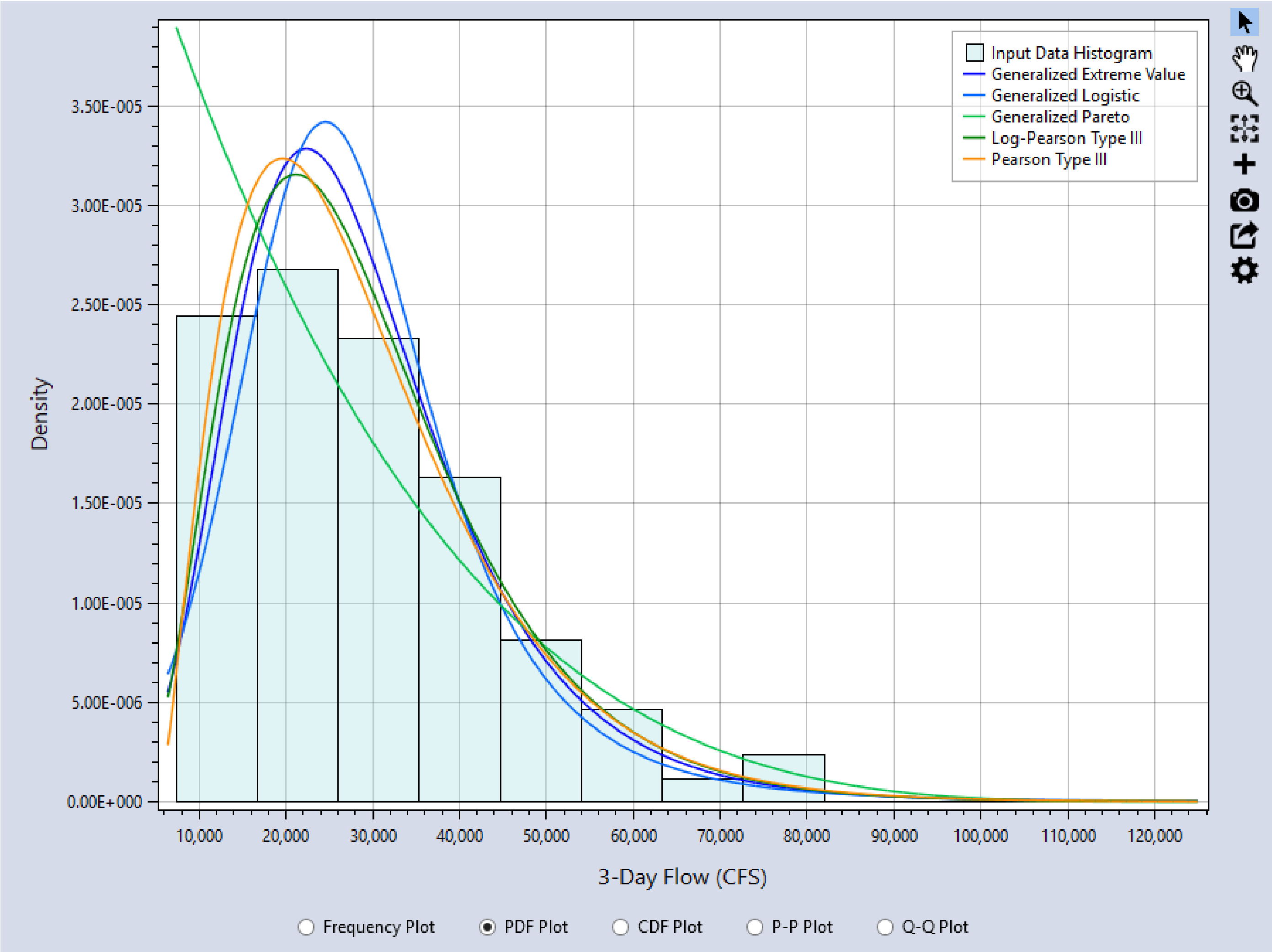


Figure 69 – Distribution Fitting Analysis PDF Plot.

CDF Plot

The **CDF Plot** compares the cumulative distribution function (CDF) of the fitted distribution to the plotting positions of the input data as shown in Figure 70. The CDF plot has a very insensitive scale, and is not very useful for comparing the location, spread, and shape of the distributions, for which the PDF plot is much better. In many cases, the CDF plot will not provide a good visual measure for the goodness-of-fit.

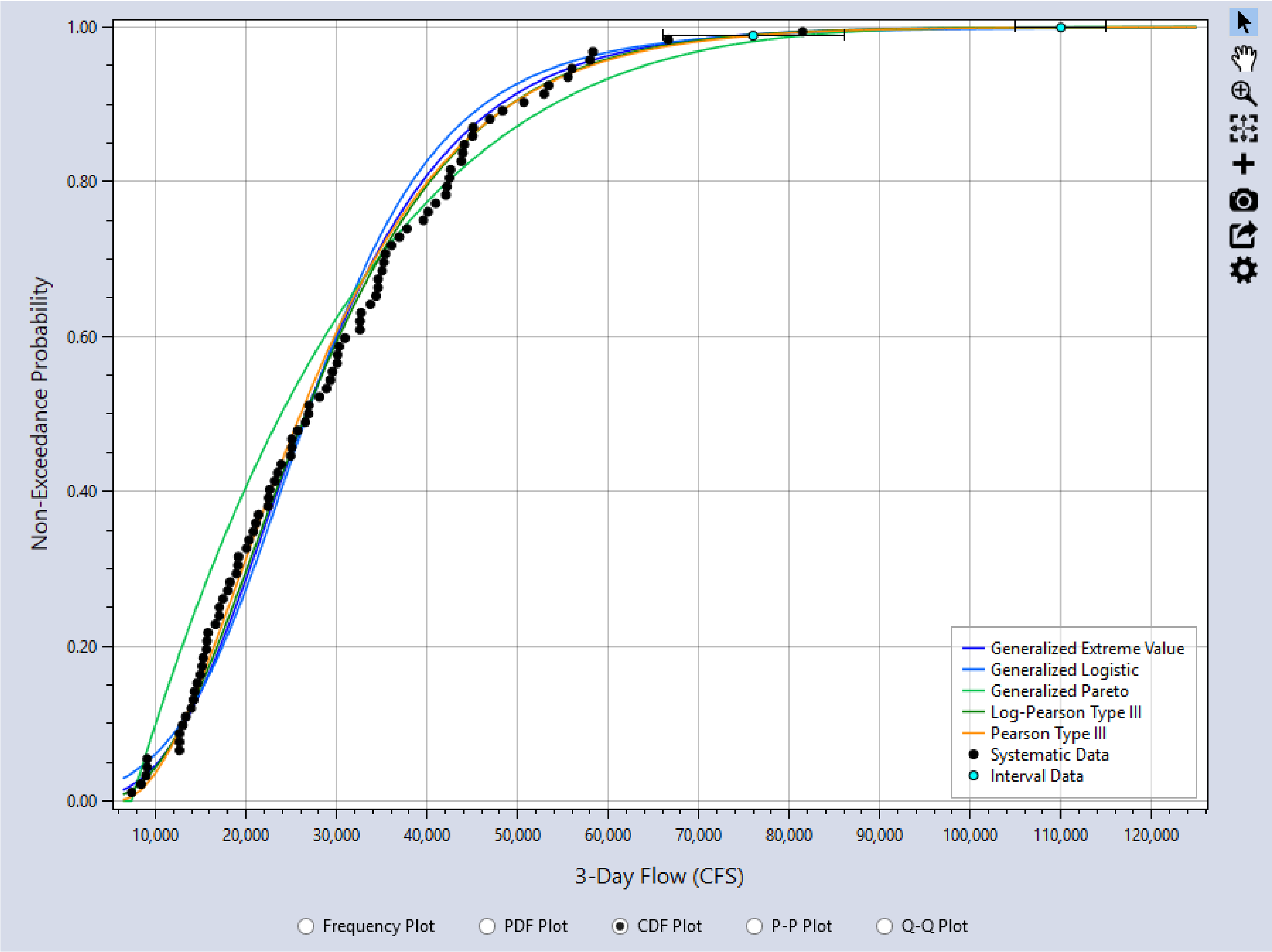


Figure 70 – Distribution Fitting Analysis CDF Plot.

P-P Plot

The Probability-Probability (P-P) plot graphs the F(x) of the model (distribution) versus the input data plotting positions. The closer the plot resembles the diagonal 1:1 line, the better the fit. The **P-P Plot** can be useful if you are interested in closely matching cumulative percentiles as it shows the differences between the middle of the fitted distributions and the input data. From the P-P plot, we can see that the GPA distribution has the poorest agreement with the data.

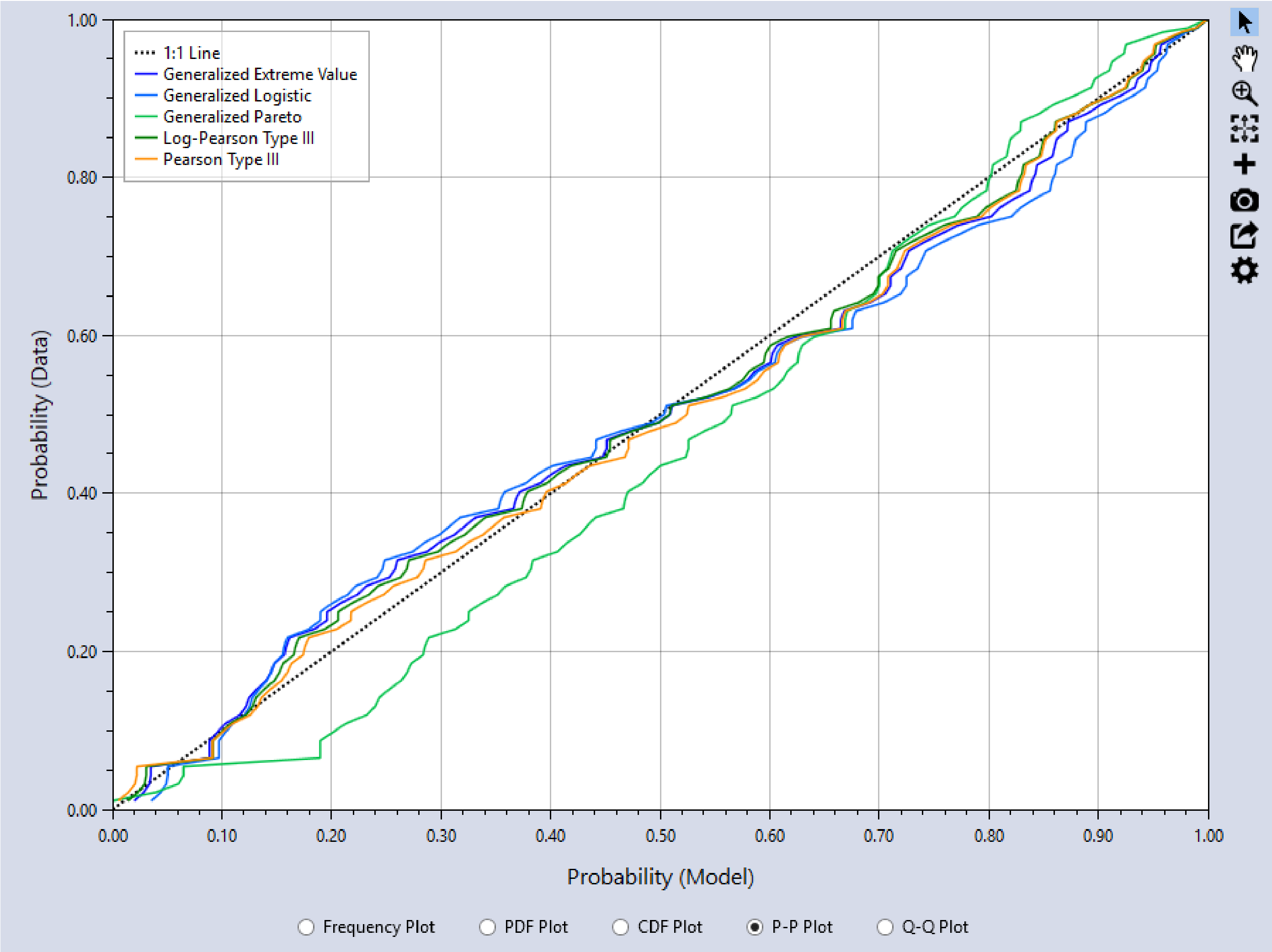


Figure 71 – Distribution Fitting Analysis P-P Plot.

Q-Q Plot

The Quantile-Quantile (Q-Q) plot graphs the inverse CDF of the model versus the percentile values of the input data. Again, the closer the plot resembles the diagonal 1:1 line, the better the fit. The **Q-Q Plot** can be useful if you are interested in closely matching cumulative percentiles as it shows the differences between the tails of the fitted distributions and the input data. From the Q-Q plot, we can see that the PIII, LPIII, and GEV distributions all fit the right-hand tail of the data really well. The Generalized Logistic (GLO) distribution has the poorest agreement with the right-hand tail of the data, followed by the GPA distribution as the second least favorable fit.

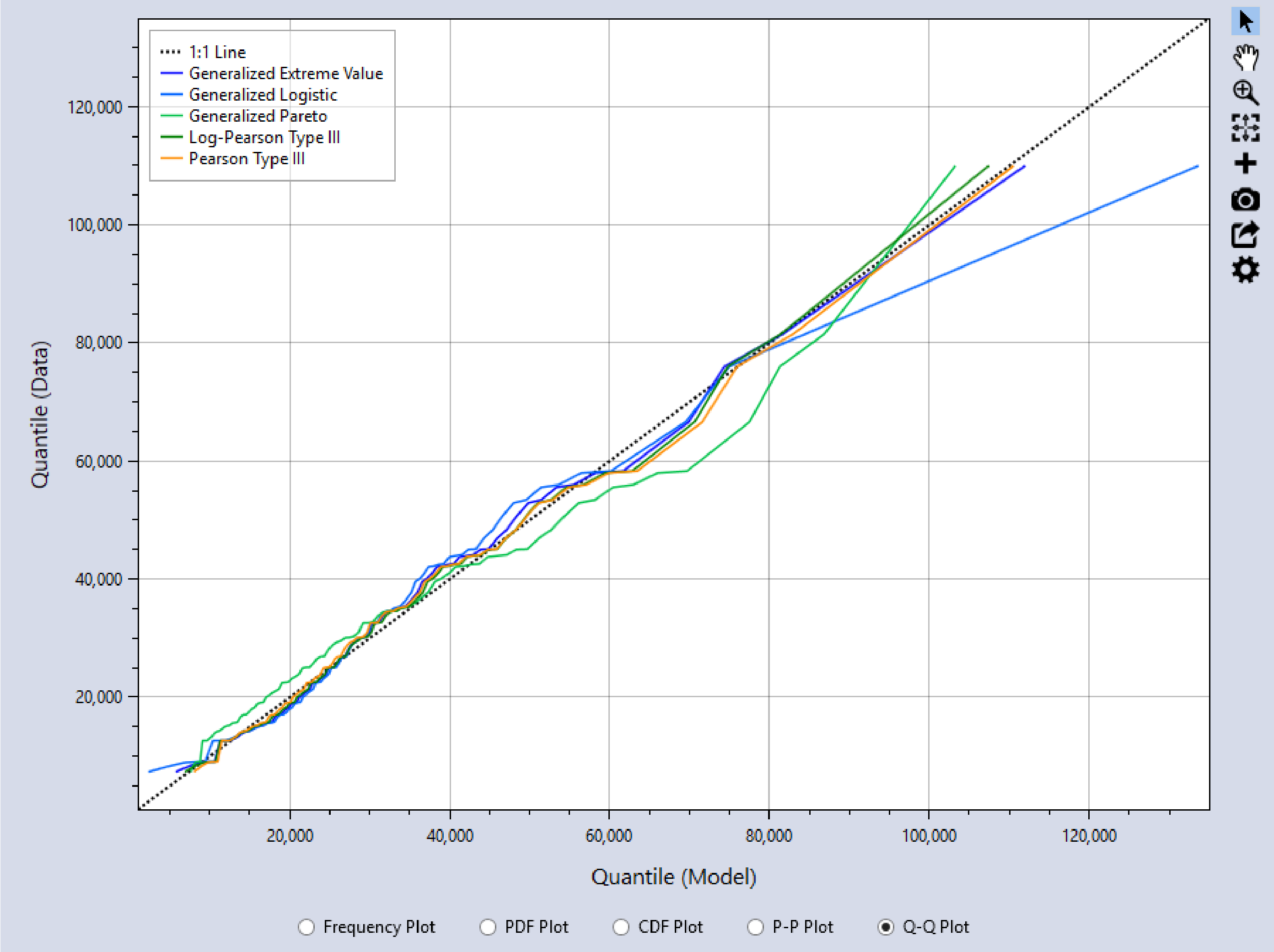
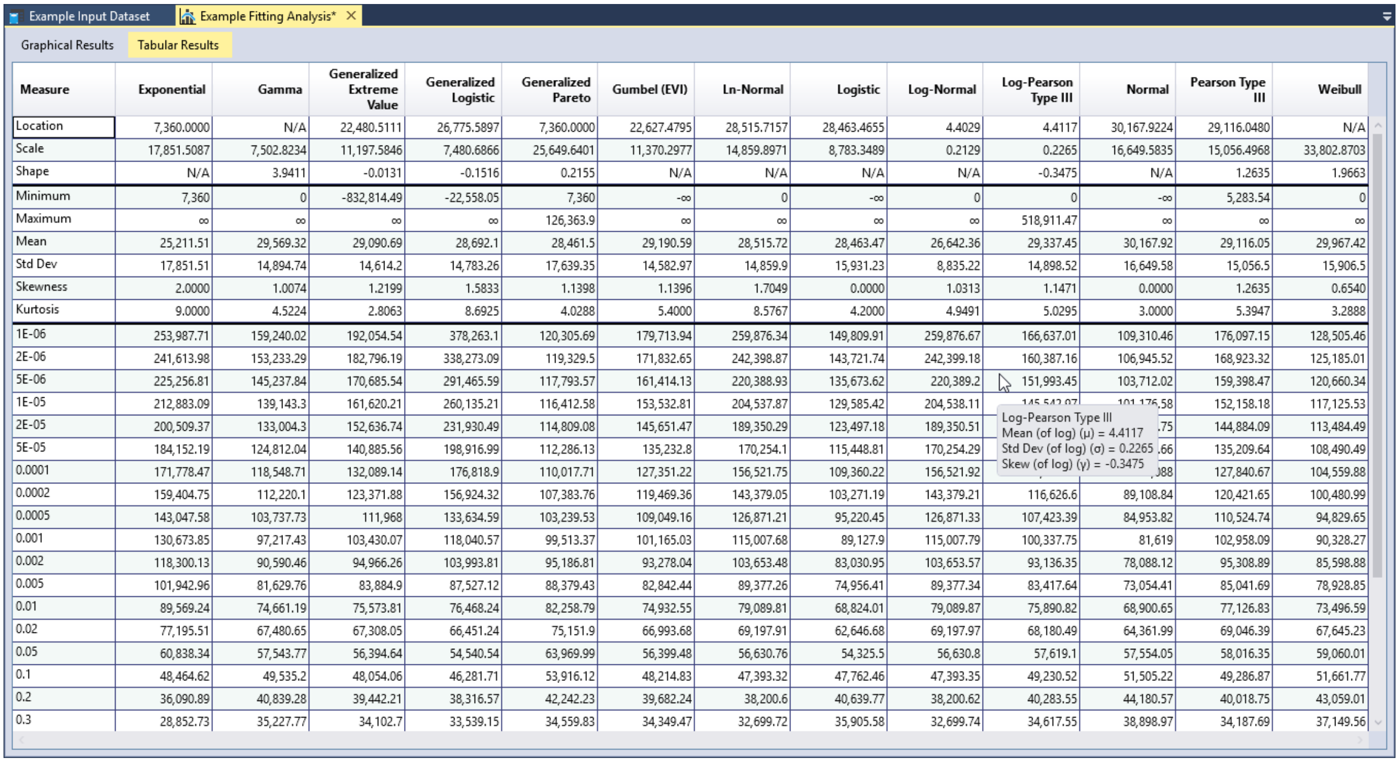


Figure 72 – Distribution Fitting Analysis Q-Q Plot.

Tabular Results

RMC-BestFit reports the parameters and basic statistics (mean, standard deviation, skewness, etc.) for each of the fitted distributions, which can also be compared to the same statistics for the input data. Quantiles for the specified output frequency AEPs are also provided in the **Tabular Results**.



Measure	Exponential	Gamma	Generalized Extreme Value	Generalized Logistic	Generalized Pareto	Gumbel (EVI)	Ln-Normal	Logistic	Log-Normal	Log-Pearson Type III	Normal	Pearson Type III	Weibull
Location	7,360.0000	N/A	22,480.5111	26,775.5897	7,360.0000	22,627.4795	28,515.7157	28,463.4655	4.4029	4.4117	30,167.9224	29,116.0480	N/A
Scale	17,851.5087	7,502.8234	11,197.5846	7,480.6866	25,649.6401	11,370.2977	14,859.8971	8,783.3489	0.2129	0.2265	16,649.5835	15,056.4968	33,802.8703
Shape	N/A	3.9411	-0.0131	-0.1516	0.2155	N/A	N/A	N/A	N/A	-0.3475	N/A	1.2635	1.9663
Minimum	7,360	0	-832,814.49	-22,558.05	7,360	-∞	0	-∞	0	0	-∞	5,283.54	0
Maximum	∞	∞	∞	∞	126,363.9	∞	∞	∞	∞	518,911.47	∞	∞	∞
Mean	25,211.51	29,569.32	29,090.69	28,692.1	28,461.5	29,190.59	28,515.72	28,463.47	26,642.36	29,337.45	30,167.92	29,116.05	29,967.42
Std Dev	17,851.51	14,894.74	14,614.2	14,783.26	17,639.35	14,582.97	14,859.9	15,931.23	8,835.22	14,898.52	16,649.58	15,056.5	15,906.5
Skewness	2.0000	1.0074	1.2199	1.5833	1.1398	1.1396	1.7049	0.0000	1.0313	1.1471	0.0000	1.2635	0.6540
Kurtosis	9.0000	4.5224	2.8063	8.6925	4.0288	5.4000	8.5767	4.2000	4.9491	5.0295	3.0000	5.3947	3.2888
1E-06	253,987.71	159,240.02	192,054.54	378,263.1	120,305.69	179,713.94	259,876.34	149,809.91	259,876.67	166,637.01	109,310.46	176,097.15	128,505.46
2E-06	241,613.98	153,233.29	182,796.19	338,273.09	119,329.5	171,832.65	242,398.87	143,721.74	242,399.18	160,387.16	106,945.52	168,923.32	125,185.01
5E-06	225,256.81	145,237.84	170,685.54	291,465.59	117,793.57	161,414.13	220,388.93	135,673.62	220,389.2	151,993.45	103,712.02	159,398.47	120,660.34
1E-05	212,883.09	139,143.3	161,620.21	260,135.21	116,412.58	153,532.81	204,537.87	129,585.42	204,538.11	145,542.97	101,176.58	152,158.18	117,125.53
2E-05	200,509.37	133,004.3	152,636.74	231,930.49	114,809.08	145,651.47	189,350.29	123,497.18	189,350.51	[illegible]	[illegible]	144,884.09	113,484.49
5E-05	184,152.19	124,812.04	140,885.56	198,916.99	112,286.13	135,232.8	170,254.1	115,448.81	170,254.29	[illegible]	[illegible]	135,209.64	108,490.49
0.0001	171,778.47	118,548.71	132,089.14	176,818.9	110,017.71	127,351.22	156,521.75	109,360.22	156,521.92	[illegible]	[illegible]	127,840.67	104,559.88
0.0002	159,404.75	112,220.1	123,371.88	156,924.32	107,383.76	119,469.36	143,379.05	103,271.19	143,379.21	116,626.6	89,108.84	120,421.65	100,480.99
0.0005	143,047.58	103,737.73	111,968	133,634.59	103,239.53	109,049.16	126,871.21	95,220.45	126,871.33	107,423.39	84,953.82	110,524.74	94,829.65
0.001	130,673.85	97,217.43	103,430.07	118,040.57	99,513.37	101,165.03	115,007.68	89,127.9	115,007.79	100,337.75	81,619	102,958.09	90,328.27
0.002	118,300.13	90,590.46	94,966.26	103,993.81	95,186.81	93,278.04	103,653.48	83,030.95	103,653.57	93,136.35	78,088.12	95,308.89	85,598.88
0.005	101,942.96	81,629.76	83,884.9	87,527.12	88,379.43	82,842.44	89,377.26	74,956.41	89,377.34	83,417.64	73,054.41	85,041.69	78,928.85
0.01	89,569.24	74,661.19	75,573.81	76,468.24	82,258.79	74,932.55	79,089.81	68,824.01	79,089.87	75,890.82	68,900.65	77,126.83	73,496.59
0.02	77,195.51	67,480.65	67,308.05	66,451.24	75,151.9	66,993.68	69,197.91	62,646.68	69,197.97	68,180.49	64,361.99	69,046.39	67,645.23
0.05	60,838.34	57,543.77	56,394.64	54,540.54	63,969.99	56,399.48	56,630.76	54,325.5	56,630.8	57,619.1	57,554.05	58,016.35	59,060.01
0.1	48,464.62	49,535.2	48,054.06	46,281.71	53,916.12	48,214.83	47,393.32	47,762.46	47,393.35	49,230.52	51,505.22	49,286.87	51,661.77
0.2	36,090.89	40,839.28	39,442.21	38,316.57	42,242.23	39,682.24	38,200.6	40,639.77	38,200.62	40,283.55	44,180.57	40,018.75	43,059.01
0.3	28,852.73	35,227.77	34,102.7	33,539.15	34,559.83	34,349.47	32,699.72	35,905.58	32,699.74	34,617.55	38,898.97	34,187.69	37,149.56

Figure 73 – Distribution Fitting Analysis Tabular Results.

Selecting a Probability Distribution

Finally, you need to select a probability distribution to carry forward to the **Bayesian Estimation Analysis**. The following steps can help you choose an appropriate distribution:

- Start by reviewing the descriptions of the probability distributions found in the Appendix. Then, look at the variable in question. Does the data have bounds? Is it symmetric or skewed? Which distributions are theoretically appropriate for the data?
- Select candidate distributions that best characterize the variable.
- Then, use the **Distribution Fitting Analysis** results to select the distribution that best describes your **Input Data**.

In the above example, we evaluated how well the three-parameter distributions fit 3-day inflow data at Blakely Mountain Dam. It was clear that the PIII, LPIII, and GEV distributions produced better fits than the GLO and GPA distributions. Flow data can span several orders in magnitude (e.g., 1,000 cfs to 1,000,000 cfs), and is typically skewed and cannot have negative values. If the skewness of the data is greater than the absolute value of 2 ($C_s > |2|$), then the maximum likelihood estimation method cannot produce a solution for the PIII and LPIII distributions. Real-space flow data can often have skews much larger than 2. Therefore, LPIII better characterizes flow data than the PIII distribution (U.S. Geological Survey, 1982) (U.S. Geological Survey, 2018). As such, the PIII distribution is not considered appropriate for characterizing the flow data.

When considering the statistical and graphical goodness-of-fit performance, and the appropriateness of the distribution, the LPIII distribution fits the **Input Data** the better than the GEV distribution. The LPIII distribution will now be carried forward to the Bayesian estimation analysis.

Bayesian Estimation Analysis

RMC-BestFit performs Bayesian estimation using a Markov Chain Monte Carlo (MCMC) algorithm to estimate distribution parameters given the specified input data and parent distribution. The Bayesian estimation method produces the most likely estimate for parameters (posterior mode) and a characterization of the parameter uncertainty.

To perform Bayesian Estimation with RMC-BestFit, there are four steps required:

- Define **Input Data** and select the parent probability distribution.
- Run the **Bayesian Estimation Analysis**.
- Diagnose convergence.
- Explore the results.

Further details of these steps are discussed in the following sections.

Create New Bayesian Analysis

Let's create a new **Bayesian Estimation Analysis**. Right-click on the **Bayesian Estimation Analysis** folder header and click **Create New…** as shown in Figure 61. Next, give the Bayesian analysis a name and click **OK**.

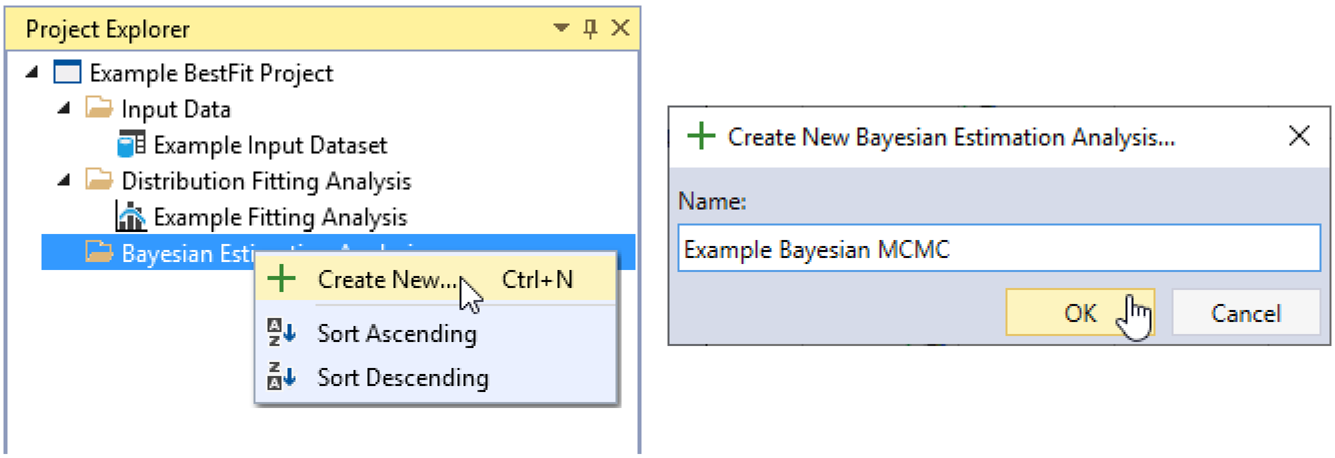


Figure 74 – Create New Bayesian Estimation Analysis.

Once, the new **Bayesian Estimation Analysis** is created, it will be automatically opened into the **Tabbed Documents** area, and the Bayesian analysis properties will be displayed in the **Properties Window**. From here, you can set the required inputs.

Bayesian Analysis Framework

In Bayesian analysis, the values of the parent probability distribution parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ converge to a distribution rather than to a single best value. The uncertainty in the parameters is represented by a *prior* probability distribution $P(\theta)$, which is established based on information available *a priori*. This prior distribution is not derived from the observed data $D = (X_1, \ldots, X_n)$, but instead comes from other sources that can be either subjective (e.g., expert opinion) or objective (e.g., previous statistical analyses, or regional information). After the prior distributions and the observed data are specified, Bayes' theorem (Equation 12) is used to combine the *a priori* information about the parameters with the observed data, using the likelihood $P(D|\theta)$ (Equation 13).

$$P(\theta|\mathrm{D}) = \frac{P(D|\theta) \cdot P(\theta)}{\int P(D|\theta) \cdot P(\theta) \cdot d\theta}$$ *Equation 12*

$$P(D|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$ *Equation 13*

where $P(\theta|\mathrm{D})$ is the posterior probability density function (PDF) of θ; $P(\theta)$ is the prior pdf of θ; and $P(D|\theta)$ is the likelihood function. The posterior cumulative distribution function (CDF) of X now follows from the total probability theorem:

$$F(X) = \int F(X|\theta, D) \cdot P(\theta|D) \cdot d\theta$$ *Equation 14*

which is a probability-weighted sum of the CDFs under different posterior parameter sets $\theta_1, \theta_2, \ldots, \theta_n$. Equation 14 is known as the Bayesian posterior predictive distribution, and is equivalent to the *expected probability of exceedance* concept first presented by (Beard, 1960). Stedinger (1983) and Kuczera (1999) refer to this integral as the design flood distribution, and it is considered the optimal estimator of an exceedance probability.

In most cases, there is not a closed form solution to the denominator of Equation 12. Therefore, Monte Carlo simulation techniques such as Markov Chain Monte Carlo (MCMC) are required. The RMC-BestFit software employs an adaptive Differential Evolution Markov Chain (DE-MC_z) population-based sampler (ter Braak & Vrugt, 2008), which has proven to be very efficient at arriving at the posterior distribution.

Figure 75 illustrates the basic steps in Bayesian analysis. The Bayesian approach offers a framework that is well suited to incorporate different sources of information, such as systematic records, historical data, regional information, and other information along with related uncertainties (Viglione, Merz, Salinas, & Bloschl, 2013). The Bayesian approach allows you to formally include your own expertise into the analysis by choosing *a priori* distributions. The possibility to combine this information with the observed data is even more important because, in practice, the observed records are usually of limited size.

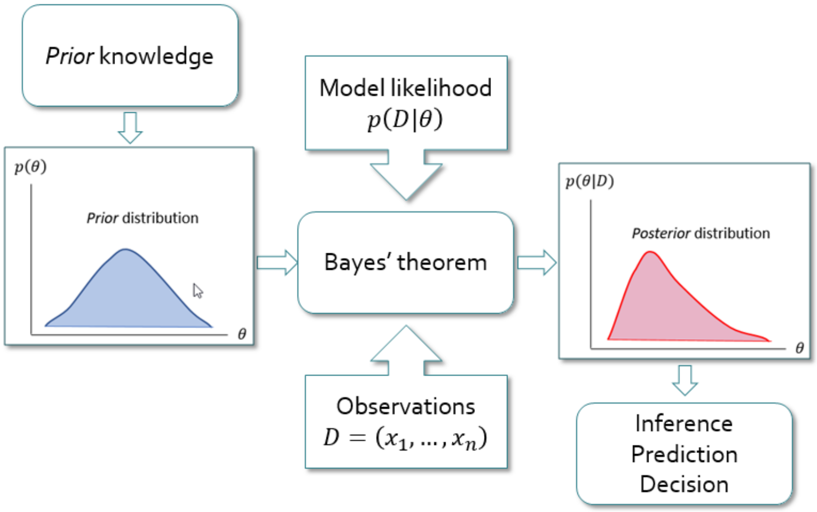


Figure 75 – Diagram Illustrating the Basic Steps in Bayesian Analysis (adapted from (Meylan, 2012), which was originally taken (Perreault, 2000)).

Define Inputs

You must select the **Input Data** and the parent probability distribution to use in the **Bayesian Estimation Analysis**. The parent distribution describes the parent population of your Input Data, which is assumed to be a sample from the parent population. By default, the parent distribution is set as the Generalized Extreme Value (GEV) distribution.

To select your data, click the **Input Data** drop-down on the **General** tab and select the desired data for the Bayesian analysis as shown in Figure 76. Next, set the parent distribution to be the distribution selected from the **Distribution Fitting Analysis**, which in this case was Log-Pearson Type III (LPIII). The prior distributions for parameters and quantiles can also be set from the **General** tab.

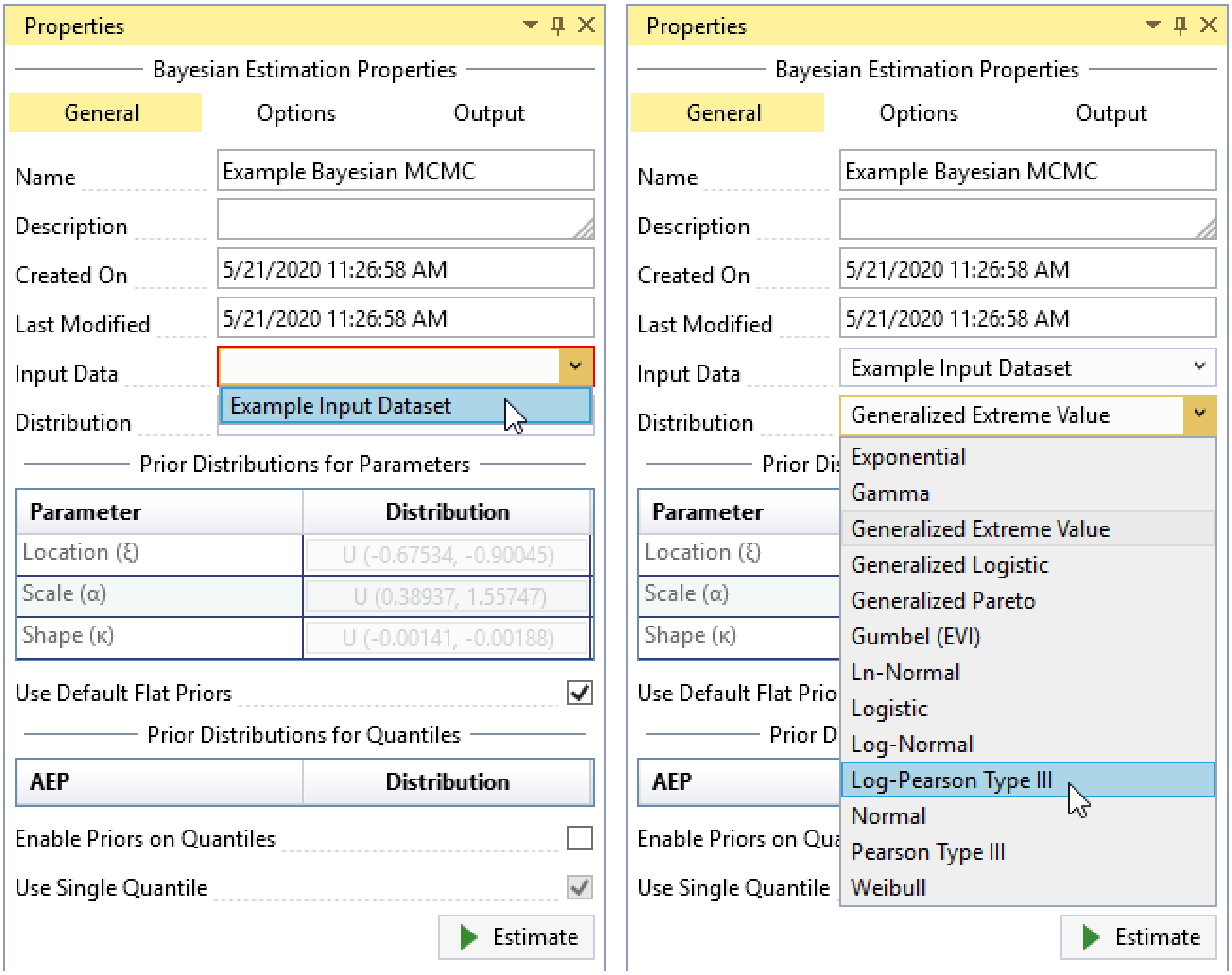


Figure 76 – Bayesian Estimation Analysis General Properties.

Default Flat Priors

After you have selected the **Input Data** and parent distribution, RMC-BestFit automatically develops default flat (uniform) priors for the selected distribution, given the data. The goal of this routine is to develop prior distributions that have minimal impact on the posterior distributions. This approach is sometimes referred to as vague priors, or weakly informative priors. Weakly informative priors contain information to keep the posterior within reasonable bounds without fully capturing one's scientific knowledge about the underlying parameter (Gelman, et al., 2014). There are two approaches to developing a weakly informative prior as described by Gelman et al (2014):

1. Start with some version of an uninformative prior distribution and then add enough information so that inferences are constrained to be reasonable.

2. Start with a strong, highly informative prior and broaden it to account for uncertainty in one's prior beliefs and in the applicability of any historically based prior distributions to new data.

RMC-BestFit develops default flat priors by first considering the parent distribution and parameter support, and then peeking at the data to determine broad upper and lower constraints for the parameters. This ensures the prior distributions for parameters are somewhat centered near the likelihood, but with a much larger spread. The typical end-user of RMC-BestFit will likely not have much advanced training in Bayesian statistics. Therefore, the routine for default flat priors ensures you will get reasonable results out of the box.

The default flat priors are shown in Figure 77. You may uncheck the **Use Default Flat Priors** checkbox to customize the priors. See the Informative Priors section for details on how to set informative priors for parameters and quantiles.

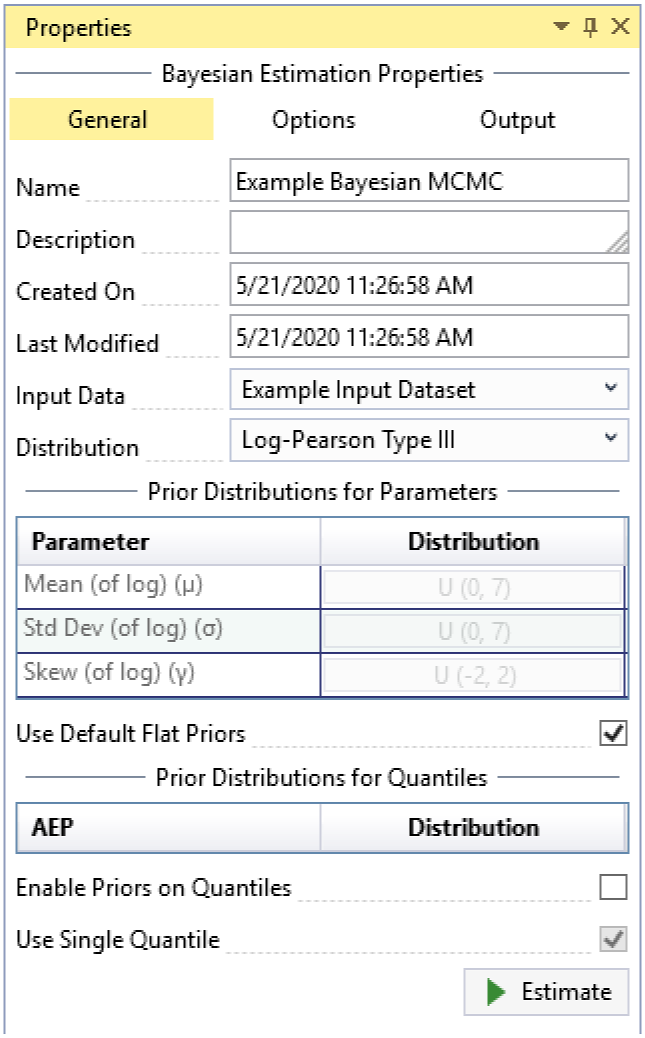


Figure 77 – Default Flat Prior Distributions for Parameters.

Simulation Options

For typical applications, the default simulation options should provide reasonable results out of the box. The **Bayesian Estimation Analysis** has the following simulation options (see Figure 78) available for advanced users:

- ***Number of Chains***: The **Bayesian Estimation Analysis** utilizes an adaptive Differential Evolution Markov Chain (DE-MC_Z) population-based sampler (ter Braak & Vrugt, 2008), in which multiple chains are run in parallel. It is recommended that the number of chains be 2 times the number of parent distribution parameters.

- ***Thinning Interval***: Determines how often the Markov Chain Monte Carlo (MCMC) evolutions will be recorded. Thinning can be used to reduce autocorrelation in the posterior distributions. A thinning interval of 20 means that every 20th iteration will be recorded.

- ***Warm Up Evolutions***: The number of thinned warm up MCMC evolutions to discard at the beginning of the simulation. It is recommended that the warm up be half the length of the number of evolutions; e.g., if the number of evolutions is 4,000, then the warm up length should be 2,000.

- ***Evolutions***: The number of thinned MCMC evolutions to simulate. If the thinning interval is 10 and the number of evolutions is 1,000, there will be a total of 10,000 iterations in the simulation. It is recommended to simulate at least 3,000 thinned evolutions.

- ***PRNG Seed***: The pseudo random number generator (PRNG) seed used within the Monte Carlo simulation. The PRNG ensures repeatability.

- ***Initial Population***: Determines the length of the initial population vector. It is recommended that the initial population be at least 100 times the number of parent distribution parameters in length.

- ***Jump Parameter*** $(\boldsymbol{\gamma})$: The jump parameter allows the simulation to jump from one mode region to another in the target distribution. It is recommended to set $\gamma = 2.38/\sqrt{2d}$, where d is the number of parent distribution parameters.

- ***Jump Threshold***: Determines how often the jump parameter (γ) switches to 1.0. It is recommended that the jump threshold be set to 0.20, which will result in adaptation 20% of the time.

- ***Noise Parameter*** $(\boldsymbol{b})$: A random noise ε is added to the proposal in the MCMC simulation. The noise follows a uniform distribution $U(-b, +b)$. It is recommended that b be very small, such as 0.001.

The simulation options are automatically set with default settings. You can uncheck the **Use Defaults** checkbox to customize the settings.

Output Options

The **Bayesian Estimation Analysis** has the following output options (see Figure 78):

- ***Credible Interval***: Sets the width of the credible interval. For a 90% credible interval, the value of interest lies with a 90% probability in the interval.

- ***Output Length***: The number of posterior parameter sets to output. It is recommended to output 10,000 parameter sets to ensure an accurate 90% credible interval.

- ***Output Frequency Ordinates***: The annual exceedance probabilities (AEP) used for plotting the results on the frequency plot. The default frequency ordinates range from 0.99 to 1E-6 AEP.

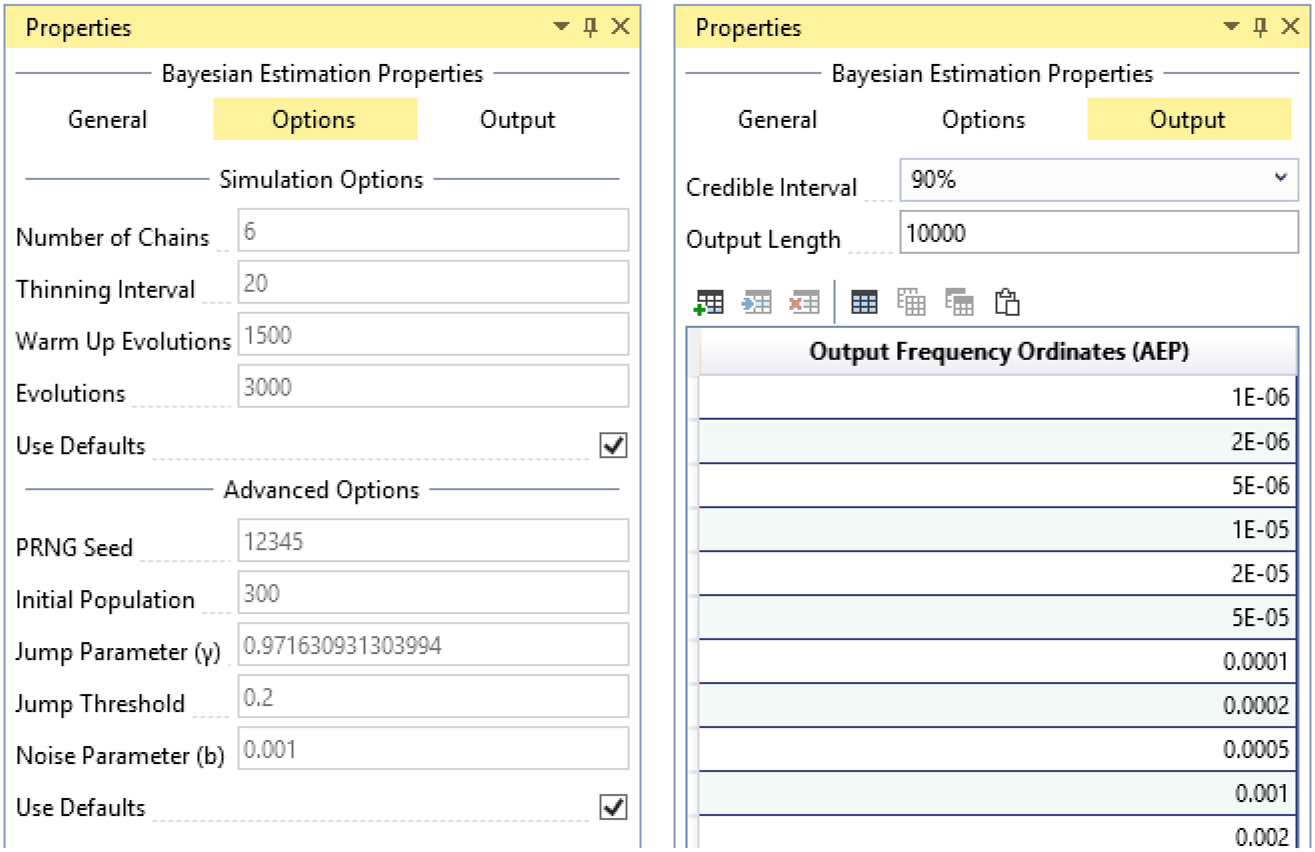


Figure 78 – Bayesian Estimation Analysis Simulation and Output Options.

Run the Bayesian Analysis

After you have defined all of your inputs and settings, click the **Estimate** command button to run the Bayesian analysis. The runtime typically takes on the order of 30 seconds, depending on your computer configurations. A progress bar will appear to the left of the Estimate command button as shown in Figure 79. When the analysis is complete, you will see the frequency curve with credible intervals appear in **Frequency Results** plot located in the **Tabbed Documents** area.

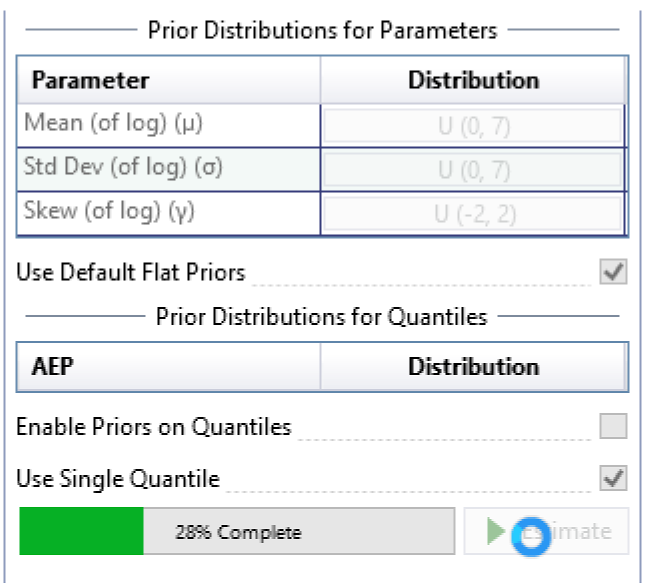


Figure 79 – Run the Bayesian Estimation Analysis.

Diagnose Convergence

After you have run the **Bayesian Estimation Analysis**, RMC-BestFit provides several useful plots for diagnosing the simulation convergence. The default simulation option (see Simulation Options) will typically ensure that you get reasonable results. However, there may be situations where you would like to adjust the simulation options to reduce runtimes while still achieving accurate results.

The **Bayesian Estimation Analysis** will open to the **Frequency Results** tab by default because this information is of most useful for a typical user. However, let's start from the bottom and work upward by first clicking on the **Markov Chain Traces** tab as shown in Figure 80.

Markov Chain Traces

Trace plots provide an important tool for assessing the mixing of a Markov chain. The trace plot is a time series plot of the Markov Chain iterations. The trace plot shows the evolution of a parameter vector over the iterations of one or all chains. You can select the distribution parameter and which chain(s) to view. In addition, you can include the warm up evolutions. Let's select the **Skew (of log)**(γ) parameter and **All Chains** and check the **Include Warm Up** checkbox, as shown in Figure 80.

We can see that for the first 100 or so evolutions, the sampler is warming up and has not yet converged. After that point, we see that the MCMC sampler seems to be mixing well because all of the chains are exploring the same region of the parameter space.

What you want to look for is any anomaly in the traces. Are there any chains significantly different from the others? Is there multimodality? In other words, do the traces jump from one modal region to another? These would be signs of poor mixing, and would indicate that you need to increase the number of evolutions.

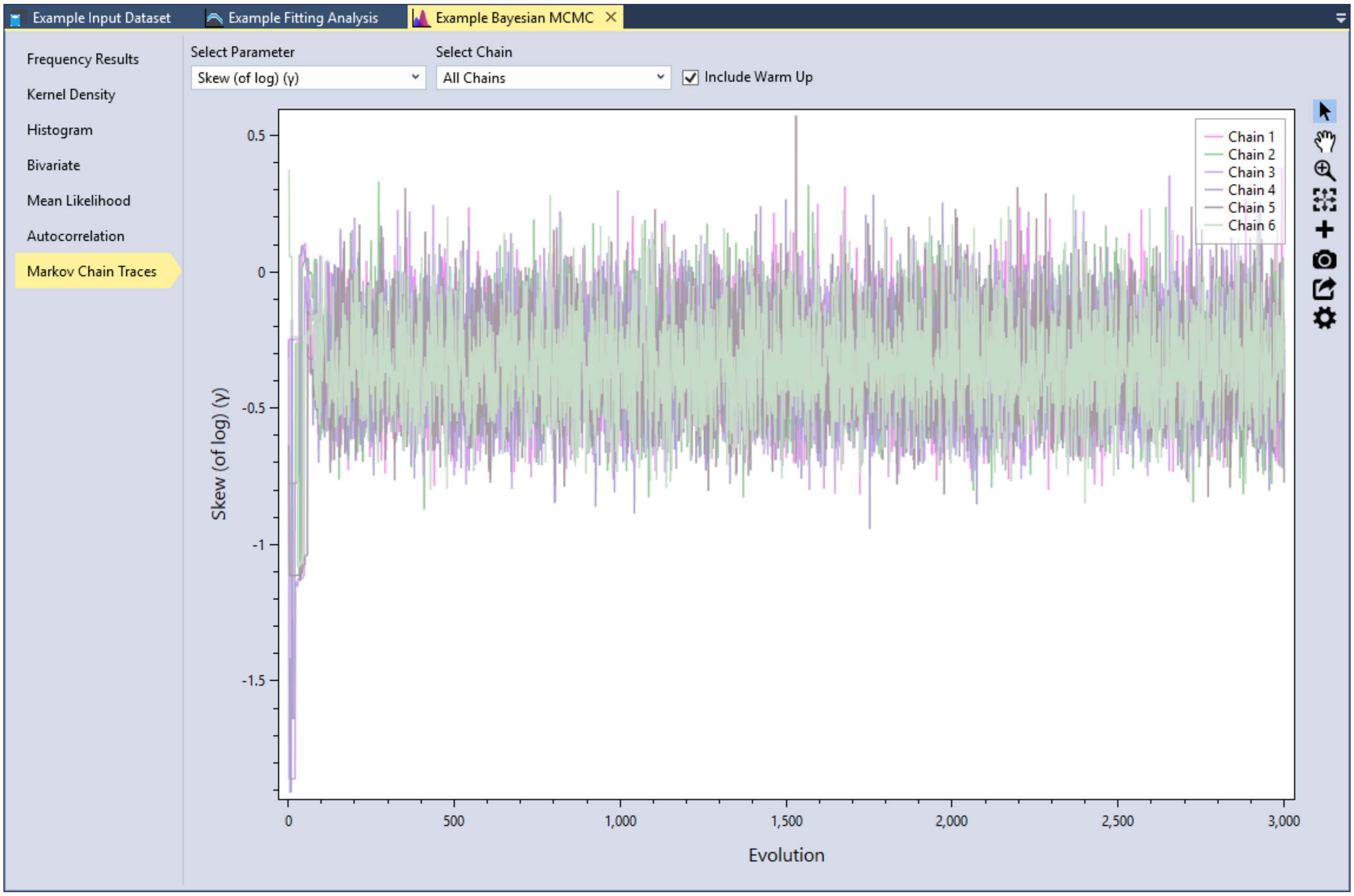


Figure 80 – Markov Chain Traces.

Autocorrelation

Another way to check for convergence is to look at the autocorrelations between the MCMC samples. MCMC samples are dependent, so there will be correlation among each consecutive iteration. This will not affect the validity of inference on the posterior samples so long as the sampler has time to fully explore the posterior distribution. However, autocorrelation will affect the efficiency of the sampler. In other words, highly correlated chains require more samples to produce the same level of precision for an estimate. Since autocorrelation tends to decrease as the lag increases, thinning samples will reduce the final autocorrelation in the sample while also reducing the total number of saved MCMC iterations required.

Select the Autocorrelation tab and select the **Skew (of log)**(γ) parameter as shown in Figure 81. This plot shows the lag-***k*** autocorrelation between every sample and the sample ***k*** steps before. The autocorrelation should decrease as ***k*** increases. When the autocorrelation is zero, the samples can be considered independent. Conversely, if the autocorrelation remains high for higher values of ***k***, then this indicates a high degree of correlation between samples and slow mixing.

The **Thinning Interval** is set to 20 by default in the simulation options. We can see that this ensures we get good mixing and near zero autocorrelation.

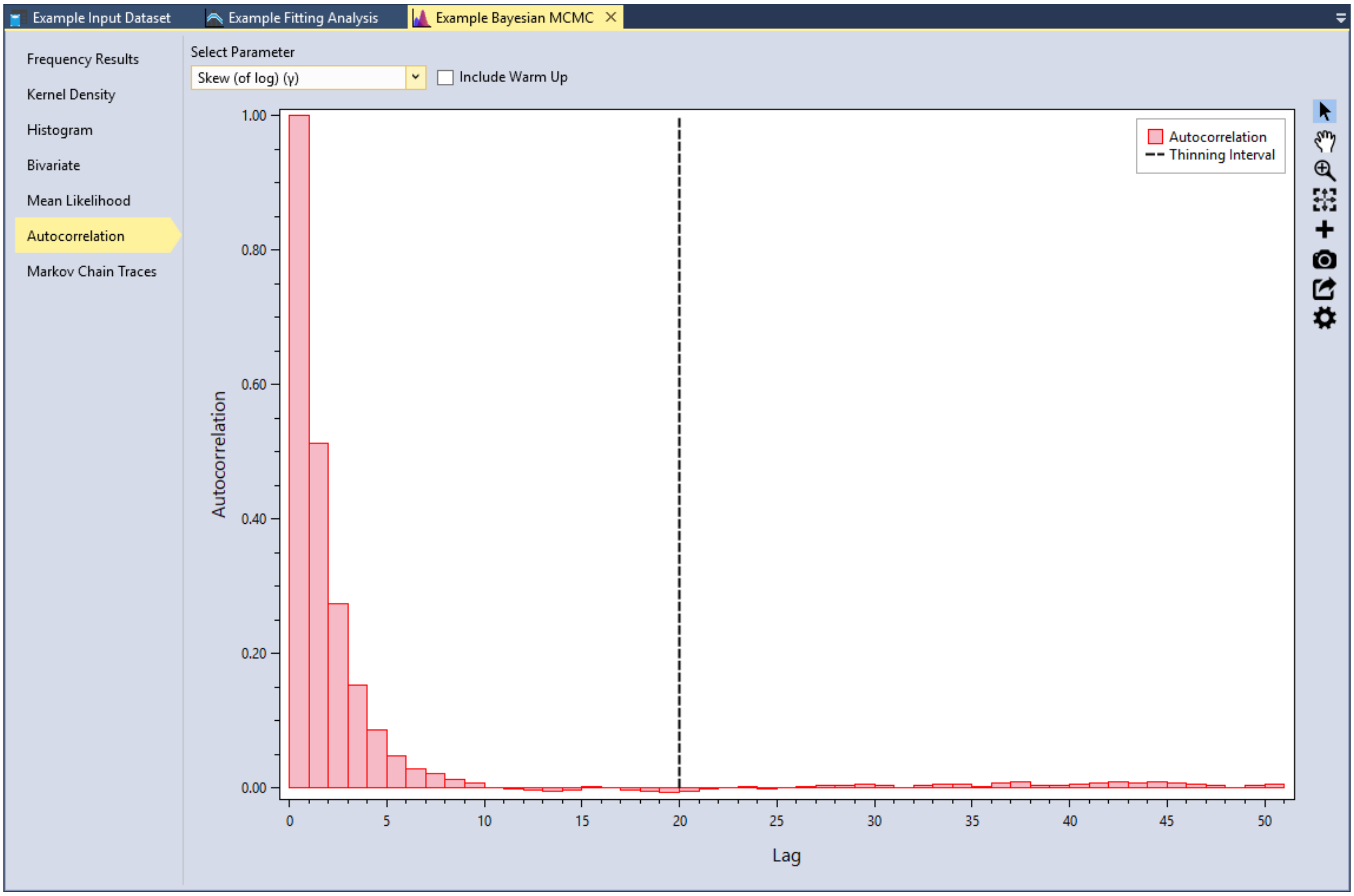


Figure 81 – Autocorrelation.

Mean Likelihood

Another way to check for convergence is to see if log-likelihood function is stable. Select the **Mean Likelihood** tab. We can see that during the first 100 or so evolutions, the sampler is warming up and has not yet converged. As we previously saw in the Markov Chain trace plot, all of the chains are exploring the same region of the parameter space, which leads to a very stable mean log-likelihood.

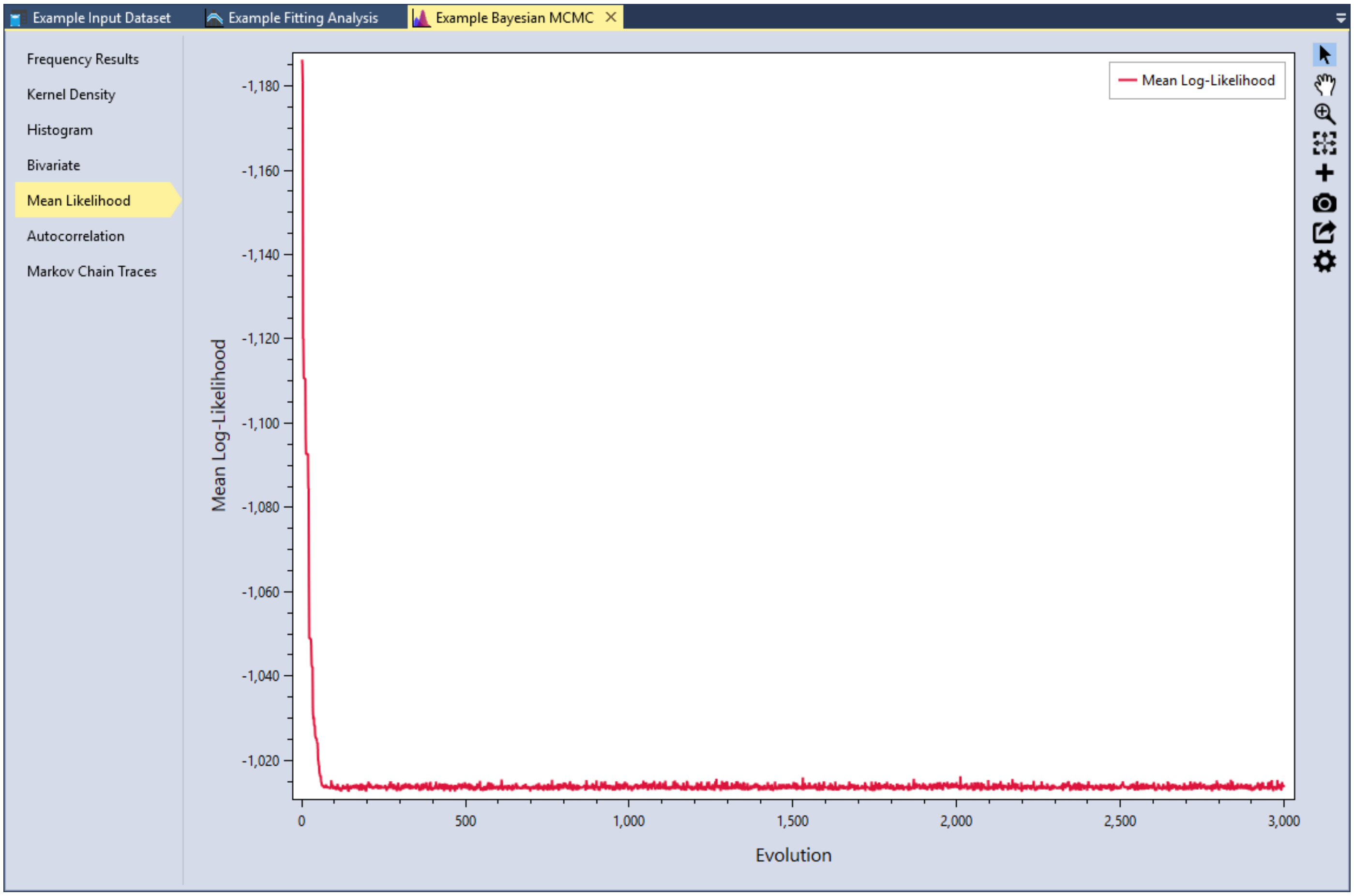


Figure 82 – Mean Likelihood.

It is recommended that you do not rely on a single diagnostic measure. It is important to use a weight-of-evidence approach and view many different diagnostics. After examining these plots, we can be confident that the MCMC simulation has converged. Now let's explore the results.

Explore Results

RMC-BestFit provides several tools for exploring the results of the Bayesian analysis. The **Bayesian Estimation Analysis** will open to the **Frequency Results** tab by default as shown in Figure 83.

Frequency Results

The posterior predictive distribution, posterior mode distribution, and the credible intervals will be plotted on the **Graphical Results** tab. By default, the posterior parent distribution is plotted as a frequency curve, with annual exceedance probabilities plotted on the X-axis using a Normal scale, and magnitude on the Y-axis using a logarithmic scale. You may edit the plot properties, flip the axes, or changing the axes scales as desired. RMC-BestFit will save and persist all of the changes you have made to the plot.

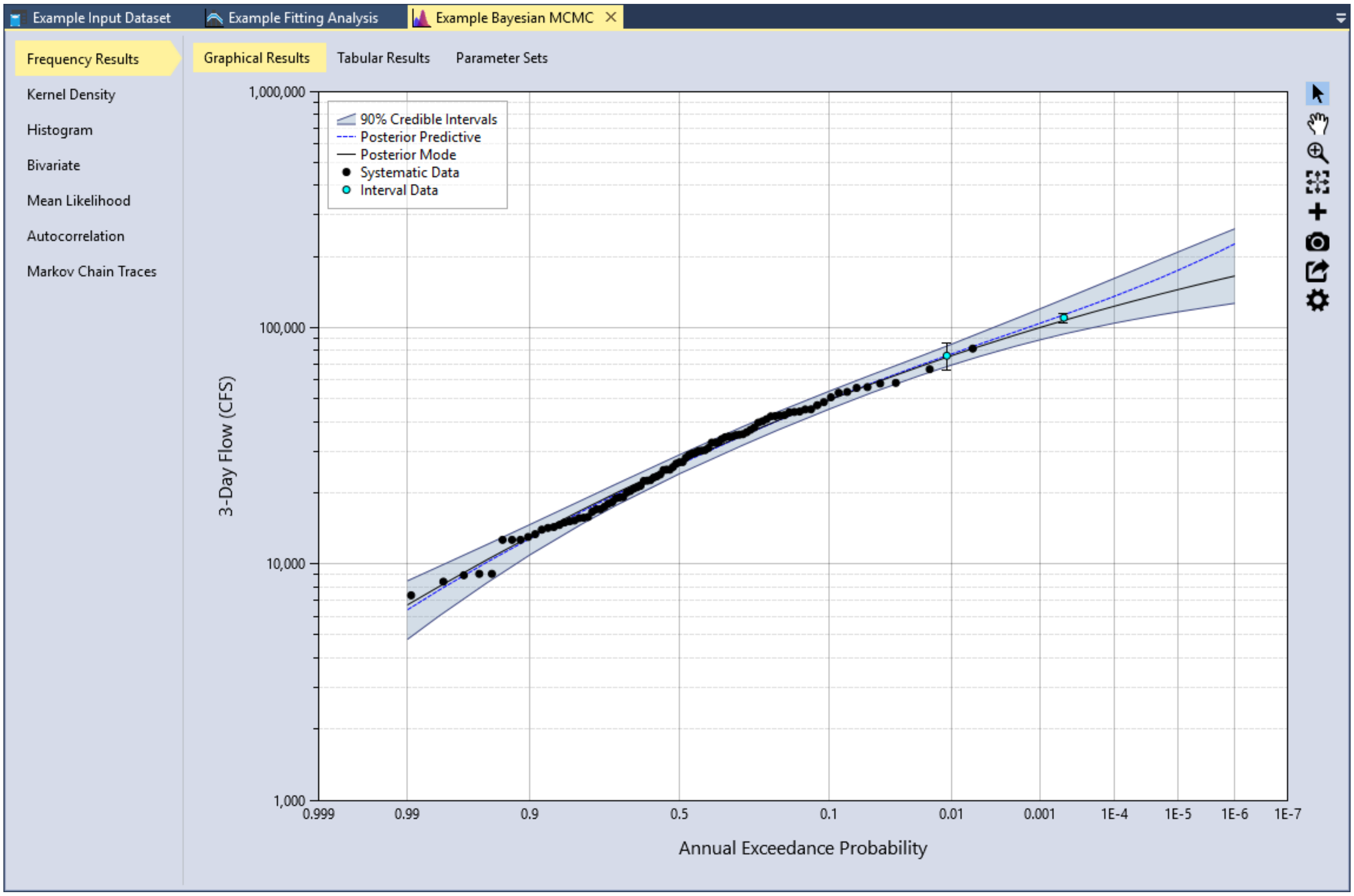
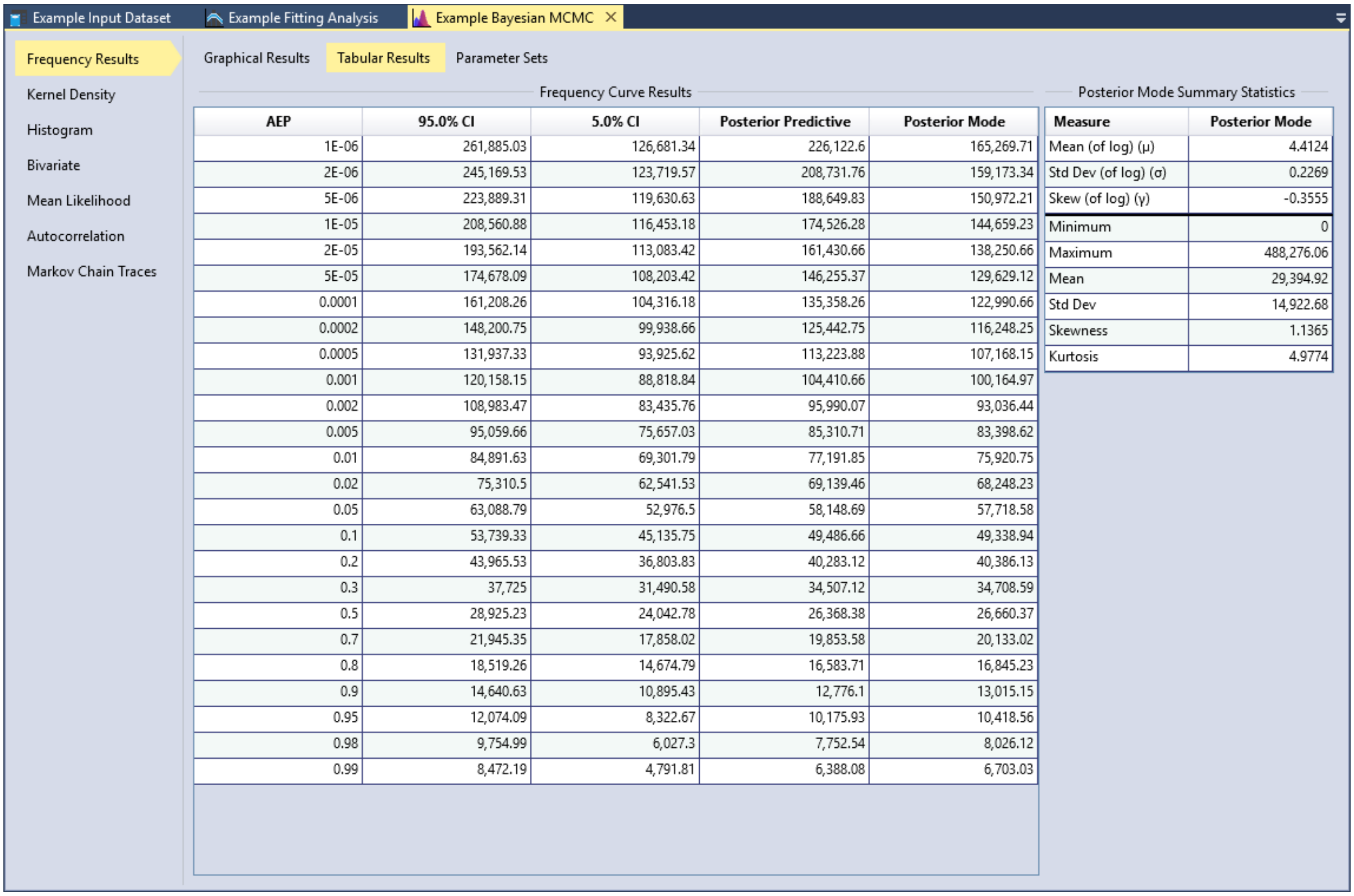


Figure 83 – Graphical Frequency Results.

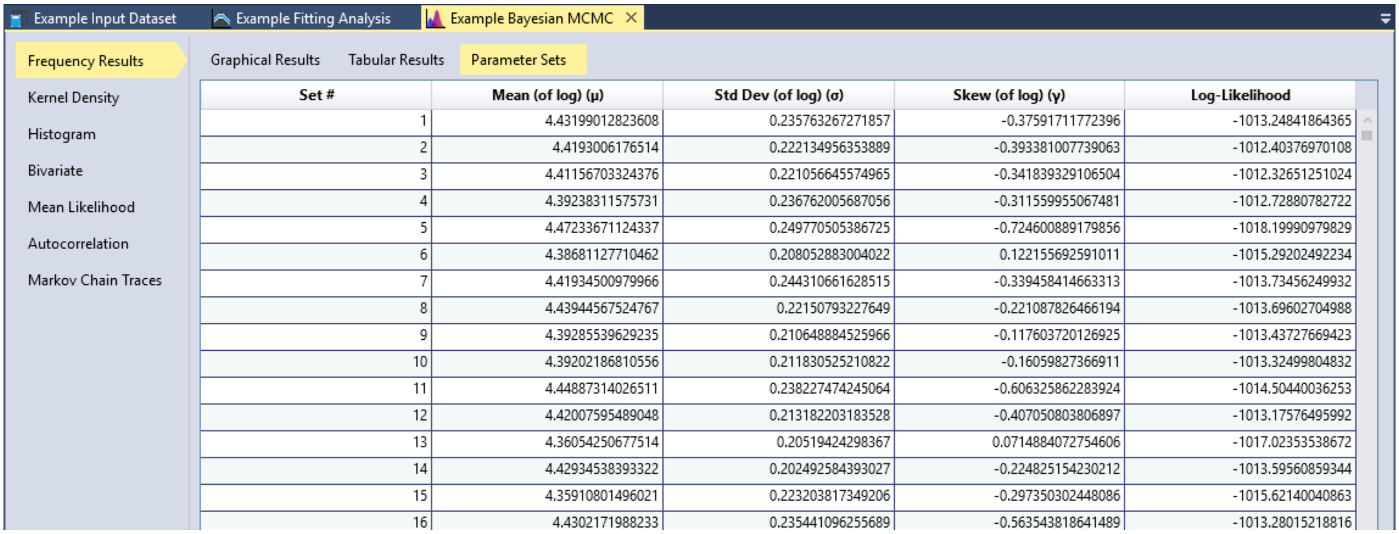
Click the **Tabular Results** tab to view the frequency curve results and the posterior mode summary statistics as shown in Figure 84. All of the outputted posterior parameter sets and associated log-likelihood values are available on the **Parameter Sets** tab as shown in Figure 85. You may sort these tables, and copy all of the values for external use.



AEP	95.0% CI	5.0% CI	Posterior Predictive	Posterior Mode
1E-06	261,885.03	126,681.34	226,122.6	165,269.71
2E-06	245,169.53	123,719.57	208,731.76	159,173.34
5E-06	223,889.31	119,630.63	188,649.83	150,972.21
1E-05	208,560.88	116,453.18	174,526.28	144,659.23
2E-05	193,562.14	113,083.42	161,430.66	138,250.66
5E-05	174,678.09	108,203.42	146,255.37	129,629.12
0.0001	161,208.26	104,316.18	135,358.26	122,990.66
0.0002	148,200.75	99,938.66	125,442.75	116,248.25
0.0005	131,937.33	93,925.62	113,223.88	107,168.15
0.001	120,158.15	88,818.84	104,410.66	100,164.97
0.002	108,983.47	83,435.76	95,990.07	93,036.44
0.005	95,059.66	75,657.03	85,310.71	83,398.62
0.01	84,891.63	69,301.79	77,191.85	75,920.75
0.02	75,310.5	62,541.53	69,139.46	68,248.23
0.05	63,088.79	52,976.5	58,148.69	57,718.58
0.1	53,739.33	45,135.75	49,486.66	49,338.94
0.2	43,965.53	36,803.83	40,283.12	40,386.13
0.3	37,725	31,490.58	34,507.12	34,708.59
0.5	28,925.23	24,042.78	26,368.38	26,660.37
0.7	21,945.35	17,858.02	19,853.58	20,133.02
0.8	18,519.26	14,674.79	16,583.71	16,845.23
0.9	14,640.63	10,895.43	12,776.1	13,015.15
0.95	12,074.09	8,322.67	10,175.93	10,418.56
0.98	9,754.99	6,027.3	7,752.54	8,026.12
0.99	8,472.19	4,791.81	6,388.08	6,703.03

Measure	Posterior Mode
Mean (of log) (μ)	4.4124
Std Dev (of log) (σ)	0.2269
Skew (of log) (γ)	-0.3555
Minimum	0
Maximum	488,276.06
Mean	29,394.92
Std Dev	14,922.68
Skewness	1.1365
Kurtosis	4.9774

Figure 84 – Tabular Frequency Results.



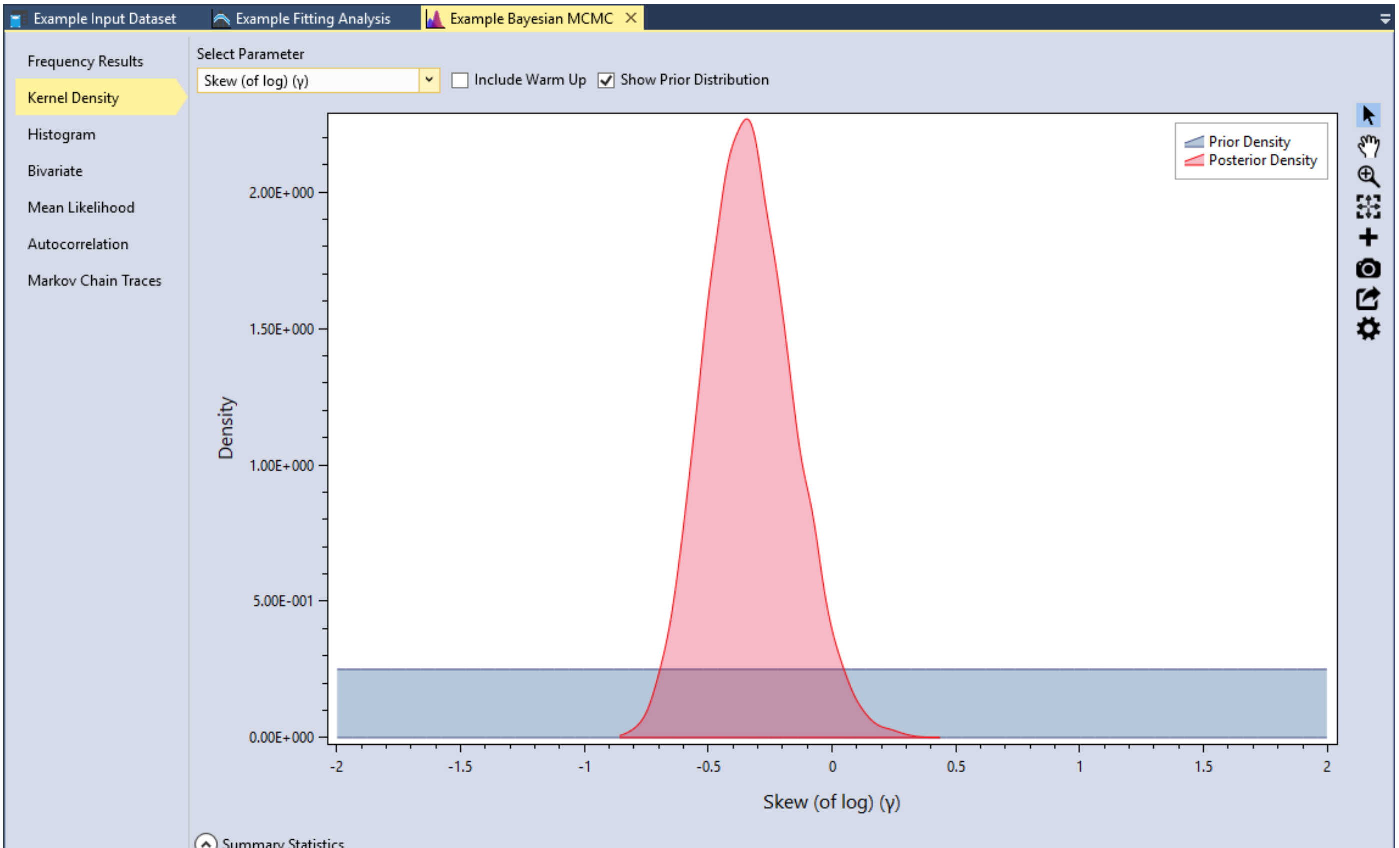
Set #	Mean (of log) (μ)	Std Dev (of log) (σ)	Skew (of log) (γ)	Log-Likelihood
1	4.43199012823608	0.235763267271857	-0.37591711772396	-1013.24841864365
2	4.4193006176514	0.222134956353889	-0.393381007739063	-1012.40376970108
3	4.41156703324376	0.221056645574965	-0.341839329106504	-1012.32651251024
4	4.39238311575731	0.236762005687056	-0.311559955067481	-1012.72880782722
5	4.47233671124337	0.249770505386725	-0.724600889179856	-1018.19990979829
6	4.38681127710462	0.208052883004022	0.122155692591011	-1015.29202492234
7	4.41934500979966	0.244310661628515	-0.339458414663313	-1013.73456249932
8	4.43944567524767	0.22150793227649	-0.221087826466194	-1013.69602704988
9	4.39285539629235	0.210648884525966	-0.117603720126925	-1013.43727669423
10	4.39202186810556	0.211830525210822	-0.16059827366911	-1013.32499804832
11	4.44887314026511	0.238227474245064	-0.606325862283924	-1014.50440036253
12	4.42007595489048	0.213182203183528	-0.407050803806897	-1013.17576495992
13	4.36054250677514	0.20519424298367	0.0714884072754606	-1017.02353538672
14	4.42934538393322	0.202492584393027	-0.224825154230212	-1013.59560859344
15	4.35910801496021	0.223203817349206	-0.297350302448086	-1015.62140040863
16	4.4302171988233	0.235441096255689	-0.563543818641489	-1013.28015218816

Figure 85 – Parameter Sets.

Kernel Density

RMC-BestFit provides kernel density plots of the marginal posterior distribution of parameters. Kernel density estimates are closely related to histograms, but provide smoothness and continuity. Select the **Kernel Density** tab, select the **Skew (of log)**(γ) parameter, and check the **Show Prior Distribution** checkbox, as shown in Figure 86. Recall that the prior for skew was set as a uniform distribution $U(-2, +2)$.

Underneath the plot, you can expand the **Summary Statistics** for the marginal distribution. Here we can see that the mean of the skew (of log) parameter is -0.3337 with a standard deviation of 0.1756. You will also notice that there is a statistic labeled **Rhat**. This is often referred to as the Gelman-Rubin statistic, which assesses the mixing of the Markov chains using the between- and within-chain variances (Gelman, et al., 2014). A Rhat equal to 1.0 indicates that the posterior distribution has likely converged.



Rhat	Mean	Std Dev	5.0%	50%	95.0%
1.001	-0.3337	0.1756	-0.6084	-0.3416	-0.0339

Figure 86 – Kernel Density.

Histogram

RMC-BestFit also provides histograms of the marginal posterior distribution of parameters. These plots provide the same information as the kernel density plots. The histogram bins are set using the Rice rule $k = 2\sqrt[3]{n}$, where k is the number of bins, and n is the number of posterior parameter sets.

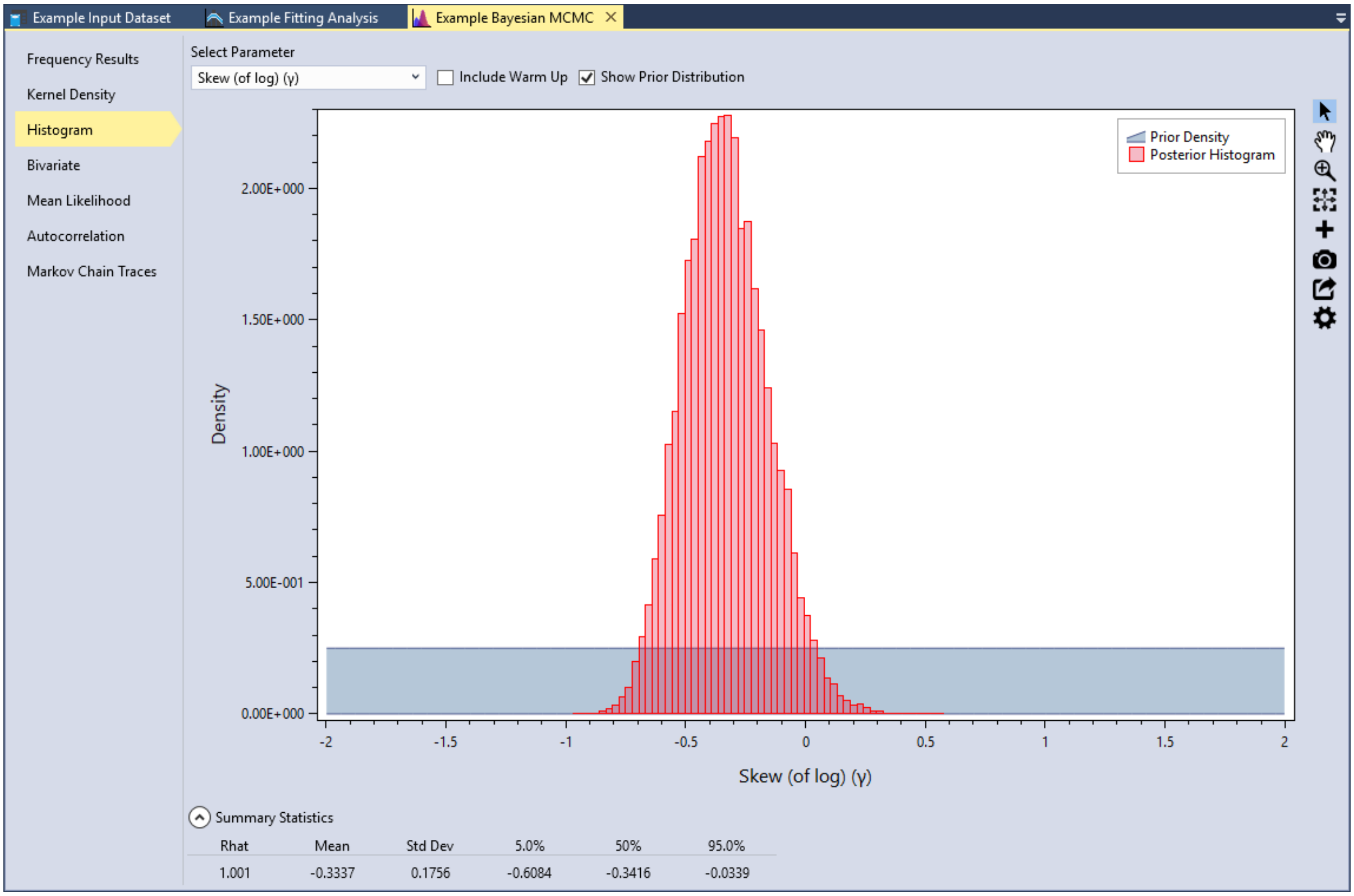


Figure 87 – Histogram.

Bivariate

RMC-BestFit provides the option to view the marginal posterior density functions for any pair of parameters as a two-dimensional heat map, with the color red indicating the highest density and blue indicated lowest density. Bivariate plots illustrate the dependence among the parent distribution parameters.

Typically, the higher order parameters are dependent on the lower order parameters. Set the **X Parameter** to be the **Std Dev (of log)** (σ) and the **Y Parameter** to be **Skew (of log)**(γ), as shown in Figure 88**.** We can see that standard deviation is negatively correlated with skew. Smaller standard deviations result in higher skews; whereas higher standard deviations results in lower skews. This tradeoff in parameters is typical with the LPIII distribution.

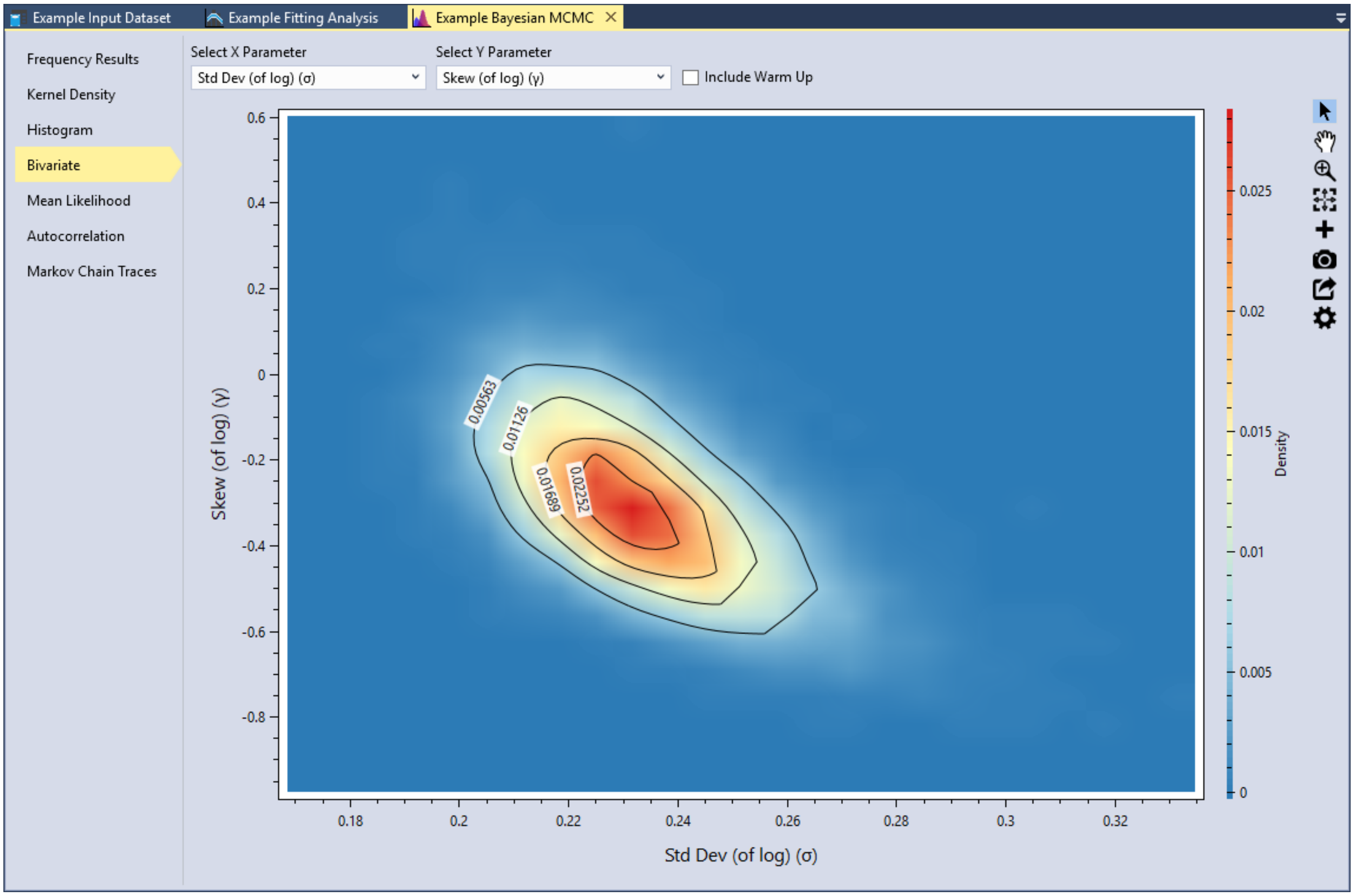


Figure 88 – Bivariate.

Informative Priors

An informative prior provides specific, scientific information about the parameter. Prior information can be obtained from regional analysis, causal modeling, or expert elicitation. In flood frequency, an example of an informative prior would be the use of a regional skew (Kuczera G. , 1983) for the LPIII distribution as described in Bulletin 17B (U.S. Geological Survey, 1982) and 17C (U.S. Geological Survey, 2018).

For the Blakely Mountain Dam, regional skew information was obtained from a USGS regional study of Arkansas, Oklahoma, and Louisiana (Wagner, Krieger, & Veilleux, 2016). From the USGS study, the regional skew was determined to be -0.17 with a mean-square error (MSE) of 0.12. This information can be incorporated into the Bayesian analysis by setting the prior for the skew parameter of LPIII to be a Normal distribution with a mean of -0.17 and standard deviation of 0.35, or $\sqrt{0.12}$.

First, uncheck the **Use Default Flat Priors** checkbox located on the **General** tab of the **Properties Window**. Then, click the distribution button for the skew (of log) parameter. A distribution selector will pop open to the left of the button, as shown in Figure 89.

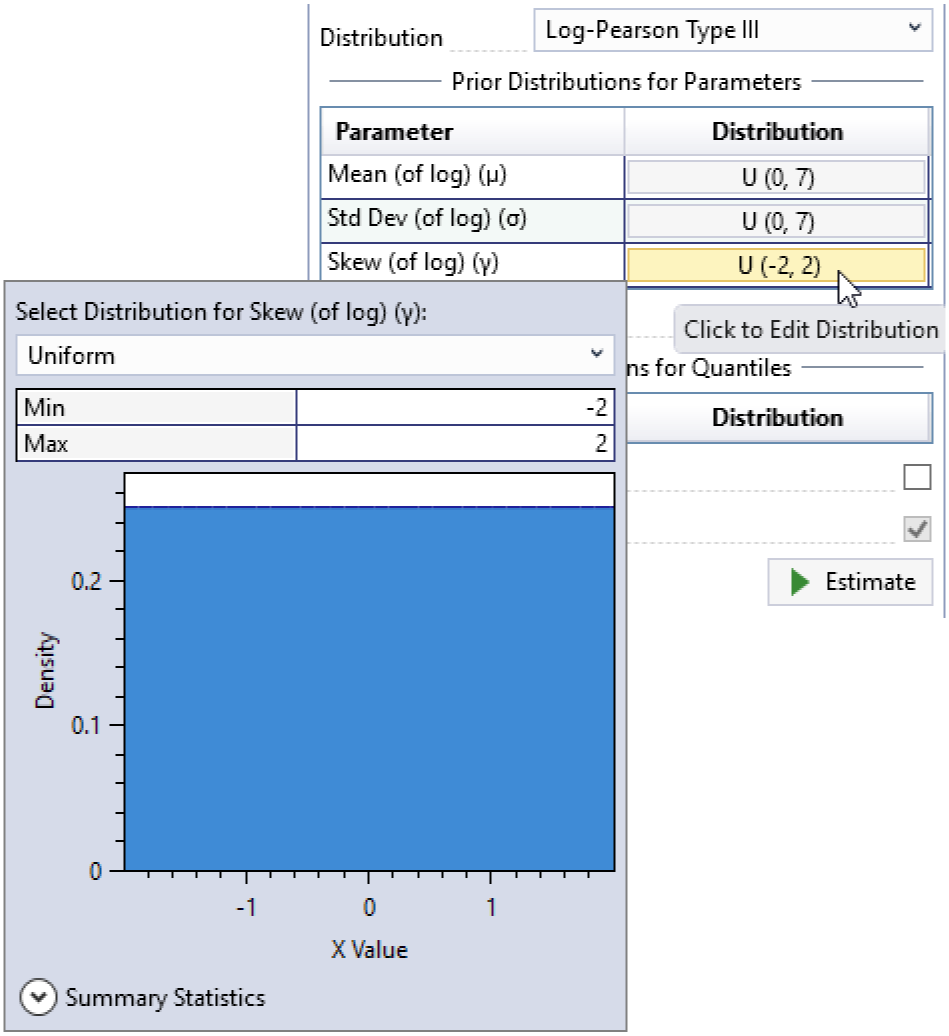


Figure 89 – Informative Prior on Skew.

Next, select the Normal distribution and set the mean to be -0.17 and the standard deviation to be 0.35 as shown below.

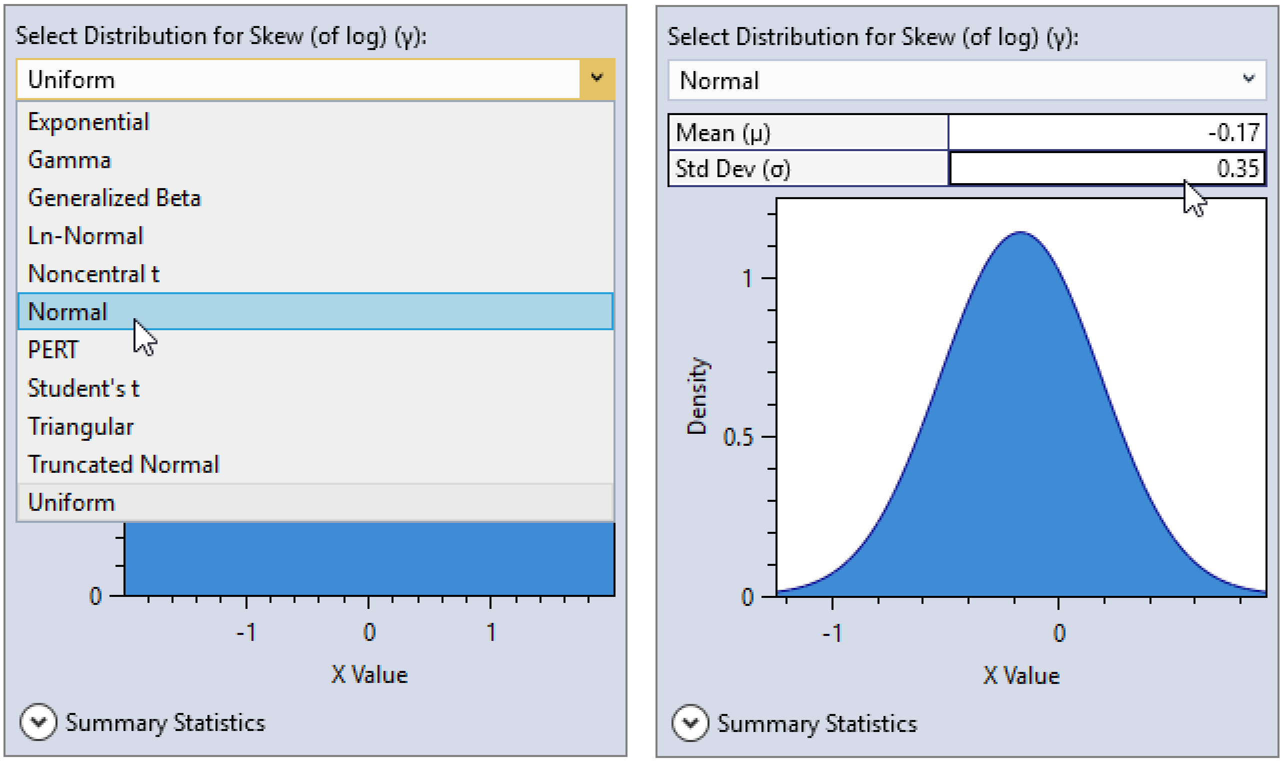


Figure 90 – Select Distribution for Skew.

Click away from the distribution selector popup and it will automatically close. You will now see that the prior distribution for skew has been set as $N(-0.17, 0.35)$. Now, click the **Estimate** command button to perform the Bayesian analysis using the informative prior.

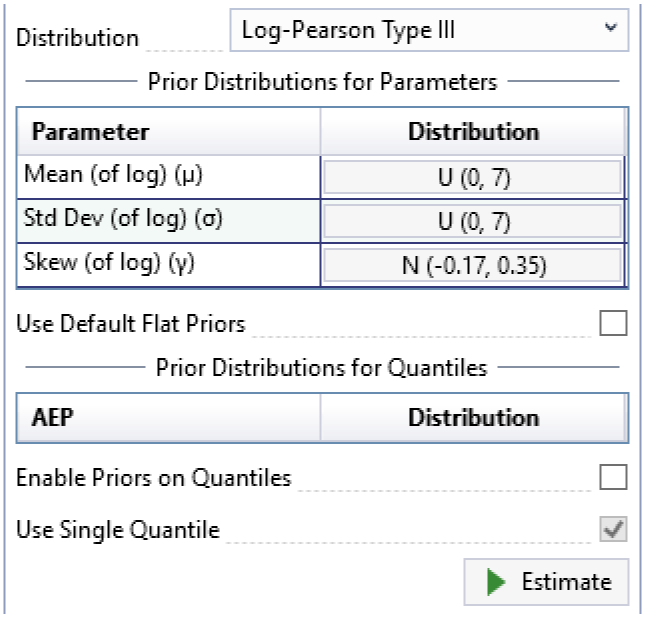


Figure 91 – Prior Distributions for Parameters.

When the simulation is complete, click on the Kernel Density tab and select the **Skew (of log)**(γ) parameter. When we used the default flat prior, the mean of the skew (of log) was -0.3337 with a standard deviation of 0.1756. Now, with the informative prior, we can see that the mean of the skew (of log) is -0.3062 with a standard deviation of 0.1616. The inclusion of the regional skew information has made the skew parameter less negative and reduced the variance.

As the sample size increases, the influence of the prior distribution on posterior inferences will decrease because the data likelihood will dominate. Taking this into account, regional prior information is most valuable when the at-site sample sizes are small relative to the effective sample size of the regional information.

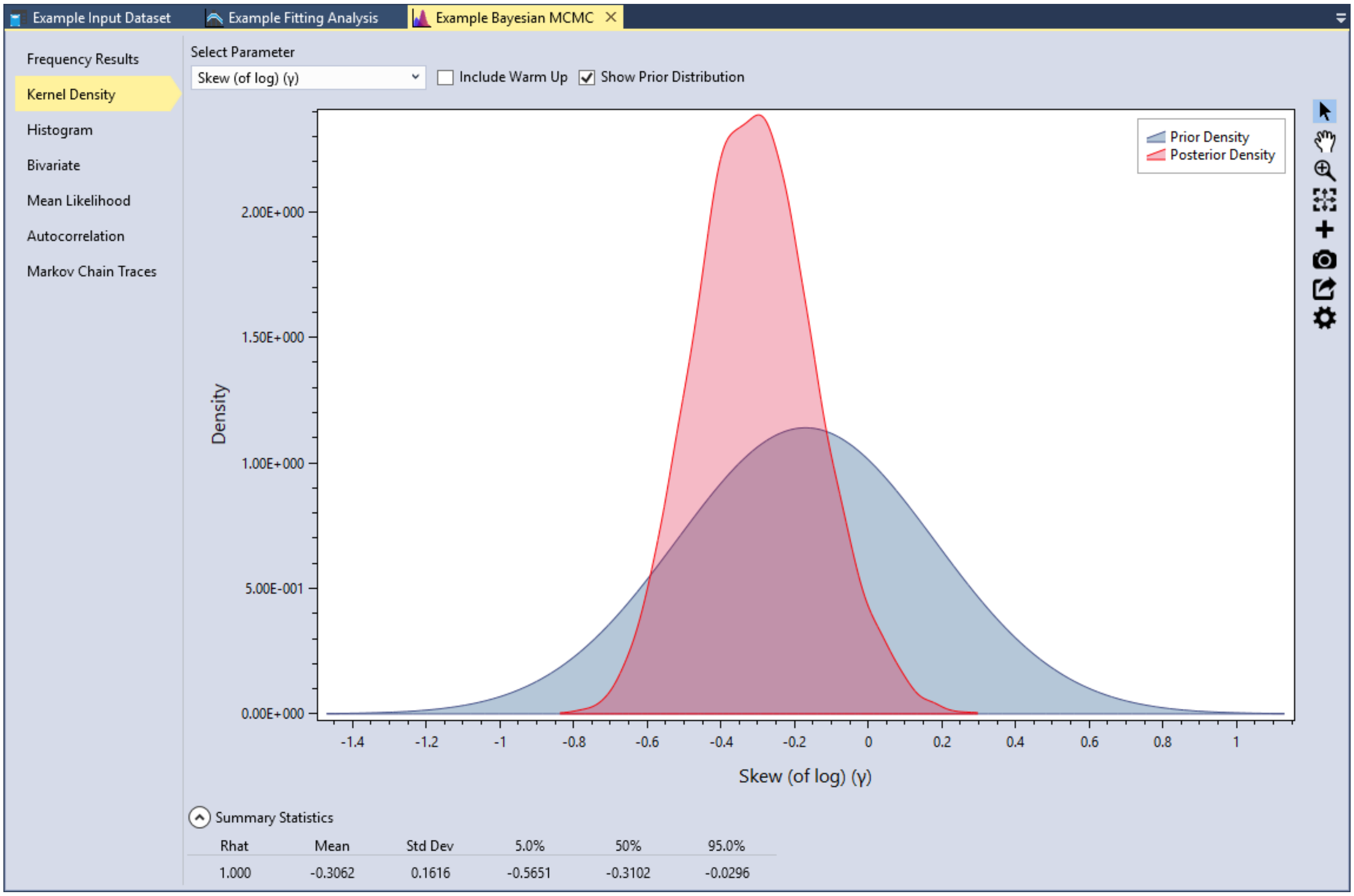


Figure 92 – View Results of Using an Informative Prior.

Priors on Quantiles

Modeled rainfall-runoff results can be incorporated into the Bayesian analysis by defining a prior distribution for a flood quantile. This approach is referred to as causal information expansion, which analyzes the generating mechanisms of floods in the catchment of interest (Merz & Bloschl, 2008).

A regional precipitation-frequency analysis was performed for the Blakely Mountain watershed, and 3-day basin-average rainfall-frequency events were routed using a calibrated HEC-HMS model. The main benefit of modeling regional rainfall-frequency information is that available rainfall records are often much longer than the at-site flood records. Therefore, the regional information combined with causal rainfall-runoff modeling can provide important prior information on the flood quantiles.

Use Single Quantile

Prior distributions for flood quantiles can be set in one of two ways. First, check the **Enable Priors on Quantiles** checkbox located on the **General** tab of the **Properties Window**. By default, the **Use Single Quantile** checkbox will also be checked. For more details on this single quantile approach, please refer to (Viglione, Merz, Salinas, & Bloschl, 2013) and (Skahill, Viglione, & Byrd, 2016).

From the analysis performed with HEC-HMS, the distribution of rainfall-runoff at the 1E-4 AEP was determined to be Normally distributed with a mean of 155,000 cfs and a standard deviation of 22,000 cfs. This rare AEP was selected because in order to add information to the fit, it needed to be rarer than the paleoflood event, while not being so rare as to be overly influential.

Click the distribution button and a distribution selector will pop open to the left of the button as shown in Figure 93. The Normal distribution is the only option available for priors on quantiles. Next, set the mean to be 155,000 and the standard deviation to be 22,000. Click away from the distribution selector popup and it will automatically close. Now, select the 0.0001 AEP as shown below in Figure 93.

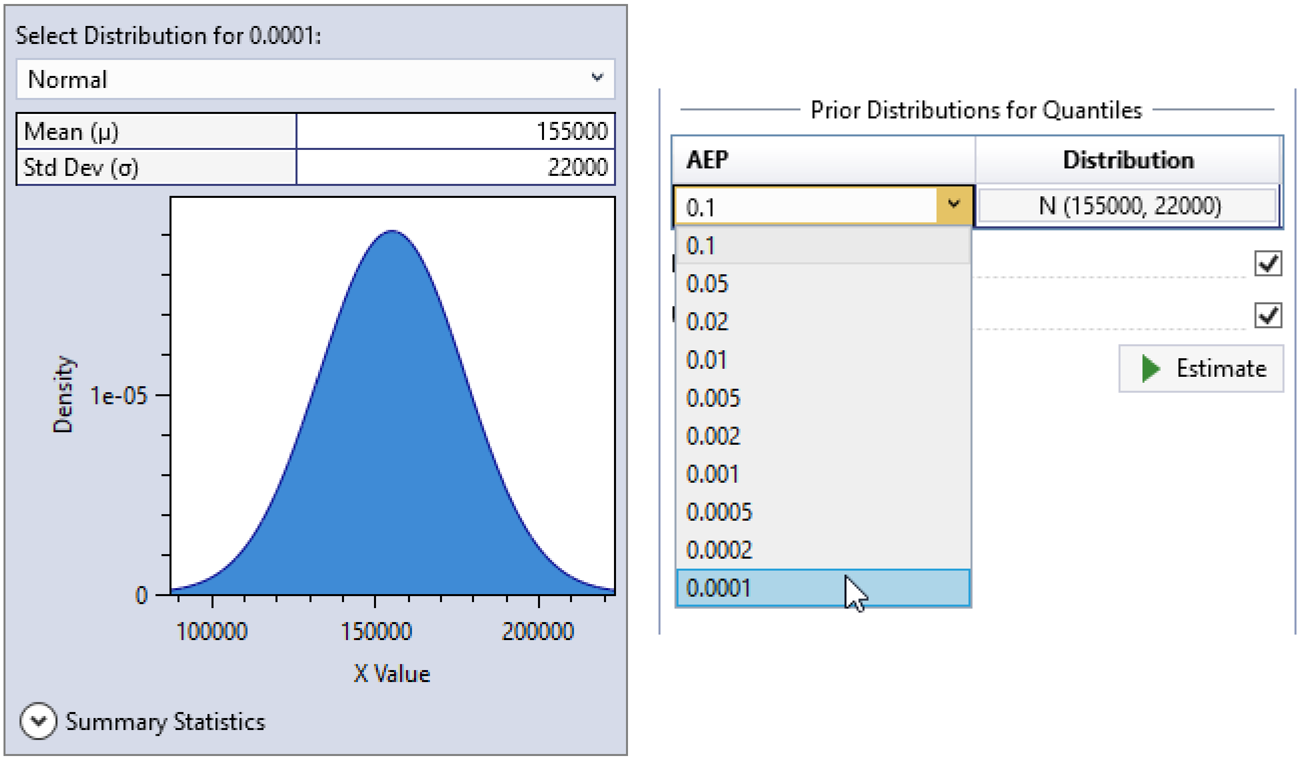


Figure 93 – Select Distribution for a Quantile.

You will now see that the prior distribution for the quantile has been set to $N(155000,\ 22000)$. Now, click the **Estimate** command button to perform the Bayesian analysis using the informative quantile prior. When the analysis is complete, you will see the frequency curve with credible intervals appear in the **Frequency Results** plot located in the **Tabbed Documents** area. The mean of the quantile prior will be displayed as a green square, and the 5th and 95th percentiles of the prior will be shown as a vertical error bar, as shown in Figure 95.

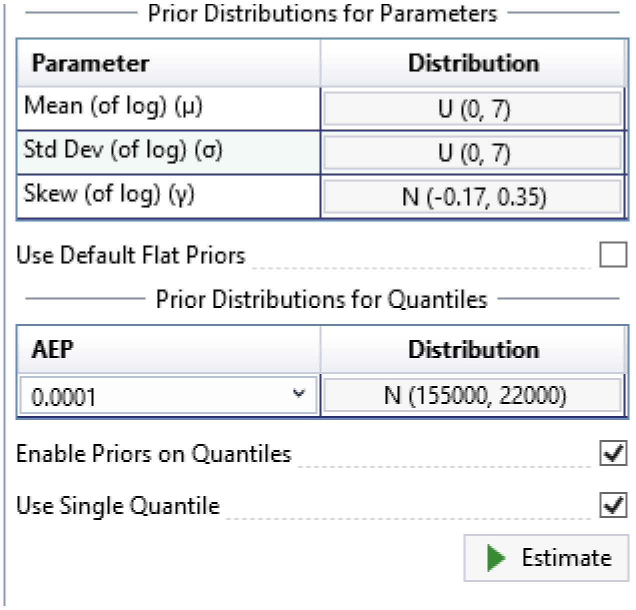


Figure 94 – Prior Distributions for a Single Quantile.

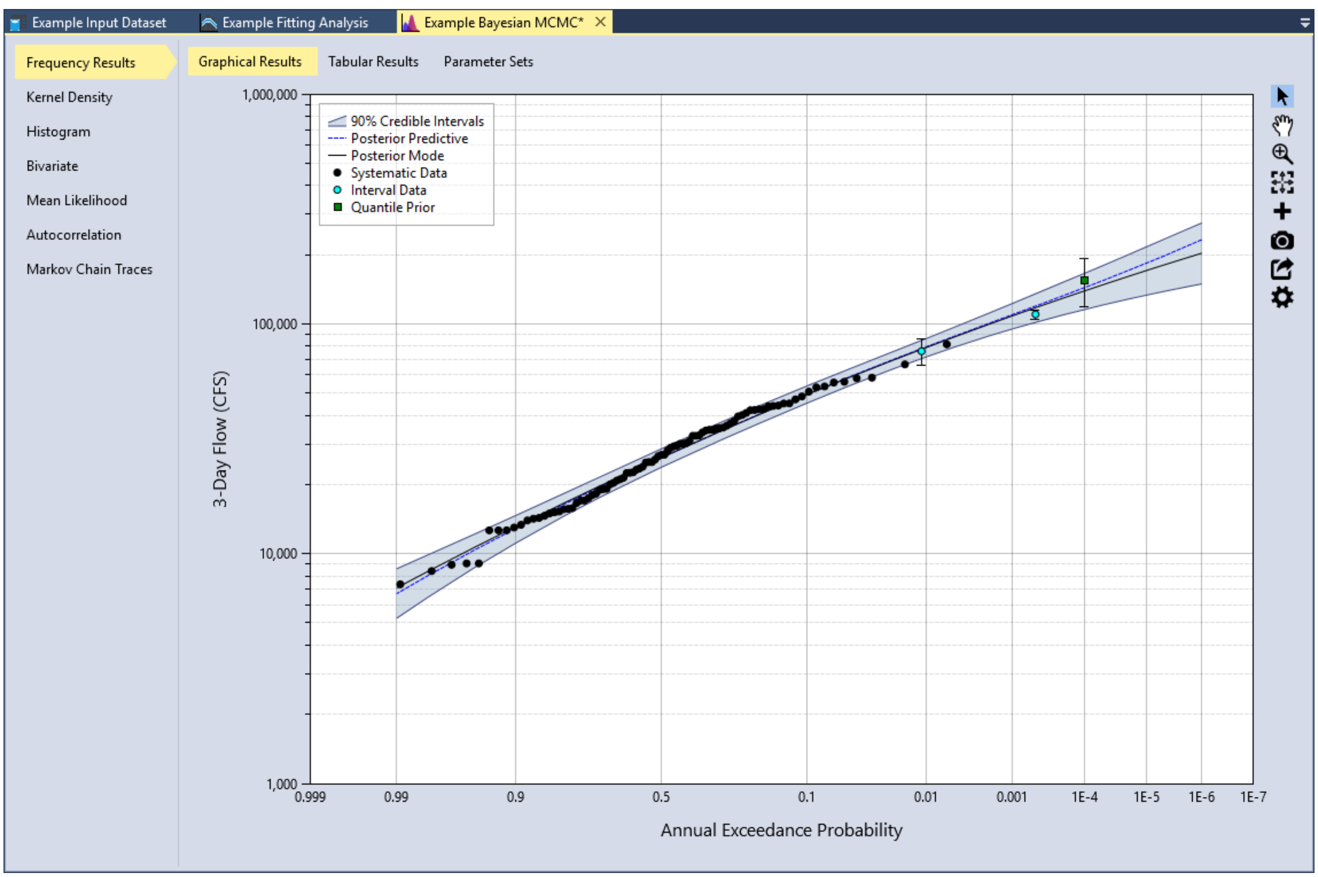


Figure 95 – Frequency Results Using a Prior Distribution for a Single Quantile.

If we click on the Kernel Density tab and select the **Skew (of log)**(γ) parameter, we can see that the mean of the skew (of log) is -0.2384 with a standard deviation of 0.1465. The inclusion of the causal rainfall-runoff information has made the skew parameter significantly less negative, reduced the variance, and reduced the width of the resulting credible intervals of the **Frequency Results**.

Summary Statistics

Rhat	Mean	Std Dev	5.0%	50%	95.0%
1.000	-0.2384	0.1465	-0.4801	-0.2389	0.0013

Figure 96 – Summary Statistics Results for the Skew Parameter after Using an Informative Prior on a Single Quantile.

Use Multiple Quantiles

The final way to set prior distributions for flood quantiles is to uncheck the **Use Single Quantile** checkbox. This option for setting priors on distribution quantiles follows the approach used in (Coles & Tawn, 1996). The number of quantile priors must be equal to the number of distribution parameters. For example, for the LPIII distribution there must be three quantile priors.

In the single quantile approach shown above, the choice of AEP for the prior information has a large influence on the resulting fit and credible intervals. If a more frequent quantile is chosen, such as 1E-2, more weight would be given to the historical and paleoflood data and the quantile prior would have less influence on the fit. Whereas, the choice of the 1E-4 quantile gives significant weight to the prior. Choosing a rare quantile implies that we have high confidence in the regional precipitation-frequency analysis and modeled rainfall-runoff results. In general, the rarer the chosen quantile for the prior information, the more influence it will have on the posterior results.

As a general rule of thumb, if you are using NOAA Atlas 14[1] (A14) precipitation-frequency data with a three parameter distribution, such as LPIII, then you should enter quantile priors for 1E-1, 1E-2, and 1E-3. However, if you have performed a custom regional precipitation-frequency analysis that is believed to be of higher quality than A14, you should enter priors for 1E-2, 1E-3, and 1E-4.

In the case of Blakely Mountain Dam, a custom regional analysis was performed for the nearby Trinity River Basin in Texas, which is described in (MetStat, Inc., 2018). The regional frequency analysis performed for the Trinity River Basin also included the geographical region where the Blakely Mountain watershed is located in Arkansas. This study incorporated advanced techniques, such as storm typing; therefore, the results are considered to provide a better extrapolation to rare exceedance probabilities than A14.

Using the results from the HEC-HMS analysis, the distribution of rainfall-runoff for the 1E-2, 1E-3, and 1E-4 AEP quantiles are shown in Figure 97. After you have entered these prior distributions, click the **Estimate** command button to perform the Bayesian analysis. The **Frequency Results** plot is shown below in Figure 98.

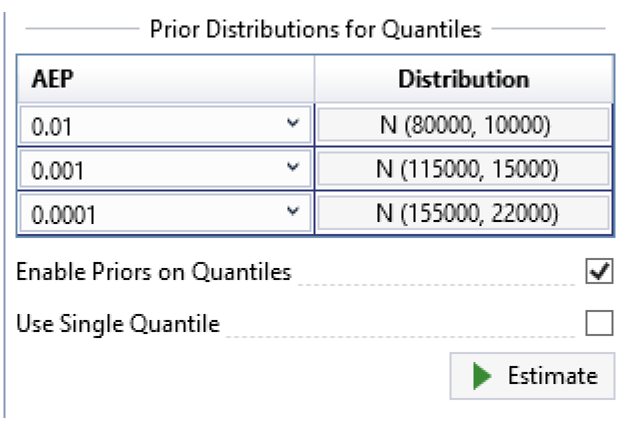


Figure 97 – Prior Distributions for Multiple Quantiles.

[1] https://hdsc.nws.noaa.gov/hdsc/pfds/pfds_map_cont.html

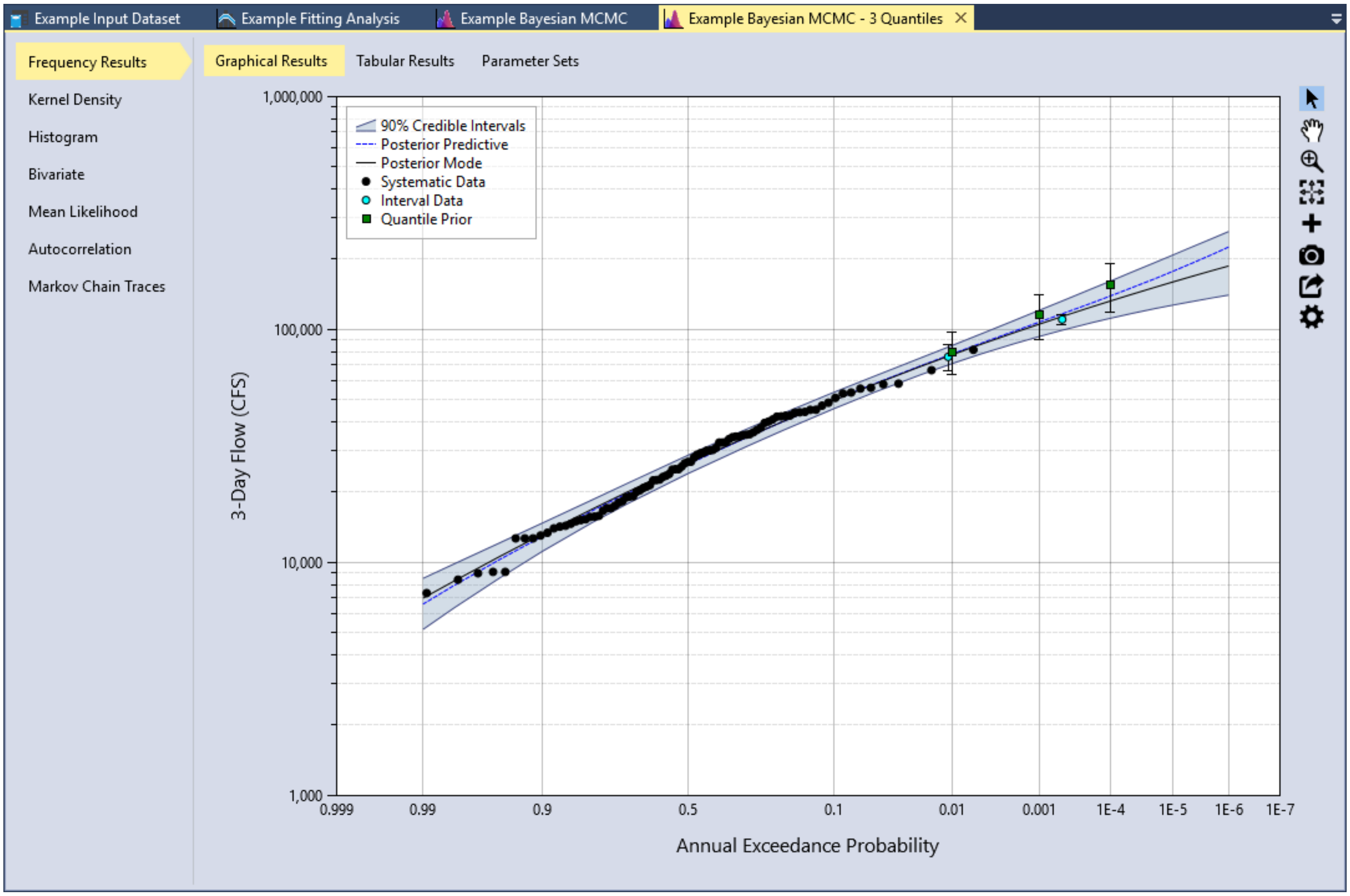


Figure 98 – Frequency Results Using a Prior Distribution for Multiple Quantiles.

If we click on the Kernel Density tab and select the **Skew (of log)**(γ) parameter, we can see that the mean of the skew (of log) is -0.2779 with a standard deviation of 0.1500. The use of multiple quantiles has still made the skew parameter less negative and reduced the variance. However, the effect is less noticeable as compared to what we saw with the single quantile option. With this in mind, the single quantile option has a potential to underestimate the true parameter and quantile variance. The multiple quantile prior method provides a more complete treatment of the priors, and is considered a better choice when the data is available.

Summary Statistics

Rhat	Mean	Std Dev	5.0%	50%	95.0%
1.000	-0.2779	0.1500	-0.5194	-0.2799	-0.0297

Figure 99 – Summary Statistics Results for the Skew Parameter after Using an Informative Prior on Multiple Quantiles.

As we saw in the regional skew example, when the at-site sample size increases, the influence of the prior distribution on posterior will decrease because the data likelihood will dominate. Taking this into account, prior information on quantiles is most valuable when the at-site sample sizes are small relative to the effective sample size of the regional precipitation-frequency and causal rainfall-runoff information. For further information on setting informative priors for quantiles, please see (Coles & Tawn, 1996), (Smith, 2005), (Viglione, Merz, Salinas, & Bloschl, 2013), and (Skahill, Viglione, & Byrd, 2016).

You have now finished the example application with RMC-BestFit. Save the project by selecting **File** > **Save**, or by clicking the **Save** button on the main window **Tool Bar**.

References

Asquith, W. H. (2011). *Distributional Analysis with L-moment Statistics using the R Environment for Statistical Computing.* Lubbock, TX: Create Space Independent Publishing Platform.

Beard, L. R. (1960, July). Probability Estimates Based on Small Normal-Distribution Samples. *Journal of Geophysical Research*.

Bobee, B., & Ashkar, F. (1991). *The Gamma Family and Derived Distributions Applied in Hydrology.* Littleton, CO: Water Resources Publications.

Cohn, T. A., England, J. F., Berenbrock, C. E., Mason, R. R., Stedinger, J. R., & Lamontagne, J. R. (2013). A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. *Water Resources Research, 49*(8), 5047-5058.

Cohn, T. A., Lane, W. M., & Baier, W. G. (1997). An Algorithm for Computing Moments-Based Flood Quantile Estimates When Historical Flood Information is Available. *Water Resources Research*, 2089-2096.

Coles, S. G., & Tawn, J. A. (1996). A Bayesian Analysis of Extreme Rainfall Data. *Journal of the Royal Statistical Society*, 463-478.

Efron, B., & Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science.* New York, NY: Cmbridge University Press.

Fisher, R. A., & Cornish, E. A. (1960). The percentile points of distributions having known cumulants. *Technometrics, 2*(2), 209-225.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (Third ed.). Boca Raton, FL: CRC Press.

Haan, C. T. (1977). *Statistical Methods in Hydrology.* Ames: The Iowa State University Press.

Hirsch, R. M., & Stedinger, J. R. (1987). Plotting Positions for Historical Floods and Their Precision. *Water Resources Research*, 715-727.

Hosking, J. R., & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge, UK: Cambridge University Press.

Jongejan, R. (2018). *Uncertainty in Hydrology: An evaluation of USACE methodology for estimating and portraying hydrologic uncertainty.*

Krishnamoorthy, K. (2016). *Handbook of Statistical Distributions with Applications.* Boca Raton, FL: CRC Press.

Kuczera, G. (1983). A Bayesian Surrogate for Regional Skew in Flood Frequency Analysis. *Water Resources Research, 19*, 821-832.

Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resources Research*, 1551-1557.

Maidment, D. R. (1992). *Handbook of Hydrology.* New York: McGraw-Hill.

Merz, R., & Bloschl, G. (2008). Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*.

MetStat, Inc. (2018). *Trinity River Hydrologic Hazards Project - Regional Extreme Precipitation-Frequency Analysis for the Trinity River Basin.* Fort Collins, CO.

Meylan, P. (2012). *Predictive Hydrology: A Frequency Analysis Approach.* Enfield, New Hampshire: Science Publishers.

O'Connell, D. R., Ostenaa, D. A., Levish, D. R., & Klinger, R. E. (2002). Bayesian flood frequency analysis with paleohydrologic bound data. *Water Resources Research, 38*(5). Retrieved from doi:10.1029/2000WR000028

Perreault, L. (2000). *Bayesian retrospective analysis of a break in sequences of hydrologic random variables.* Paris: ENGREF.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2017). *Numerical Recipes: The Art of Scientific Computing.* (3rd ed.). Cambridge, UK: Cambridge University Press.

Rao, A., & Hamed, K. H. (2000). *Flood Frequency Analysis.* Boca Raton, FL: CRC Press LLC.

Reis, D. S., & Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, 97-116.

Skahill, B. E., Viglione, A., & Byrd, A. (2016). *ERDC/CHL CHETN-X-1 A Bayesian Analysis of the Flood Frequency Concept.* Vicksburg, MS: U.S. Army Engineer Research and Development Center, [Coastal and Hydraulics Laboratory].

Smith, E. (2005). *Bayesian Modelling of Extreme Rainfall Data.* Newcastle, NSW: University of Newcastle.

Stedinger, J. R. (1982). Confidence Intervals For Design Events. *Journal of Hydraulic Engineering*, 13-27.

Stedinger, J. R. (1983). Design Events With Specified Flood Risk. *Water Resources Research*, 511-522.

Stedinger, J. R., & Cohn, T. A. (1986). Flood frequency analysis with historical and paleoflood information. *Water Resources Research, 22*(5).

Stephenson, A., & Ribatet, M. (2006). *A User's Guide to the evdbayes Package (Version 1.1).* Newcastle, NSW: Macquarie University.

ter Braak, C. J., & Vrugt, J. A. (2008). Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, 435-446.

U.S. Geological Survey. (1982). *Guidelines for Determining Flood Flow Frequency Bulletin 17B.*

U.S. Geological Survey. (2018). *Guidelines for Determining Flood Flow Frequency Bulletin 17C.* https://doi.org/10.3133/tm4B5.

Viglione, A., Merz, R., Salinas, J. L., & Bloschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research, 49*(2). Retrieved from doi:10.1029/2011WR010782

Wagner, D. M., Krieger, J. D., & Veilleux, A. G. (2016). *Methods for estimating annual exceedance probability discharges for streams in Arkansas, based on data through water year 2013.* U.S. Geological Survey. Reston: U.S. Geological Survey. doi:http://dx.doi.org/10.3133/sir20165081

Appendix – Probability Distributions

Normal and Related Distributions

Normal

Description

The Normal (or Gaussian) distribution is a very widely used two-parameter probability distribution. The Normal distribution is fundamental to most statistical modeling because of the Central Limit Theorem, which states that the mean of independent random variables trends towards a Normal distribution, even if the original variables themselves are not Normally distributed. The Normal distribution fits many natural phenomena, such as body heights, blood pressure, measurement error, and annual rainfall.

Parameters

location: μ
scale: σ

Support

$-\infty < x < +\infty$

$-\infty < \mu < +\infty$

$0 < \sigma < +\infty$

Distribution Functions

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$F(x|\mu,\sigma) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$F^{-1}(p|\mu,\sigma) = \mu + \sigma\sqrt{2}\,\text{erf}^{-1}(2p-1)$$

Log-Normal

Description

The log-Normal distribution is a two-parameter positively skewed distribution that describes a random variable whose logarithm is Normally distributed. A log-Normal process arises from the multiplicative product of many independent random variables, each of which is positive. In hydrology, the log-Normal distribution is used for frequency analysis of annual maximum discharge. In reliability analysis, it is often used to model times to repair a system.

RMC-BestFit contains two log-Normal distributions. The first, named "Ln-Normal" is based on the natural logarithm, or log base e. This distribution is parameterized using real-space moments to be more intuitive for multi-disciplinary users of the software. The other distribution, named "Log-Normal" uses log base 10 and is parameterized using $\log_{10}$ moments, which is consistent with typical practice in hydrologic frequency analysis. Both of these log-Normal distributions are functionally identical, and will produce the same statistical inference.

Parameters

location: μ
scale: σ

Support

$0 < x < +\infty$

$-\infty < \mu < +\infty$

$0 < \sigma < +\infty$

Distribution Functions

$$f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}$$

$$F(x|\mu,\sigma) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$F^{-1}(p|\mu,\sigma) = e^{\mu+\sigma\sqrt{2}\operatorname{erf}^{-1}(2p-1)}$$

Gamma Family of Distributions

Exponential

Description

The two-parameter (or shifted) Exponential distribution is a special case of the Gamma family of distributions, which includes the two-parameter Gamma, Pearson Type III, and Log-Pearson Type III. The Exponential distribution describes the time between events in a Poisson process; i.e., a process in which events occur continuously and independently at a constant average rate. The Exponential distribution is often used in reliability applications, where it can be used to model data with a constant failure rate. This distribution is useful for modeling highly positively skewed data that have a non-zero lower bound.

Parameters

location: ξ
scale: α

Support

$\xi \leq x < +\infty$

$-\infty < \xi < +\infty$

$0 < \alpha < +\infty$

Distribution Functions

$$f(x|\xi, \alpha) = \alpha^{-1} e^{-\left(\frac{x-\xi}{\alpha}\right)}$$

$$F(x|\xi, \alpha) = 1 - e^{-\left(\frac{x-\xi}{\alpha}\right)}$$

$$F^{-1}(p|\xi, \alpha) = \xi - \alpha \ln(1 - p)$$

Gamma

Description

The Gamma distribution is a two-parameter, positively-skewed distribution. The Exponential, Erlang, and Chi-squared distributions are special cases of the Gamma distribution. There are three different parameterizations for the Gamma distribution in common use:

1) With a scale parameter θ and a shape parameter κ.

2) With an inverse scale parameter $\beta = 1/\theta$, called a rate parameter, and a shape parameter $\alpha = \kappa$.

3) With a mean parameter $\mu = k\theta = \alpha/\beta$ and a shape parameter κ.

RMC –BestFit uses the first parameterization, with a scale parameter θ and a shape parameter κ.

The Gamma distribution has applications in hydrology, econometrics, and other applied fields. It is a flexible distribution that is capable of modeling many kinds of data with a mild positive skew and lower bound of zero.

Parameters

scale: θ
shape: κ

Support

$0 < x < +\infty$

$0 < \theta < +\infty$

$0 < \kappa < +\infty$

Distribution Functions

$f(x|\theta,\kappa) = \frac{1}{\Gamma(\kappa)\theta^{\kappa}} x^{\kappa-1} e^{-\frac{x}{\theta}}$;where $\Gamma(\cdot)$ is the gamma function.

$F(x|\theta,\kappa) = \frac{1}{\Gamma(\kappa)} \gamma\left(\kappa, \frac{x}{\theta}\right)$;where γ is the lower incomplete gamma function.

$F^{-1}(p|\theta,\kappa)$ has no explicit analytical form.

Pearson Type III

Description

The Pearson Type III (PIII) distribution is a three-parameter distribution that is widely used in hydrologic frequency analysis. It has also been used to model the probability of wind speed and rainfall intensity. The PIII distribution is deduced from the two-parameter gamma distribution, and converges to a Normal distribution as its skewness (γ) approaches zero.

In RMC-BestFit, the PIII distribution is parameterized using the central moments of the distribution mean (μ), standard deviation (σ), and skewness (γ). The true parameters (location, scale, and shape) are computed from the specified moments. This is done because the moments of the distribution are more intuitively defined by end-users familiar with Bulletin 17B (U.S. Geological Survey, 1982) and Bulletin 17C (U.S. Geological Survey, 2018). RMC-BestFit uses the same parameterization as (Hosking & Wallis, 1997), with the underlying location parameter ξ, the scale parameter β, and the shape parameter α.

Parameters

mean: μ
standard deviation: σ
skewness: γ

location: $\xi = \mu - \frac{2\sigma}{\gamma}$
scale: $\beta = \frac{1}{2}\sigma\gamma$
shape: $\alpha = \frac{4}{\gamma^2}$

Support

If $\gamma < 0$, $\quad -\infty < x \leq \xi$

If $\gamma = 0$, $\quad -\infty < x < +\infty$

If $\gamma > 0$, $\quad \xi \leq x < +\infty$

$-\infty < \mu < +\infty$

$0 < \sigma < +\infty$

$-\infty < \gamma < +\infty$;however, the method of maximum likelihood can only produce a solution if $-2 \leq \gamma \leq +2$

Distribution Functions

If $\gamma < 0$,

$f(x|\xi,\beta,\alpha) = g(\xi - x|\beta,\alpha)$,

$F(x|\xi,\beta,\alpha) = 1 - G(\xi - x|\beta,\alpha)$,

$F^{-1}(p|\xi,\beta,\alpha) = \xi - G^{-1}(1 - p|\beta,\alpha)$

where $g(\cdot)$ is the Gamma distribution probability density function (PDF); $G(\cdot)$ is the Gamma cumulative distribution function (CDF); and G^{-1} is the Gamma inverse CDF.

If $\gamma = 0$, then the distribution in Normal

$f(x|\mu,\sigma) = n(x|\mu,\sigma)$,

$F(x|\mu,\sigma) = N(x|\mu,\sigma)$,

$F^{-1}(p|\mu,\sigma) = N^{-1}(p|\mu,\sigma)$

where $n(\cdot)$ is the Normal distribution PDF; $N(\cdot)$ is the Normal CDF; and N^{-1} is the Normal inverse CDF.

If $\gamma > 0$,

$f(x|\xi,\beta,\alpha) = g(x - \xi|\beta,\alpha)$,

$F(x|\xi,\beta,\alpha) = G(x - \xi|\beta,\alpha)$,

$F^{-1}(p|\xi,\beta,\alpha) = \xi + G^{-1}(p|\beta,\alpha)$

Log-Pearson Type III

Description

The log-Pearson Type III (LPIII) distribution is a flexible three-parameter distribution that describes a random variable whose logarithm is PIII distributed. The LPIII distribution was originally used to model annual maximum flood flows.

In RMC-BestFit, the LPIII uses log base 10 and is parameterizes using $\log_{10}$ moments of the distribution mean (μ), standard deviation (σ), and skewness (γ). The true parameters (location, scale, and shape) are computed from the specified moments. This is done because the moments of the distribution are more intuitively defined by end-users familiar with Bulletin 17B (U.S. Geological Survey, 1982) and Bulletin 17C (U.S. Geological Survey, 2018).

Parameters

mean (of log): μ
standard deviation (of log): σ
skewness (of log): γ

location: $\xi = \mu - \frac{2\sigma}{\gamma}$
scale: $\beta = \frac{1}{2}\sigma\gamma$
shape: $\alpha = \frac{4}{\gamma^2}$

Support

If $\gamma < 0$, $\quad 0 < x \leq \xi$

If $\gamma = 0$, $\quad 0 < x < +\infty$

If $\gamma > 0$, $\quad \xi \leq x < +\infty$

$-\infty < \mu < +\infty$

$0 < \sigma < +\infty$

$-\infty < \gamma < +\infty$;however, the method of maximum likelihood can only produce a solution if $-2 \leq \gamma \leq +2$

Distribution Functions

If $\gamma < 0$,

$$f(x|\xi,\beta,\alpha) = g(\log_{10}(\xi - x)\,|\beta,\alpha)\left(\frac{K}{x}\right),$$

$$F(x|\xi,\beta,\alpha) = 1 - G(\log_{10}(\xi - x)\,|\beta,\alpha),$$

$$F^{-1}(p|\xi,\beta,\alpha) = e^{\left(\xi - G^{-1}(1-p|\beta,\alpha)\right)/K}$$

where $g(\cdot)$ is the Gamma distribution probability density function (PDF); $G(\cdot)$ is the Gamma cumulative distribution function (CDF); G^{-1} is the Gamma inverse CDF; and $K = \log_{10} e$.

If $\gamma = 0$, then the distribution in Normal

$$f(x|\mu,\sigma) = n(\log_{10}(x)\,|\mu,\sigma)\left(\frac{K}{x}\right),$$

$$F(x|\mu,\sigma) = N(\log_{10}(x)\,|\mu,\sigma),$$

$$F^{-1}(p|\mu,\sigma) = e^{\left(N^{-1}(p|\mu,\sigma)\right)/K}$$

where $n(\cdot)$ is the Normal distribution PDF; $N(\cdot)$ is the Normal CDF; and N^{-1} is the Normal inverse CDF.

If $\gamma > 0$,

$$f(x|\xi,\beta,\alpha) = g(\log_{10}(x - \xi)\,|\beta,\alpha)\left(\frac{K}{x}\right),$$

$$F(x|\xi,\beta,\alpha) = G(\log_{10}(x - \xi)\,|\beta,\alpha),$$

$$F^{-1}(p|\xi,\beta,\alpha) = e^{\left(\xi + G^{-1}(p|\beta,\alpha)\right)/K}$$

Extreme Value Distributions

Gumbel (Extreme Value Type I)

Description

The Gumbel, or Extreme Value Type-I (EVI) distribution, is a two-parameter distribution with a fixed positive skewness of ≈ 1.14. The Gumbel distribution is used to describe the maximum (or minimum) of a number of samples, and has seen widespread use in hydrologic frequency analysis. The Gumbel distribution is a particular case of the Generalized Extreme Value (GEV) distribution when the GEV shape parameter is zero. The Gumbel distribution is also used for a probability plotting scale because it exaggerates the extreme tails of the data.

Parameters

location: ξ
scale: α

Support

$-\infty < x < +\infty$

$-\infty < \xi < +\infty$

$0 < \alpha < +\infty$

Distribution Functions

$f(x|\xi,\alpha) = \frac{1}{\alpha} e^{-(z+e^{-z})}$;where $z = \frac{x-\xi}{\alpha}$

$F(x|\xi,\alpha) = e^{-e^{-z}}$

$F^{-1}(p|\xi,\alpha) = \xi - \alpha \ln(-\ln(p))$

Weibull

Description

The Weibull distribution is a two-parameter distribution commonly used in reliability analysis. It is related to a number of other probability distributions. In particular, the Weibull distribution interpolates between the exponential distribution (for $\kappa = 1$) and the Rayleigh distribution (when $\kappa = 2$). If the quantity x is the "time to failure", the Weibull distribution gives the distribution for which the failure rate is proportional to a power of time.

Parameters

scale: λ
shape: κ

Support

$0 \leq x < +\infty$

$0 < \lambda < +\infty$

$0 < \kappa < +\infty$

Distribution Functions

$$f(x|\lambda,\kappa) = \frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} e^{-\frac{x^{\kappa}}{\lambda}}$$

$$F(x|\lambda,\kappa) = 1 - e^{-\frac{x^{\kappa}}{\lambda}}$$

$$F^{-1}(p|\lambda,\kappa) = \lambda \ln\left(\frac{1}{1-p}\right)^{\frac{1}{\kappa}}$$

Generalized Extreme Value

Description

The Generalized Extreme Value (GEV) distribution is a three-parameter distribution that subsumes the three extreme-value distributions: Gumbel (EVI), Fréchet (EVII), and Weibull (EVIII). The shape parameter κ determines which sub-distribution the GEV represents. The extreme value theorem states that the GEV distribution is the limit distribution of maxima of a sequence of independent and identically distributed random values. GEV is used for hydrologic frequency analysis, insurance, and financial risks.

RMC-BestFit uses the Hosking's parameterization (Hosking & Wallis, 1997), in which a negative shape parameter κ has no upper bound, and a positive κ has a fixed upper bound. Other sources adopt the opposite convention, where a negative κ implies an upper bound.

Parameters

location: ξ
scale: α
shape: κ

Support

If $\kappa < 0$, $\quad \xi + \alpha/\kappa \leq x < +\infty \qquad -\infty < \xi < +\infty$

If $\kappa = 0$, $\quad -\infty < x < +\infty \qquad 0 < \alpha < +\infty$

If $\kappa > 0$, $\quad -\infty < x \leq \xi + \alpha/\kappa \qquad -\infty < \kappa < +\infty$

Distribution Functions

$$\begin{cases} z = -\dfrac{1}{\kappa}\ln\left(1 - \kappa\left(\dfrac{x-\xi}{\alpha}\right)\right) & if\ \kappa \neq 0, \\ z = \dfrac{x-\xi}{\alpha} & if\ \kappa = 0. \end{cases}$$

$$f(x|\xi,\alpha,\kappa) = \frac{1}{\alpha} e^{-(1-\kappa)z - e^{-z}}$$

$$F(x|\xi,\alpha,\kappa) = e^{-e^{-z}}$$

$$F^{-1}(p|\xi,\alpha,\kappa) = \begin{cases} \xi - \dfrac{\alpha}{\kappa}\left(1 - (-\ln(p))^{\kappa}\right) & if\ \kappa \neq 0, \\ \xi - \alpha\ln(-\ln(p)) & if\ \kappa = 0. \end{cases}$$

Generalized Pareto

Description

The Generalized Pareto (GPA) distribution is a three-parameter distribution with a fixed lower bound. The GPA is upper bounded when the shape parameter is positive. When κ is zero, the distribution reduces to a shifted Exponential distribution. The GPA distribution is most often used with peaks-over-threshold data. Hydrologic frequency analysis uses the GPA in cases where rainfall or flow maxima exceed a specified threshold.

RMC-BestFit uses the Hosking's parameterization (Hosking & Wallis, 1997), in which a negative shape parameter κ has no upper bound, and a positive κ has a fixed upper bound.

Parameters

location: ξ
scale: α
shape: κ

Support

If $\kappa \leq 0$, $\quad \xi \leq x < +\infty$ $\qquad -\infty < \xi \leq minimum\ of\ data$

If $\kappa > 0$, $\quad \xi \leq x \leq \xi + \alpha/\kappa$ $\qquad 0 < \alpha < +\infty$

$-\infty < \kappa < +\infty$

Distribution Functions

$$\begin{cases} z = -\dfrac{1}{\kappa}\ln\left(1 - \kappa\left(\dfrac{x-\xi}{\alpha}\right)\right) & if\ \kappa \neq 0, \\ z = \dfrac{x-\xi}{\alpha} & if\ \kappa = 0. \end{cases}$$

$$f(x|\xi,\alpha,\kappa) = \frac{1}{\alpha}e^{-(1-\kappa)z}$$

$$F(x|\xi,\alpha,\kappa) = 1 - e^{-z}$$

$$F^{-1}(p|\xi,\alpha,\kappa) = \begin{cases} \xi - \dfrac{\alpha}{\kappa}\left(1-(1-p)^{\kappa}\right) & if\ \kappa \neq 0, \\ \xi - \alpha\ln(1-p) & if\ \kappa = 0. \end{cases}$$

Logistic Distributions

Logistic

Description

The Logistic distribution is a two-parameter, symmetric distribution with heavier tails (higher kurtosis) than the Normal distribution. The Logistic distribution has applications in hydrology for long duration discharge or rainfall, such as monthly or yearly totals. The most common application is in logistic regression where the errors follow a Logistic distribution.

Parameters

location: ξ
scale: α

Support

$-\infty < x < +\infty$

$-\infty < \xi < +\infty$

$0 < \alpha < +\infty$

Distribution Functions

$f(x|\xi, \alpha) = \frac{1}{\alpha} e^{-z}(1 + e^{-z})^{-2}$;where $z = \frac{x-\xi}{\alpha}$

$F(x|\xi, \alpha) = (1 + e^{-z})^{-1}$

$F^{-1}(p|\xi, \alpha) = \xi + \alpha \ln(\frac{p}{1-p}))$

Generalized Logistic

Description

The Generalized Logistic (GLO) distribution is a heavy-tailed, three-parameter distribution. The GLO distribution has been used to fit values of extremes, such as stock return fluctuations and sea levels. It has been used extensively for modeling annual rainfall maxima, and for flood frequency analysis.

RMC-BestFit uses the Hosking's parameterization (Hosking & Wallis, 1997), in which a negative shape parameter κ has no upper bound, and a positive κ has a fixed upper bound.

Parameters

location: ξ
scale: α
shape: κ

Support

If $\kappa < 0$, $\quad \xi + \alpha/\kappa \le x < +\infty \qquad -\infty < \xi < +\infty$

If $\kappa = 0$, $\quad -\infty < x < +\infty \qquad 0 < \alpha < +\infty$

If $\kappa > 0$, $\quad -\infty < x \le \xi + \alpha/\kappa \qquad -\infty < \kappa < +\infty$

Distribution Functions

$$\begin{cases} z = -\dfrac{1}{\kappa}\ln\left(1 - \kappa\left(\dfrac{x-\xi}{\alpha}\right)\right) & if\ \kappa \neq 0, \\ z = \dfrac{x-\xi}{\alpha} & if\ \kappa = 0. \end{cases}$$

$$f(x|\xi,\alpha,\kappa) = \tfrac{1}{\alpha}e^{-(1-\kappa)z}(1+e^{-z})^{-2}$$

$$F(x|\xi,\alpha,\kappa) = (1+e^{-z})^{-1}$$

$$F^{-1}(p|\xi,\alpha,\kappa) = \begin{cases} \xi - \dfrac{\alpha}{\kappa}\left(1 - \left(\dfrac{1-p}{p}\right)^{\kappa}\right) & if\ \kappa \neq 0, \\ \xi - \alpha\ln(\dfrac{1-p}{p})) & if\ \kappa = 0. \end{cases}$$