



Solo

Software User's Guide

Version 1.1

Release Information

<i>Document Version Number</i>	<i>Solo-5.8-UG002</i>
<i>Software Version</i>	<i>5.8</i>
<i>Document Status</i>	<i>Final</i>
<i>Document Release Date</i>	<i>July 26, 2010</i>

Copyright

© 2010. Eigenvector Research Inc., All rights reserved.

The information contained herein is proprietary and confidential and is the exclusive property of Eigenvector Research Inc. It may not be copied, disclosed, used, distributed, modified, or reproduced, in whole or in part, without the express written permission of Eigenvector Research Inc.

Limit of Liability

Eigenvector Research Inc. has used their best effort in preparing this guide. Eigenvector Research Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of Eigenvector Research Inc. or any of its affiliates. The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

Trademarks

The Eigenvector logo, the name “*Eigenvector*,” *Solo*, and *PLS_Toolbox* are registered trademarks of Eigenvector Research Inc. All other products and company names mentioned herein may be trademarks or registered trademarks of their respective owners.

Customer Support

Customer support is available to organizations that purchase and have a current Annual Service Agreement (ASA). Contact Eigenvector Research Inc. at:

Eigenvector Research Inc.
3905 West Eaglerock Drive
Wenatchee, WA 98801
(509)-662-9213 (Phone)
509.662.9214 (Fax)
[*helpdesk@eigenvector.com*](mailto:helpdesk@eigenvector.com)
[*www.eigenvector.com*](http://www.eigenvector.com)

Table of Contents

Chapter 1: Solo Quick Start	13
Quick Start steps for an analysis in Solo	13
Chapter 2: Launching Solo	17
Chapter 3: Solo Windows.....	19
Chapter 4: Common Application Features	21
Options dialog box	21
FigBrowser.....	23
Chapter 5: Workspace Browser.....	25
Title bar	25
Main menu	25
Toolbar.....	26
Base Workspace.....	26
Chapter 6: Workspace Browser Preferences	27
To specify the Workspace Browser Shortcut icons.....	27
To edit the Workspace Browser options	28
To specify the Window docking settings.....	29
Chapter 7: Importing Data into the Workspace Browser	31
To import a .mat file into the Workspace Browser	32
To import a data file (other than a .mat file) into the Workspace Browser	34
To save imported data to a .mat file	35
Chapter 8: Icons in the Workspace Browser.....	37
Item icons.....	37
Saving, loading, and deleting items in the workspace	38
To save items to a .mat file.....	38
To load saved items into the base workspace.....	38
To delete items from the base workspace.....	38
Manipulating items	39
Viewing information about the item	39
Opening the item for viewing or editing	39

Dragging and dropping items	42
Chapter 9: DataSet Editor Window	45
DataSet Editor window layout	45
Info tab	46
Data tab	47
Row Labels tab/Column Labels tab	48
Edit menu	49
Chapter 10: Plot Controls Window	51
Plot window	52
Data plotting options.....	55
Data selection and editing options	55
Other options.....	56
Chapter 11: Analysis Window	59
Analysis window main menu	61
Edit menu.....	61
Tools menu	62
Analysis window toolbar	63
Analysis window Status pane.....	65
Analysis window Control pane	67
Analysis window Help pane.....	69
Analysis window Flowchart pane	71
Analysis window Model Cache pane.....	73
Model Cache pane view	73
Cache management.....	75
Manipulating cached items.....	76
Chapter 12: Preprocessing Methods	77
Preprocessing window	77
Chapter 13: Analysis Phases	81
Calibration phase	81
Test and Validation phase.....	81
Model Application phase	82

Chapter 14: Building the Model in the Calibration Phase	83
Loading the calibration data and building the initial model	83
Changing the number of components.....	85
Examining and refining the model.....	86
Chapter 15: Plotting Eigenvalues for a Calibration Model.....	89
Eigenvalues plot options	91
Chapter 16: Plotting Scores and Statistical Values for a Calibration Model.....	93
Changing the plot display.....	94
Refining the model by excluding samples.....	94
Chapter 17: Plotting Loads and Variable Statistics for a Calibration Model.....	97
Changing the plot display.....	98
Refining the model by removing variables	99
Chapter 18: Applying the Model in the Test and Validation Phase... 	103
Loading the validation data and applying the model to the data.....	103
Examining and refining the model.....	105
Chapter 19: Cross-Validation Tool	107
Chapter 20: Model Robustness Tool.....	111
Shifts	111
Interferences	112
Chapter 21: Correlation Map Tool	113

Preface

Welcome to the *Solo Software User's Guide*. The purpose of the *Solo Software User's Guide* is to provide a high-level overview of Solo's key functions so that you can begin to use the application efficiently and effectively.

Using the guide

You will find the *Solo Software User's Guide* easy to use. You can simply look up the topic that you need in the table of contents. Later, in this Preface, you will find a brief discussion of each chapter to further assist you in locating the information that you need.

Special information about the guide

The *Solo Software User's Guide* has a dual purpose design. It can be distributed electronically and then printed on an as-needed basis, or it can be viewed online in its fully interactive capacity. If you print the document, for best results, it is recommended that you print it on a duplex printer; however, single-sided printing will also work. If you view the document online, a standard set of bookmarks appears in a frame on the left side of the document window for navigation through the document. For better viewing, decrease the size of the bookmark frame and use the magnification box to increase the magnification of the document to your viewing preference.



The content of this guide was single-sourced for multiple outputs (printed documentation and a Wiki). To accommodate the production requirements for the different outputs, chapters in this manual might contain one or more blank pages.

Conventions used in the guide

The *Solo Software User's Guide* uses the following conventions:

- Information that can vary in a command—variable information—is indicated by alphanumeric characters enclosed in angle brackets; for example, <Preprocessing Method>. Do not type the angle brackets when you specify the variable information.
- A new term, or term that must be emphasized for clarity of procedures, is *italicized*.
- Page numbering is “online friendly.” Pages are numbered from 1 to x, *starting with the cover* and ending on the last page of the index.
- This guide is intended for both print and online viewing.
 - If information appears in [blue](#), it is a hyperlink. Table of Contents entries are also hyperlinks. Click the hyperlink to advance to the referenced information.

Assumptions for the guide

Solo is a stand-alone chemometrics suite that is based on PLS_Toolbox's Graphical User Interfaces and algorithms. Solo does not require that MATLAB be installed. Although this documentation references only Solo, all of the information is completely applicable to PLS_Toolbox. The *Solo Software User's Guide* assumes that:

- You are using Solo on a Windows operating system. If you are using Solo on a different operating system (Mac or Linux), you will note some differences.
- You are familiar with windows-based applications and basic Windows functions and navigational elements.

Organization of the guide

In addition to this Preface, the *Solo Software User's Guide* contains the following chapters:

- [Chapter 1, “Solo Quick Start,” on page 13](#) is designed to help you get started fast with Solo by explaining, at a high-level, how you typically carry out one of the most common analyses in Solo—a Principal Components analysis (PCA). The analysis is outlined as a series of easy steps with each step linked to the appropriate chapter in the guide that details the step.
- [Chapter 2, “Launching Solo,” on page 17](#) explains how to launch Solo.
- [Chapter 3, “Solo Windows,” on page 19](#) provides a high-level overview of the four primary windows in Solo—the Workspace Browser window, the DataSet Editor window, the Analysis window, and the Plot Controls window.
- [Chapter 4, “Common Application Features,” on page 21](#) details the application features that are common to both the Workspace Browser window and the Analysis window as well as some other windows.
- [Chapter 5, “Workspace Browser,” on page 25](#) details the layout and organization of the Workspace Browser, which is the starting interface for Solo.
- [Chapter 6, “Workspace Browser Preferences,” on page 27](#) details how to customize the Workspace Browser options to better suit your working needs.
- [Chapter 7, “Importing Data into the Workspace Browser,” on page 31](#) describes how to import data into the Workspace Browser for analysis.
- [Chapter 8, “Icons in the Workspace Browser,” on page 37](#) details the different icons that are used to represent the different types of items and data in the Workspace Browser and how you can work with and manipulate these icons.
- [Chapter 9, “DataSet Editor Window,” on page 45](#) details the layout and organization of the DataSet Editor window, which is the standard interface that you use for creating and managing a DataSet in Solo. It also provides a high-level overview of the functions that are available from the window.

- [Chapter 10, “Plot Controls Window,” on page 51](#) details the layout and organization of the Plot Controls window, which is the principal data visualization tool for Solo. It also provides a high-level overview of the functions that are available from the window.
- [Chapter 11, “Analysis Window,” on page 59](#) details the layout and organization of the Analysis window, which serves as the core interface to the Solo data modeling and analysis functions. It also provides a high-level overview of the functions that are available from the window.
- [Chapter 12, “Preprocessing Methods,” on page 77](#) describes the basic steps for setting up preprocessing rules for an analysis and verifying that the rules that you have set up are as you want them.
- [Chapter 13, “Analysis Phases,” on page 81](#) provides a high-level overview of the three phases that are required to completely carry out modeling and analysis in the Analysis window—the Calibration phase, the Test and Validation phase, and the Model Application phase.
- [Chapter 14, “Building the Model in the Calibration Phase,” on page 83](#) further details the Calibration phase, which is one of the three phases that are required to completely carry out modeling and analysis in the Analysis window.
- [Chapter 15, “Plotting Eigenvalues for a Calibration Model,” on page 89](#) details the Plot Eigenvalues function, which is a function that is common to most analysis methods.
- [Chapter 16, “Plotting Scores and Statistical Values for a Calibration Model,” on page 93](#) details the plotting of scores and statistical values for a calibration model, which is a function that is common to most analysis methods.
- [Chapter 17, “Plotting Loads and Variable Statistics for a Calibration Model,” on page 97](#) details the plotting of loads and variable statistics for a calibration model, which is a function that is common to most analysis methods.
- [Chapter 18, “Applying the Model in the Test and Validation Phase,” on page 103](#) further details the Test and Validation phase, which is one of the three phases that are required to completely carry out modeling and analysis in the Analysis window.
- [Chapter 19, “Cross-Validation Tool,” on page 107](#) provides a high-level overview of the Cross-Validation tool, which is a tool that you use to assess the optimal complexity of a model and to estimate the performance of a model when you apply the model to unknown data.
- [Chapter 20, “Model Robustness Tool,” on page 111](#) provides a high-level overview of the Model Robustness tool, which is a tool that you use to measure the sensitivity of a regression model to artifacts in new spectroscopic measurements.
- [Chapter 21, “Correlation Map Tool,” on page 113](#) provides a high-level overview of the Correlation Map tool, which is a tool that you use to show the degree of correlation among the variables after you have loaded x block data.

Preface
Solo Software User's Guide

Chapter 1: Solo Quick Start

Welcome to the Solo Quick Start chapter. This chapter is designed to help you get started fast with Solo by explaining, at a high level, the basic steps for carrying out a Principal Component Analysis (PCA), which is one of the most commonly carried out analyses in Solo. In addition, the basic steps for a PCA also touch on the basic steps for almost all of the other analysis methods that are available in Solo. Each step contains one or more links to the appropriate chapters in this guide that provide detailed information about the step. Some important points to note about this Quick Start chapter are the following:

- The steps detailed here are designed to compliment the steps that are defined in “PCA in Wine Data,” which is an online tutorial that walks you through the step by step basics of carrying out a PCA in Solo. The tutorial is located at <http://www.eigenvector.com/eigenguide.php>. You can use the procedure outlined in this chapter to follow along with the tutorial. At any time, you can pause the online tutorial, and click on a link next to a step to go to the indicated chapter to learn more about the step.
- The tutorial and the quick start steps are based on using the wine DataSet, which is demo data that is loaded during the installation of Solo. You can repeat the steps listed here using this DataSet, or you can use another smaller DataSet, for example, arch.

Note: If you running Solo on a Windows OS, the demo data is loaded in C:\Program Files\EVRI\Solo\Demo_Data. Contact Eigenvector for assistance in locating the demo data for other OSs and/or for selecting a different DataSet.

Note: Before you carry out the quick start steps, it might be helpful to have a detailed overview of the windows and the modeling and analysis phases in Solo. See [Chapter 3, “Solo Windows,”](#) on page 19 and [Chapter 13, “Analysis Phases,”](#) on page 81.

Quick Start steps for an analysis in Solo

1. Launch Solo. See [Chapter 2, “Launching Solo,”](#) on page 17.

After you launch Solo, the Workspace Browser opens automatically. The Workspace Browser is your starting interface for Solo. The interface provides quick access to all of the data analysis tools. For information about the Workspace Browser its layout and its options—see:

- [Chapter 5, “Workspace Browser,”](#) on page 25.
- [Chapter 6, “Workspace Browser Preferences,”](#) on page 27.

2. Import data into the Workspace Browser.

The selected data files are loaded into the Workspace Browser. After you import the data, different icons are displayed in the Workspace Browser for the different data types. You can save these data items to a workspace, and you can manipulate this data in the browser before you analyze it. See:

- [Chapter 7, “Importing Data into the Workspace Browser,”](#) on page 31.
- [Chapter 8, “Icons in the Workspace Browser,”](#) on page 37.

3. Load the imported data into the PCA tool for analysis by dragging the data icon onto the Decompose (PCA) shortcut icon.

The Drag and Drop method is only one of the variety of methods that are available for opening an Analysis window and loading data. All of the available methods are discussed in detail in the appropriate chapters in this guide. For a detailed discussion of the Analysis window, see [Chapter 11, “Analysis Window,” on page 59](#).

4. Optionally, to view the raw data in a spreadsheet layout prior to analysis, and if necessary, edit the data prior to analysis, open the data in the DataSet Editor window. See [Chapter 9, “DataSet Editor Window,” on page 45](#).

Information that you glean in this view can help you understand the patterns that you will see later when generating plots and other visual aids of sample relationships and variable relationships.

5. Optionally, plot the raw data for review prior to analysis. See [Chapter 10, “Plot Controls Window,” on page 51](#).

6. Select the appropriate preprocessing methods.

Data preprocessing describes any type of processing procedures that are performed on raw data to prepare it for *another* processing procedure and ultimately, analysis. Preprocessing linearizes the relationships among the variables in your DataSet and removes extraneous sources of variation that are of no interest to the analysis. A variety of preprocessing methods are available in Solo. See [Chapter 12, “Preprocessing Methods,” on page 77](#).

7. Generate the calibration (initial) model.

The Calibration phase consists of model building and exploratory analysis. In this phase, which affects only the Calibration side of the Status pane, you identify any patterns or trends in the data, and any other information that you consider relevant, for example, any relationships that might exist between the x data and the y data, and use this information to build a model. See [Chapter 14, “Building the Model in the Calibration Phase,” on page 83](#).

8. Create plots and other visual aids that assist you in examining and refining the model by excluding certain samples and/or variables to enhance the model performance.

- To generate plots and other visual aids that show the relationship among the *samples* in your data, see [Chapter 16, “Plotting Scores and Statistical Values for a Calibration Model,” on page 93](#).
- To generate plots and other visual aids that show the relationship among the *variables* in your data, see [Chapter 17, “Plotting Loads and Variable Statistics for a Calibration Model,” on page 97](#).

9. Apply the model to new data to verify that the model will provide acceptable results for the analysis of *validation* data, which is data with known physical and/or chemical characteristics. See [Chapter 18, “Applying the Model in the Test and Validation Phase,” on page 103](#).

10. Save the model to the Workspace Browser or to a file and use it at a later date ((File > Save Model on the Analysis window main menu), or export the model to a file or a predictor. (File > Export Model on the Analysis window main menu.)

Note: See http://wiki.eigenvector.com/index.php?title=Exporting_Models

Chapter 1
Solo Quick Start

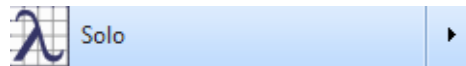
Chapter 2: Launching Solo

After installation, a shortcut icon for Solo is placed on your desktop. An option for the application is also available from your Start menu. You can double-click the desktop icon to launch Solo, or you can select the option from your Start menu.

Figure 2-1: Solo application desktop shortcut icon



Figure 2-2: Solo application Start menu option



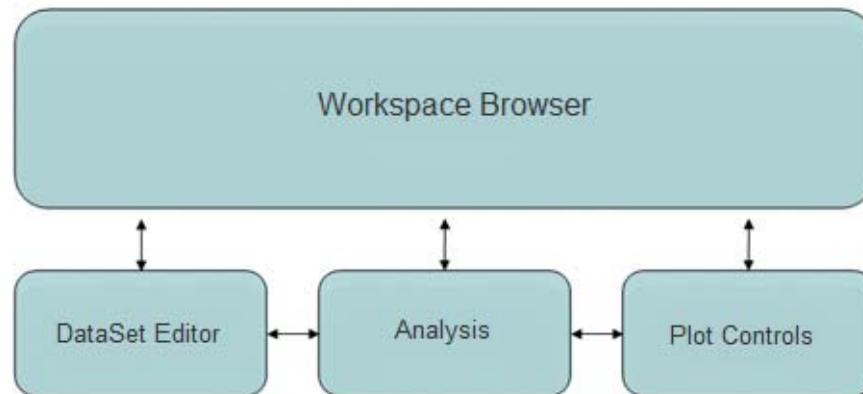
After you launch Solo, the Workspace Browser opens automatically. See [Chapter 5, "Workspace Browser,"](#) on page 25.

Chapter 2
Launching Solo

Chapter 3: Solo Windows

Solo is organized around four primary windows—the Workspace Browser window, the DataSet Editor window, the Analysis window, and the Plot Controls window. Each window provides functions that are dedicated to a specific step in the data analysis process.

Figure 3-1: Solo windows



- Workspace Browser window—The Workspace Browser is your starting interface for Solo. The interface provides quick access to all of the data analysis tools. The browser also serves as your “scratch pad” in Solo—you can pre-process data for analysis in the browser, or you can import data into the browser and then manipulate the data prior to analysis. See [“Workspace Browser” on page 25](#).
- DataSet Editor window—The Dataset Editor window is the primary data handling window in Solo. It provides a variety of functions and tools for loading, editing, and saving data. See [“DataSet Editor Window” on page 45](#).
- Analysis window—The Analysis window serves as the core interface to the Solo data modeling and analysis functions. You create your models in an Analysis window, analyze and explore the models in this window, and also apply models in this window. See [“Analysis Window” on page 59](#).
- Plot Controls window—The Plot Controls window is the principal data visualization tool for Solo. It provides an extensive number of tools for labeling, manipulating, and publishing plots that you generate in Solo. See [“Plot Controls Window” on page 51](#).

Any window in Solo can be a floating window or it can be docked. You can specify Window docking settings either through options in the Workspace Browser or through options in an Analysis window. See [“Workspace Browser Preferences” on page 27](#) or [“Analysis window main menu” on page 61](#).

Chapter 3
Solo Windows

Chapter 4: Common Application Features

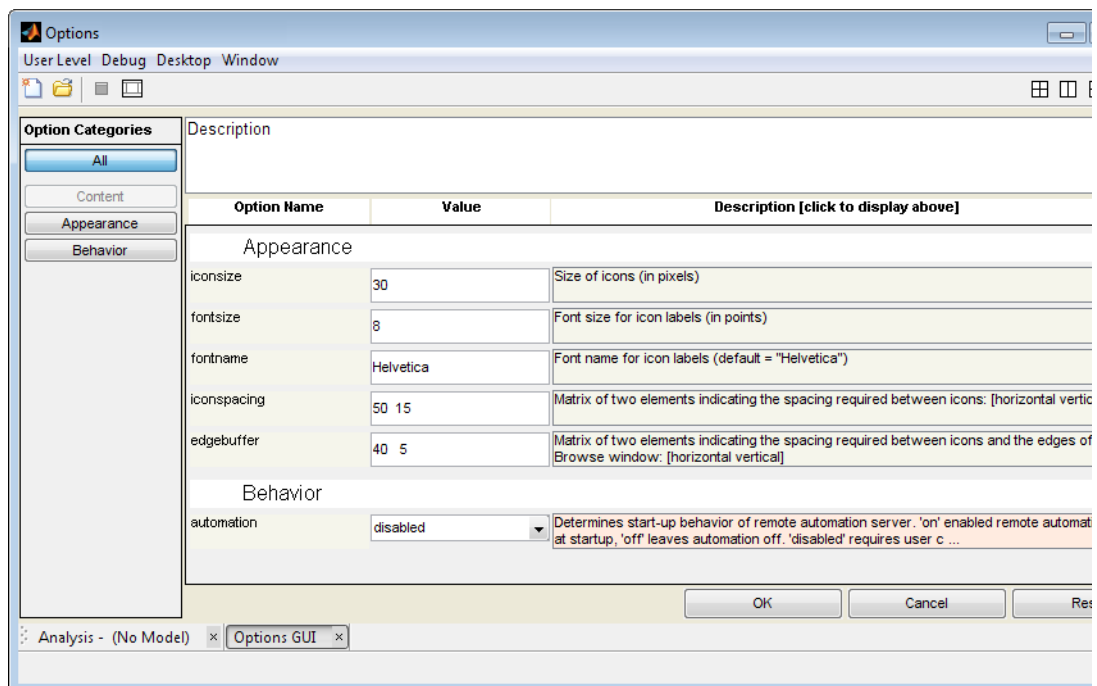
Solo contains several application features that are common to both the Workspace Browser window and the Analysis window as well as some other windows. Two of these features are the Options dialog box and the FigBrowser. See:

- “Options dialog box” below.
- “FigBrowser” on page 23.

Options dialog box

Options are settings that affect the behavior of a function or window in Solo. You can modify the default values for options so that the behavior of the functions and windows in Solo better suits your working needs. When an Options dialog box first opens, the dialog box lists *all* of the options that you can modify for the function or window, grouped by category. The dialog box also lists the current value for each option and a description of the option. For example, [Figure 4-1](#) shows the Options dialog box for the Workspace Browser window. The options that are listed in the dialog box affect the *display* properties for the Workspace Browser (such as icon size, font size, and icon font, which are grouped in the Appearance category) and the *interactivity* properties for the Workspace Browser (such as single-click behavior versus double-click behavior and dragging functions, which are grouped in the Behavior category).

Figure 4-1: Options dialog box

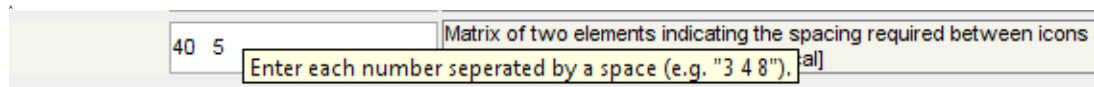


Chapter 4 Common Application Features

A variety of options are available for working with the dialog box:

- The User Level setting is a filter that determines which categories are displayed in the Options dialog box and which are not. By default, the User Level is set to “Intermediate” in the Options dialog box, which meets the majority of users’ needs. You can, however, change the level to “Advanced” to display additional categories for which you can modify the option values, or you can or change the level to “Beginner” to simplify the number of options that are displayed.
- If a description is not displayed in its entirety in the Options dialog box, it is highlighted in pink. You can click on the description to view it in its entirety in the Description pane at the top of the dialog box, or you can simply resize the window.
- To view the options for only a specific category, you can click the appropriate category in the Option Categories pane. For example, in the Options dialog box shown above, to view only the appearance options for the Workspace Browser, click Appearance in the Option Categories pane.
- When you are entering or modifying a value for an option, you can hold your mouse pointer over a field to view specific instructions for entering the value in the field.

Figure 4-2: Viewing instructions for entering a value in an Options dialog box field



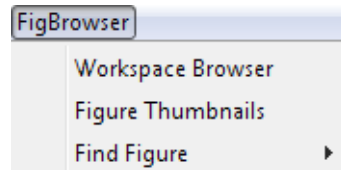
After you enter the value, and then click OK, the Options dialog box closes and you return to the opened window or tool for which you made the changes. Any changes that you made are effective immediately.

Note: Most options are persistent (remembered from one Solo session to another) while other options apply to only the current session, or only the current window (such as the Method options in the Analysis window.)

FigBrowser

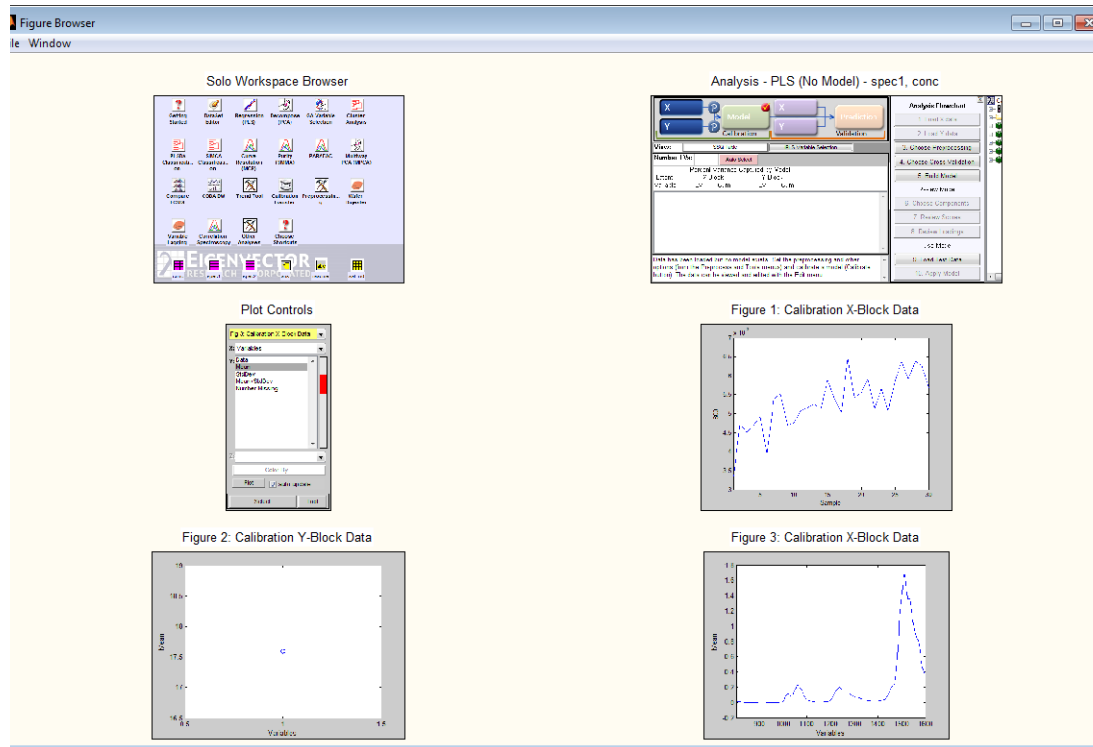
As you carry out an analysis, you can often generate multiple plots and other figures. You can use the FigBrowser utility for managing and viewing these multiple figures. This utility is available on the main menu of the Workspace Browser window, on the main menu of any analysis window, on the main menu of the Plot Controls window, and on the main menu of a Plot window.

Figure 4-3: FigBrowser option



- **Workspace Browser**—Brings the open Workspace Browser window to the front and makes it the active window.
- **Figure Thumbnails**—Opens the Figure Browser window which contains all of your currently opened figures and provides you a quick way for navigating among them. (The Figure Browser window also displays the Workspace Browser window, any open analysis window, and the Plot Controls window.) Click on a figure thumbnail to bring the figure to the front and make it the currently active figure. After you click on a thumbnail, the Thumbnail window closes.

Figure 4-4: Figure Browser window

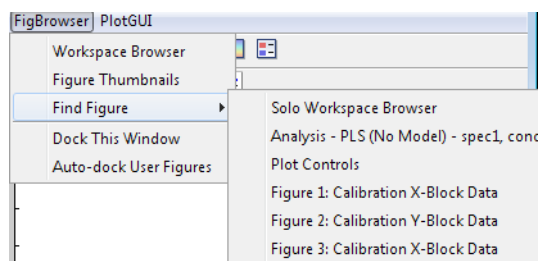


Chapter 4 Common Application Features

- Find Figure—Lists all of the currently opened figures in a list by Figure title.

Note: Although this option provides another means of navigating among multiple figures, if you have two of the same kind of plot open, then the only difference by which you can discern plots is the figure title.

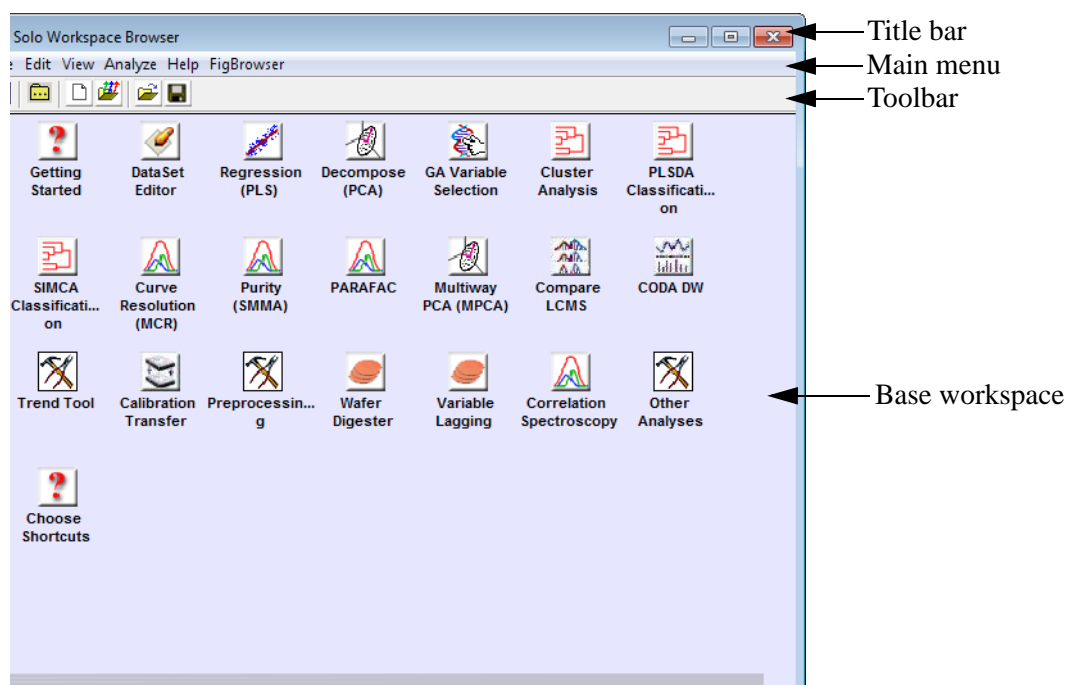
Figure 4-5: Find Figure option



Chapter 5: Workspace Browser

The Workspace Browser opens automatically when you launch Solo.

Figure 5-1: Workspace Browser



The Workspace Browser is your starting interface for Solo. The interface provides quick access to all of the data analysis tools. The browser also serves as your “scratch pad” in Solo—you can pre-process data for analysis in the browser, or you can import data into the browser and then manipulate the data prior to analysis. The Solo Workspace Browser has four major components—the [title bar](#), the [main menu](#), the [toolbar](#), and the [base workspace](#).

Title bar

The phrase “Solo Workspace Browser” appears in the title bar at the top of the Solo Workspace Browser, as well as the standard Window Minimize, Maximize, and Close buttons.

Main menu







The main menu is set up in a standard Windows format with commands grouped into menus (File, Edit, View, Analyze, Help and FigBrowser) across the menu bar. Some of these menu commands are available in other areas of the browser.

Note: The File > Remote Automation command allows the control of Solo by another program. If Solo is running, certain third party programs can connect to Solo and directly “dump” data into Solo and even start an Analysis window with the data loaded. This

prevents you from having to carry out the manual process of opening a file, importing the file, and bringing the file into the Analysis window. Contact Eigenvector for assistance with this function.

Toolbar

The toolbar provides quick access to some of the most commonly used Workspace Browser functions. Place your mouse pointer over a toolbar button to open tooltip text for the button. Some of these functions are available in other areas of the browser.

Button	Function
	Refresh Browser icon - Refreshes the current display for the Workspace Browser.
	Change Working Directory icon - Opens the Browse for Folder dialog box which you use to browse to and select a different <i>working</i> directory (the directory that is associated with the current Workspace Browser process.)
	New Dataset icon - Opens the New DataSet dialog box which you use to specify the size and initial value for a new DataSet, which is the object used in Solo for managing data.
	Import Data icon - Opens the Import dialog box which you use to select a specific data file type for importing into Solo.
	Load Workspace icon - Opens the Load Workspace dialog box from which you can select a .mat file to load into the Workspace Browser. (See Save Workspace icon below.)
	Save Workspace icon - Saves the currently loaded items to a single .mat file. You can save as many .mat files as needed to fit your working requirements, and then use the Load Workspace icon to load a saved .mat file into the base workspace.

Base Workspace

The Workspace Browser shows the contents of the *base workspace*. The base workspace contains the shortcut icons for *all* of the Analysis tools. Some of these shortcut icons open the Analysis window in a specific analysis mode (for example, if you click Decompose (PCA), the Analysis window opens with PCA as the selected analysis mode) and some of these icons open specific windows (for example, if you click GA Variable selection, the Genetic Algorithm Variable Selection window opens). The base workspace also contains a Getting Started icon which launches the Eigenvector Research Documentation wiki, and a Choose Shortcuts icon, which opens the Select Shortcuts to Show dialog box. You can use the options in this dialog box to specify which shortcut icons are displayed in the base workspace for your Workspace Browser. (See [“Workspace Browser Preferences” on page 27.](#))

Chapter 6: Workspace Browser Preferences

When Solo first opens, it opens with a default set of Shortcut icons for *all* of the Analysis tools. You can specify which Shortcut icons are to be displayed in your Workspace Browser. Also, because Solo runs on Windows, LINUX, and MacIntosh operating systems, you might need to customize some of the browser options to better suit your working needs. Finally, you can also specify Windows docking settings which determine how interfaces and data figures can be moved and resized. See:

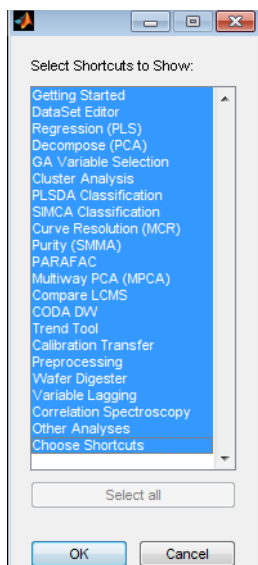
- “To specify the Workspace Browser Shortcut icons” below.
- “To edit the Workspace Browser options” on page 28.
- “To specify the Window docking settings” on page 29.

To specify the Workspace Browser Shortcut icons

1. Do one of the following:
 - Click Choose Shortcut icons.
 - On the Workspace Browser menu, click Edit > Options > Workspace Shortcuts.

The Select Shortcuts to Show dialog box opens. By default, all Shortcut icons are selected.

Figure 6-1: Select Shortcuts to Show dialog box



2. Clear the selections for the Shortcut icons that you do *not* want to show in the Workspace Browser, and then click OK.

The Select Shortcuts to Show dialog box closes and you return to the Workspace Browser. Any changes that you made are effective immediately.

To edit the Workspace Browser options

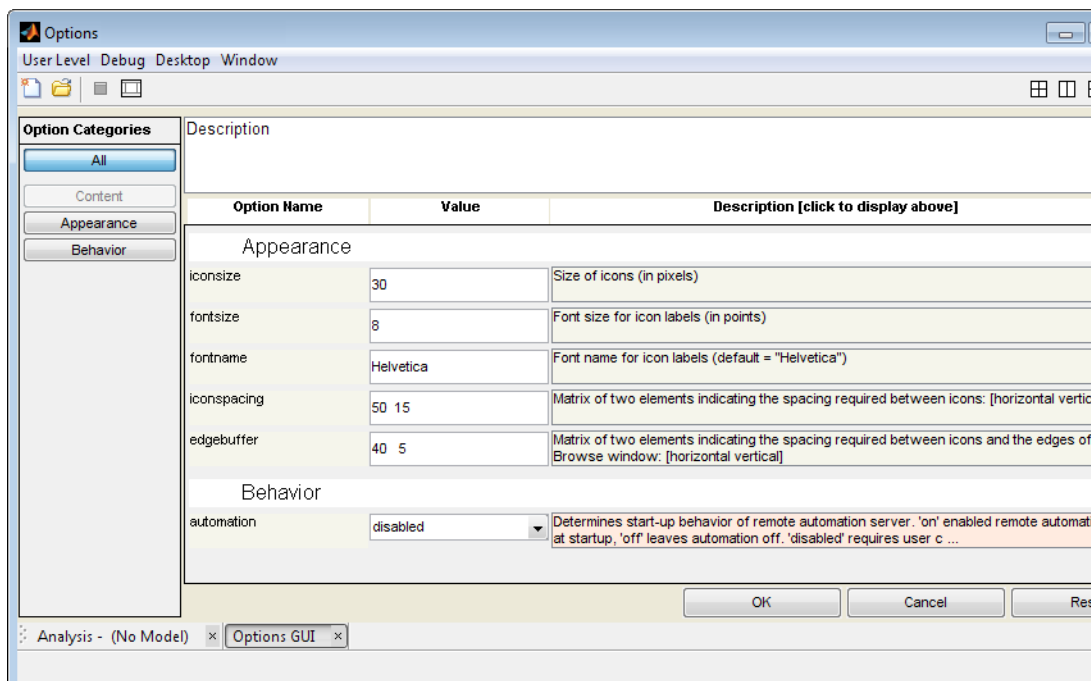
Note: For a detailed discussion about the Options dialog box, see “Options dialog box” on page 21.

Workspace Browser options affect the *display* properties for the Workspace Browser (such as icon size, font size, and icon font) and the *interactivity* properties for the Workspace Browser (such as single-click behavior versus double-click behavior and dragging functions).

1. On the Workspace Browser menu, click Edit > Options > Workspace Browser options.

The Options dialog box for the Workspace Browser window opens.

Figure 6-2: Options dialog box



2. Modify the value for any option, and then click OK.

The Options dialog box closes and you return to the Workspace Browser. Any changes that you made are effective immediately.

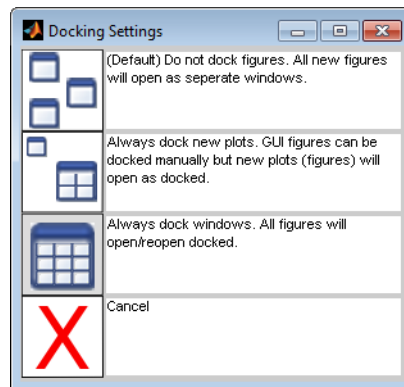
To specify the Window docking settings

By default, when you first open Solo, every data figure and interface in Solo is a *floating* window, which is a window that you can drag to any position on your desktop. You can also resize a floating window. You can select different Window docking settings to change the floating behavior of data figures, interfaces, or both.

1. On the Workspace Browser main menu, click Edit > Options > Window Docking Settings.

The Docking Settings dialog box opens. The first docking setting—All data figures and interfaces open as separate windows—is selected by default.

Figure 6-3: Docking Settings dialog box



2. To select a different Window docking setting, click the setting.

The Docking Settings dialog box closes and you return to the Workspace Browser. The new docking setting is effective after you close and reopen a window.

Chapter 6
Workspace Browser Preferences

Chapter 7: Importing Data into the Workspace Browser

To carry out an analysis, the first thing that you must do is import data for analysis. Two options are available for importing or loading data:

- You can import data directly into the Workspace Browser.
- You can import data directly into an Analysis tool.

Although both of these options are available to you, after you import data, certain functions or actions are easier to carry out in the Workspace Browser than within an Analysis tool.

Because Solo supports the importing of many different file types and the analysis of many different data types, the requirements for importing data into the Workspace Browser depend on the data source. For example, a single Excel file typically has multiple rows and multiple columns and therefore, typically contains all of the data that you need to import for analysis. An X,Y delimited text file, on the other hand, is usually analogous to a single row in an Excel file, and therefore, you might need to import multiple data files and assemble them into a single data object for analysis. Out of all the file types that you can import into Solo, the native MATLAB file format (a binary format) is the format that can be read the fastest by Solo and that requires the least disk space for storing. As a result, after you import a file, Eigenvector recommends that, regardless of the original file format, you save the imported data to the native MATLAB file format (i.e., a .mat file).

Note: The .mat file that is created is compatible with version 6.5 or later of MATLAB.

When you import data items into the workspace, be aware of the following:

- Although Solo places no restrictions on the number of items that you can import into the base workspace, memory is allocated to the loaded items. Having an excessive number of items loaded in the base workspace can limit the application's ability to carry out certain analyses.
- The steps for importing a .mat file are slightly different than the steps for importing other allowed file types. See:
 - [“To import a .mat file into the Workspace Browser” on page 32.](#)
 - [“To import a data file \(other than a .mat file\) into the Workspace Browser” on page 34.](#)

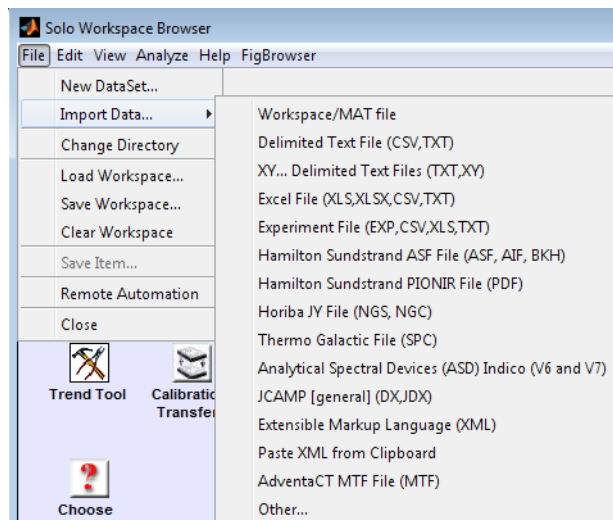
To import a .mat file into the Workspace Browser

Note: Although you initially click the “Import Data” option to select a .mat file for importing, the dialog boxes that open during the importing of a .mat file use the term “Load.” As a result, the term “Import” is typically used in reference to any file other than a .mat file, while the term “Load” is used in reference to a .mat file. The two terms, however, refer to the same action and can be used interchangeably.

1. On the main menu, click File > Import Data.

A list of available file types that you can import opens. The first option in the list is the Workspace/MAT file option.

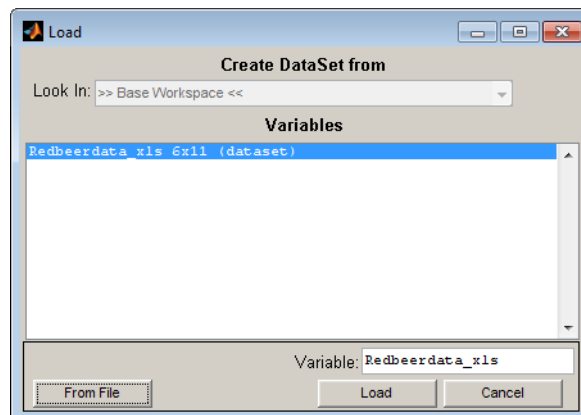
Figure 7-1: List of available file types



2. Click Workspace/MAT file.

The Load dialog box opens. By default, this dialog box references the *currently opened workspace*, and therefore, lists all of the items that are currently loaded in the workspace.

Figure 7-2: Load dialog box

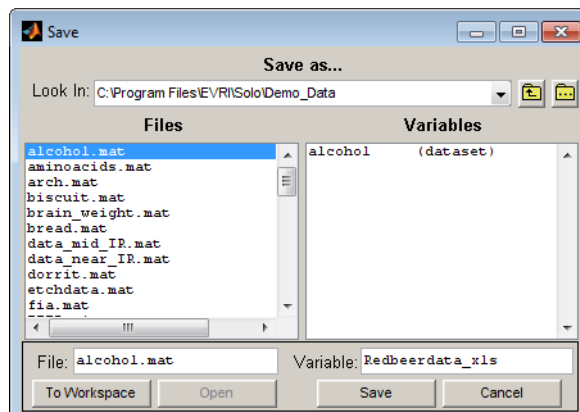


3. Click From File.

The button now shows From Workspace. The dialog box is refreshed to show the .mat file that was last loaded and all of the items contained in the .mat file.

Note: In MATLAB, an item is called a variable, and it is the type of data that can be stored in a .mat file. An item can be a DataSet, a matrix, a character array, and so on. Although this documentation uses the term “item,” the term “variable” is used on various windows and dialog boxes and in some lists in Solo. Multiple variables, DataSet objects, and so on can be stored in a single .mat file.

Figure 7-3: Save dialog box



4. If needed, in the Look In field, change to the directory from which you are importing the .mat file.
5. In the Files column, select the .mat file that you are importing.
6. Do one of the following:
 - If the .mat file contains a single variable, click Load.
 - If the .mat file contains more than one variable, then by default, the first variable in the list is selected. Click Load to import this variable, or select a different variable, and then click Load.
 - If the .mat file contains more than one variable, to import all variables, click All, and then click Load. Click Yes at the prompt to load all variables, and then click Load again.

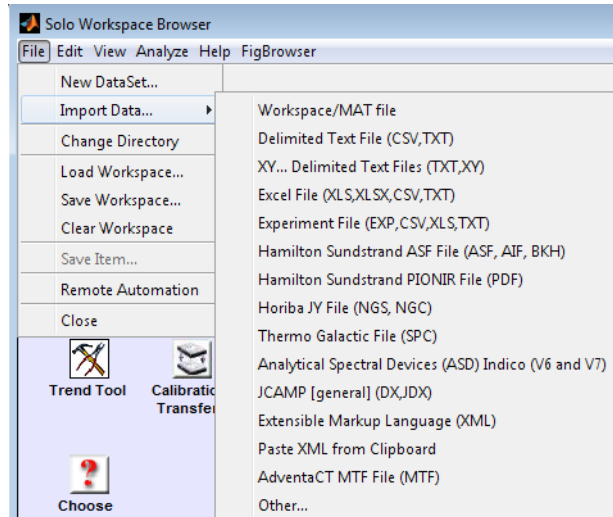
The selected data files are loaded into the Workspace Browser. After you import the data, different icons are displayed in the Workspace Browser for the different data types. You can save these data items to a workspace, and you can manipulate this data in the browser before you analyze it. See [“Icons in the Workspace Browser”](#) on page 37.

Note: If you select the All option through the Workspace Browser, then all of the variables are loaded into the base workspace. If you select the All option in an Analysis tool, then all of the variables are loaded into the base workspace. You are then pointed at the base workspace and prompted to select one of the variables that you just loaded.

To import a data file (other than a .mat file) into the Workspace Browser

1. On the main menu, click File > Import Data.
A list of available file types that you can import opens.

Figure 7-4: List of available file types



2. Select the type of file that you are importing.
3. In the Open <File Type> dialog box, scroll to and select the file that you are importing.
4. Click Import.

The data is imported into the Workspace Browser. After you import the data, different icons are displayed in the Workspace Browser for the different item types. You can save these data items to a workspace, and you can manipulate this data in the browser before you analyze it. See [“Icons in the Workspace Browser” on page 37](#).

5. Optionally, save the imported data to a .mat file. See [“To save imported data to a .mat file” on page 35](#).

Note: Remember, the native MAT file format is the format that can be read the fastest by Solo and that requires the least disk space for storing.

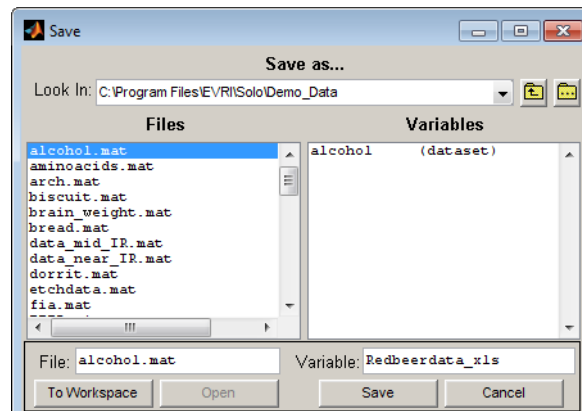
To save imported data to a .mat file

1. Do one of the following in the Workspace Browser:
 - Click the icon for the data that you are saving, and then on the main menu, click File > Save.
 - Right-click the icon for the data that you are saving and on the context menu that opens, click Save.

The Save dialog box opens. This dialog box shows the .mat file that was last loaded and all of the items contained in the .mat file.

Note: In MATLAB, an item is called a variable, and it is the type of data that can be stored in a .mat file. An item can be a DataSet, a matrix, a character array, and so on. Although this documentation uses the term “item,” the term “variable” is used on various windows and dialog boxes and in some lists in Solo. Multiple variables, DataSet objects, and so on can be stored in a single .mat file.

Figure 7-5: Save dialog box



2. Specify a location in which to save the data file and the name for the data file.
3. Click Save.






Chapter 7
Importing Data into the Workspace Browser

Chapter 8: Icons in the Workspace Browser

After you import or load items into the Workspace Browser, or analyze data, different icons are used to represent the different types of items and data. Also, after you import or load items into the Workspace Browser, you can save all of the items to a single .mat file that you can load into the base workspace. You can create as many .mat files as needed to support your work requirements, and then load any of these .mat files into the base workspace. In addition, after you import or load an item into the Workspace Browser, you can manipulate the item. (For example, you can rename the item, or you can open the item for viewing and editing.) Finally, the Workspace Browser is “drag and drop” enabled which means that you can drag one icon onto another icon or onto any other shortcut icon to manipulate or analyze the item. See:

- [“Item icons”](#) below.
- [“Saving, loading, and deleting items in the workspace”](#) on page 38.
- [“Manipulating items”](#) on page 39.
- [“Dragging and dropping items”](#) on page 42.

Item icons

Icon	Function
	DataSet icon—Indicates that the imported or loaded data is a <i>DataSet</i> , which is the object used in Solo for managing data.
	Matrix icon - Indicates that the imported or loaded data is a matrix (for example, a double array).
	Character Array icon - Indicates that the imported or loaded data is a character array.
	Model icon - Indicates that the item is a model that was created in one of the Analysis tools.
	Preprocessing icon—Indicates that the item is a set of preprocessing instructions created by the Preprocessing tool.

Saving, loading, and deleting items in the workspace

After you import or load items into the Workspace Browser, you can save all of the items to a single .mat file. You can create as many .mat files as needed to support your work requirements, and then load any of these files into the base workspace. You can also delete all items from the base workspace.

To save items to a .mat file

1. Import or load the required items into the Workspace Browser.
2. Optionally, manipulate the items as needed. For example, rename an item, modify the data for one or more items, and so on.
3. On the main menu, click File > Save Workspace.

The Save Workspace dialog box opens.

4. Enter a name for the file (by default, the Save as type is a .mat file, and you cannot change this), and then click Save.

All of the items in the base workspace are saved to a single .mat file.

To load saved items into the base workspace

When you load saved items, all of the items are loaded exactly as they were saved—they have the same names, they are loaded with all of their data, and so on.

1. On the main menu, click File > Load Workspace.
2. In the Load Workspace dialog box, scroll to and select the saved .mat file that you are loading, and then click Open.

To delete items from the base workspace

When you delete loaded items from a workspace, all of the items are deleted in a single step. Be aware, however, that any unsaved data items will be lost. If you want to delete items from a workspace, but not the data that an item contains, make sure to save all of the data items in the workspace first.

1. On the main menu, click File > Clear Workspace.
A dialog box opens, asking you if you want to delete all items from the workspace.
2. Click Yes to confirm the deletion.

Manipulating items

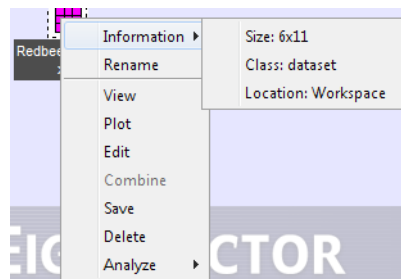
After you have imported or loaded data items into the Workspace Browser, a variety of options are available for manipulating the item, including:

- “Viewing information about the item” below.
- “Opening the item for viewing or editing.”
- “Dragging and dropping items” on page 42.

Viewing information about the item

You can right-click on an icon and on the context menu that opens, select from options for viewing information about the item, renaming the item, and viewing details about the item. For a data item, options are also available for plotting the imported data, editing the item, and analyzing the data. In addition, if multiple data items are selected, then an option to combine the data items is also available.

Figure 8-1: Context menu for an item in the Workspace Browser

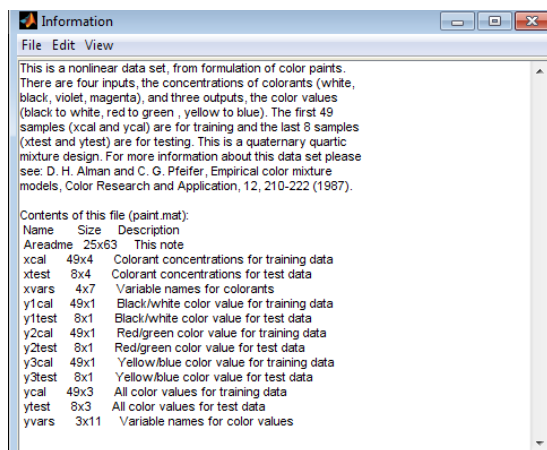


Opening the item for viewing or editing

Double-click an icon to open the item for viewing only or for editing.

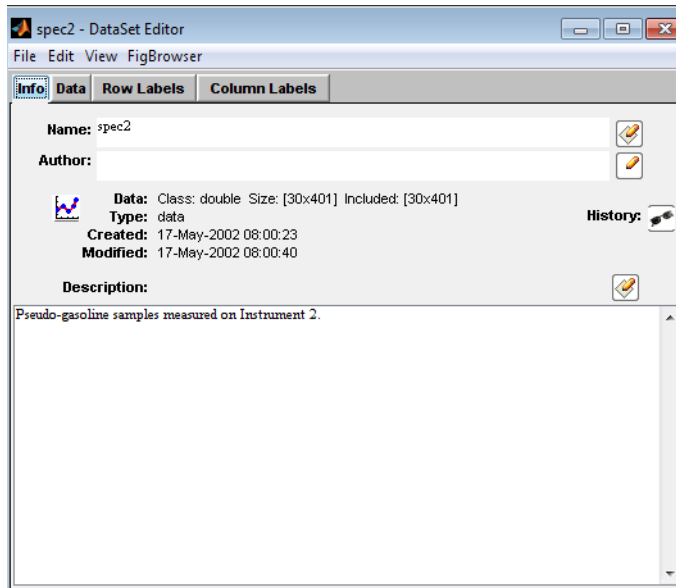
- If the item is not editable, when you double-click the icon for it, an Information dialog box opens. You can view information about the non-editable item in this dialog box.

Figure 8-2: Information dialog box



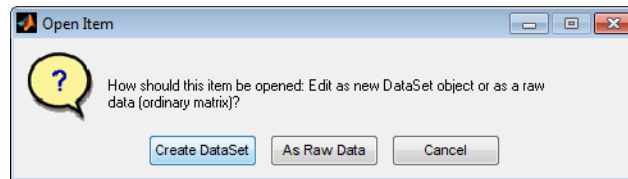
- If the item is a *DataSet*, when you double-click its icon, the DataSet Editor window opens. You can edit the DataSet (data, row labels, column labels, and so on) as needed in this window. (See “[DataSet Editor Window](#)” on page 45.)

Figure 8-3: *DataSet Editor window*



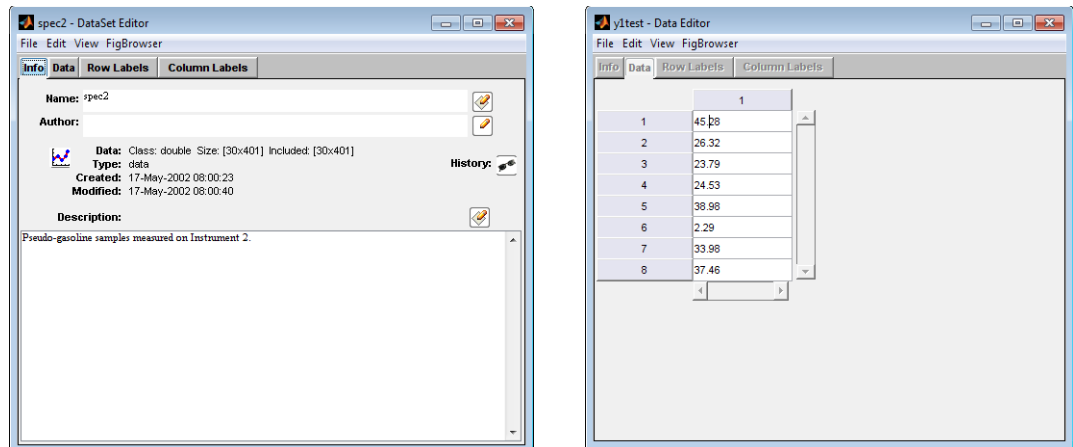
- If the item is not a DataSet, but another type of editable data, when you double-click its icon, the Open Item dialog box opens, asking you how you want to open the item—either as a new dataset or as a raw data (which means editing the data as a simple matrix without adding labels or other DataSet information).

Figure 8-4: *Open Item dialog box*



Because, in general, any data that is loaded into Solo must ultimately be converted to a DataSet object, you can click Create DataSet to proactively carry out this conversion; otherwise, you can click As Raw Data to edit the item “as is.” The same window opens whether you click Create DataSet or As Raw Data; however, as shown in [Figure 8-5 on page 41](#), if you click Create DataSet, all of the tabs are enabled for the DataSet Editor window while if you click As Raw Data, only the Data tab is open for the Data Editor window.

Figure 8-5: DataSet Editor window and Data Editor window

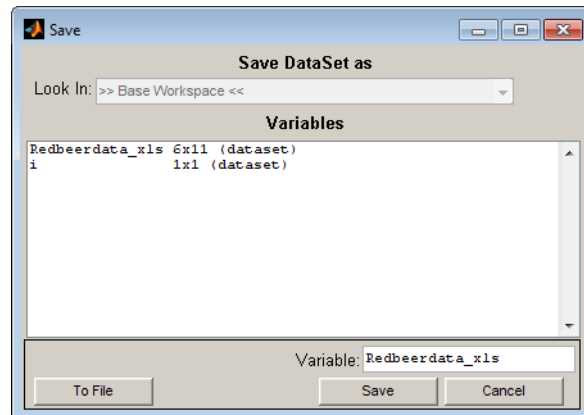


With some exceptions, if you edit a data item, you must explicitly request to overwrite the data item in the Workspace Browser with the changes. To save changes to a data item:

1. In the appropriate Editor window, edit the item as needed, and then on the Editor window menu, click File > Save.

The Save dialog box opens. The Variable name field is automatically populated with the name of the data item.

Figure 8-6: Save dialog box



2. Do one of the following:
 - Click Save to override the selected data item in the Workspace Browser with the modified data item.
 - In the Variable field, enter a new name for the data item, and then click Save to save the modified data item as a new item in the Workspace Browser.

Dragging and dropping items

- You can drag a data icon to a shortcut icon to open the Analysis window and analyze the data.
- You can drag a model icon to an shortcut icon to open the Analysis window to load a model, and optionally, apply it to new data.
- If the size of data items matches in at least one dimension, (either the same number of rows or the same number of columns), or if data items are identical in size, you can drag a data icon onto another data icon or onto an open Editor window to combine these two data items and create a single data item. You can repeat this step as many times as needed to combine all of the necessary data items.

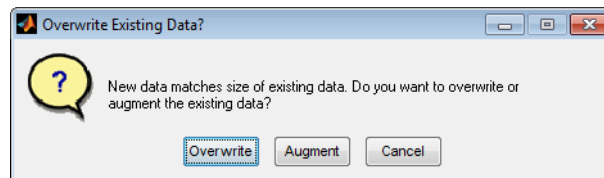
Note: You cannot join data items that do not match in at least one dimension.

Consider the following:

- DataSet item: A, 300 rows x 20 cols
- DataSet item: B, 200 rows x 20 cols
- DataSet item: C, 300 rows x 1 col
- DataSet item: A_copy, 300 rows x 20 cols

You can join A with B because these DataSets have the same number of columns, or you can join A with C because these DataSets have the same number of rows. For example, when you join A with B, you are given two options:

Figure 8-7: Overwrite existing data dialog box



You can overwrite A with the B data, or you can add the B data to A. In this case, the data is automatically joined as additional rows, and a 500 row x 20 column dataset is created.

Similarly, if you join A with C, you can overwrite A with the C data, or you can add the C data to A. In this case, the data is automatically joined as additional columns, and a 300 row x 21 column dataset is created.

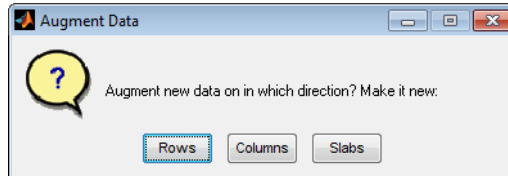
You can also join A with A_copy because these two data items are identical. You are again given two options for joining the data:

Figure 8-8: Overwrite existing data dialog box



You can overwrite A with the A_copy data, or you can add the A_copy data to A. If you choose to add the A_copy data to A, you have three options for joining the data:

Figure 8-9: Augment data dialog box



- You can join the data by rows. In this case, a 600 row x 20 column DataSet is created. The 300 new rows are considered as new samples.
- You can join the data by columns. In this case, a 300 row x 40 column DataSet is created. The 20 new columns are considered as new variables for the same samples.
- You can join the data as slabs. In this case, one DataSet is essentially placed behind the other to create a 300 column x 20 row x 2 DataSet. (You typically join data as slabs if the data is to be used in multi-way data analysis methods.)

Chapter 8
Icons in the Workspace Browser

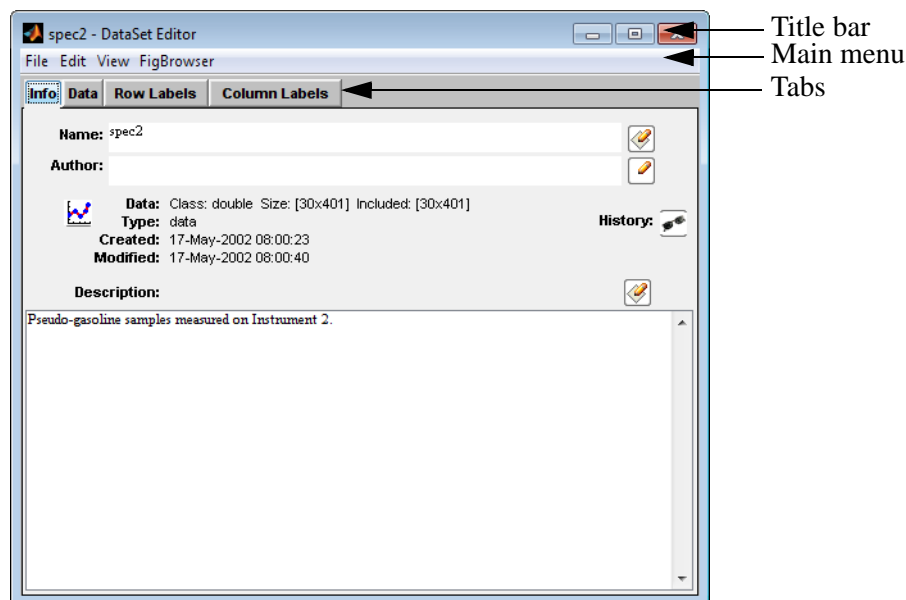
Chapter 9: DataSet Editor Window

The DataSet Editor window is the standard interface that you use for creating and managing a DataSet in Solo. Multiple options are available for opening the DataSet Editor window. You can:

- Right-click on an item in the Workspace Browser window and on the context menu that opens, click Edit.
- Drag an item in the Workspace Browser window onto the DataSet Editor icon.
- Double-click the DataSet Editor icon in the Workspace Browser window, and on the window's main menu, click File > Load or File > Import From.
- Right-click on a data component in an analysis window (for example, the X calibration control in the PCA analysis window), and on the context menu that opens, click Edit.
- With data loaded in an analysis window, (for example, the X calibration control in the PCA analysis window or the X and Y calibration control in the PLS analysis window), on the Solo main menu, click Edit > Calibration > X-Block Data or Edit > Calibration > Y-Block Data.

DataSet Editor window layout

Figure 9-1: DataSet Editor window



The DataSet Editor window has three major components—the title bar, the main menu, and the tabs.

- Title bar - The phrase “DataSet Editor” appears in the title bar at the top of the DataSet Editor window, as well as the standard Window Minimize, Maximize, and Close buttons. The name of the DataSet that is currently loaded in the DataSet Editor window also

appears in the title bar. For example, in [Figure 9-1](#) above, the DataSet named “conc” is currently loaded in the window.




Note: If you launch the DataSet Editor window any other way than from an analysis window, and then modify the DataSet in any way, an asterisk (*) is displayed next to the DataSet name in the title bar. The asterisk indicates that modifications to the data are pending. Before you can close the DataSet Editor window, you must answer a prompt about saving the modified data. If you launch the DataSet Editor window from an analysis window, any modifications that you make to the data are immediate (no asterisk is displayed next to the DataSet name in the title bar) and you can close the window without having to answer a prompt about saving the data.

- Main menu - The DataSet Editor main menu is set up in a standard Windows menu format with menu commands grouped into menus (File, Edit, View, and FigBrowser) across the menu bar. The Load and Import options on the File menu are identical to the options on the File menu on the Workspace Browser window and the Analysis window. You use these options to load or import data from the Workspace Browser or from a file. You use the File > Save Data option to save a DataSet to the Workspace Browser or to a file. You use the File > Export to export a DataSet to a .csv or .xml file.
- Tabs—The DataSet Editor window has four tabs—Info, Data, Row Labels, and Column Labels—each of which provides access to different content in the DataSet.

Info tab

When the DataSet Editor window opens, the Info tab is the active tab. (See [Figure 9-1 on page 45.](#)) The Info tab provides a high-level overview of the DataSet, including the DataSet name, the DataSet author, the data type and size in the DataSet for both included and excluded data, the DataSet creation date and time, the DataSet modification date and time, and a description of the DataSet.

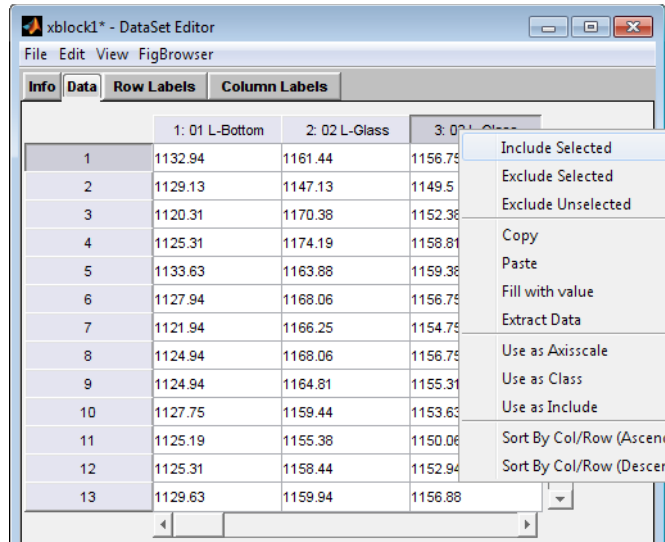
The Info tab is interactive:

- To edit the DataSet name, author, or description, click the Edit button  next to the appropriate field.
- To plot the included data in the DataSet, click the Plot button .
- To view the history of the DataSet, click the History button .

Data tab

The Data tab displays the data in the DataSet in a spreadsheet format.

Figure 9-2: DataSet Editor window, Data tab



The tab is interactive. You can:

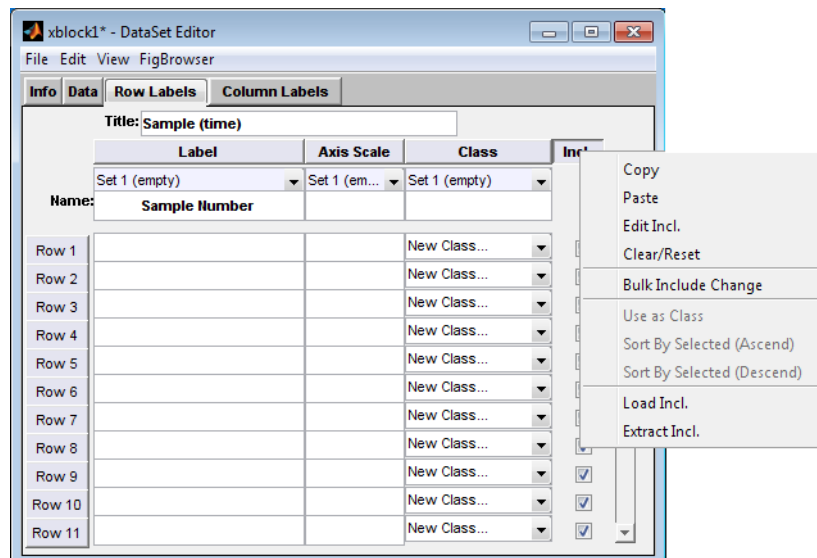
- Edit the data directly on this tab.
- Copy and paste rows and/or columns to and from other programs.
- Include and exclude rows and/or columns of data.
- Designate rows and/or columns as axis scales, classes, or the Include field. The results are reflected on either the Row Labels tab or the Column Labels tab.

All actions are available either from the Edit menu, or by right-clicking on a row or on a column header (as shown in [Figure 9-2](#) above) to open a context menu.

Row Labels tab/Column Labels tab

In a typical two-way DataSet, data mode 1 (the rows) represents the data samples and data mode 2 (the columns) represents the variables. The Row Labels tab and the Columns Labels tab—also known as the Mode Labels—provide access to the auxiliary “context” data for the Dataset, such as the labels for each sample, the axis scale, the data classes, and the Include status for data. (Multi-way data has additional Label tabs for each mode of the data.)

Figure 9-3: DataSet Editor window, Row Labels tab



You use the information in these fields for:

- Managing the data. (For example, the Include field indicates whether a given row or column is to be included in an analysis.)
- Plotting the data (For example, some correction algorithms plot against the axis scale of the columns.)
- Analyzing the data. (For example, classification algorithms use the information in the Class field to identify class assignments.)

A variety of options are available for specifying and working with the information for these fields. You can:

- Manually enter the information in each field.
- Assign a name to a field set to assist in identifying content.
- Create sets for loading multiple versions of a field into a single DataSet.
- Load the fields from files or variables in the base workspace (as long as the information that is being loaded is of the correct size.)

Note: When you load field information from one DataSet object (the *source* DataSet) into another DataSet (the *target* DataSet), the information is always loaded from the corresponding field and mode (row/column) of the source DataSet. To load information from a different mode, you must first Extract or Copy the contents from the source DataSet object,

and then Load or Paste the content into the target DataSet object. You can always load or paste information from a non-DataSet object, even when the information is from an external program such as Microsoft Excel.

- Copy and paste information to and from the Label field, the Axis Scale field, the Classes field, and the Include field.
- Edit multiple fields in a single step.

All actions are available either from the Edit menu, or by right-clicking on a row or on a column header (as shown in [Figure 9-3 on page 48](#)) to open a context menu.

Edit menu

The DataSet Editor Edit menu has many powerful data manipulation options. Some of the more commonly used options include the following:


- Transpose—Switches the rows and columns for a 2-way DataSet object.
- Exclude Data—Mark rows or columns as “Excluded.” (Also known as a “Soft Delete.”)
- Hard Delete Excluded—Permanently removes excluded data from a DataSet object.
- Exclude Excessive Missing—Automatically excludes rows or columns or multi-dimensional indices in which the number of missing data values exceeds the allowable amounts, as defined by the missing data replacement algorithm “mdcheck.” You use this option to remove samples or variables which do not have enough information to be used in modeling.
- Permute modes—Changes the order of the data modes.
- Concatenate—Provides the option of concatenating “old” data (data already loaded in the DataSet Editor) with “new” data (data that are either loaded or imported using the Load or Import From options on the DataSet Editor File menu). The old data and new data are compared and if they match in size in at least one dimension, (either the same number of rows or the same number of columns), the data can be concatenated. You can use this option to build a larger DataSet from several smaller DataSets.

Chapter 9
DataSet Editor Window

Chapter 10: Plot Controls Window

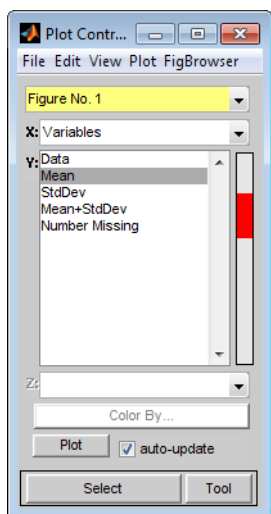
The Plot Controls window is the principal data visualization tool for Solo. It contains an extensive number of tools for labeling, manipulating, and publishing plots that are contained in a Plot window. (See “Plot window” on page 52.)

You can open the Plot Controls window in one of two ways:

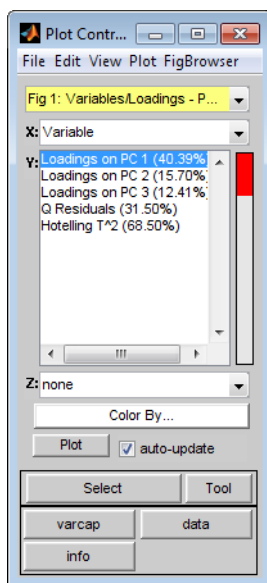
- You can right-click on any set of data (for example, on a set of data in the Workspace Browser window or on the X calibration control after you have loaded data into the control in an Analysis window) and on the context menu that opens, click Plot to open the Plot Controls window. (See “Data plotting options” on page 55, “Data selection and editing options” on page 55, and “Other options” on page 56.)
- You can click on an active Analysis window toolbar button that is specific for a plot (for example, an active Plot Eigenvalues button ).

Note: If you open the Plot Controls window by clicking on an Analysis window toolbar button, then the options that are available on the Plot Controls window are specific to the analysis method and the plot that is generated. Analysis-specific plot options are not discussed in this chapter.

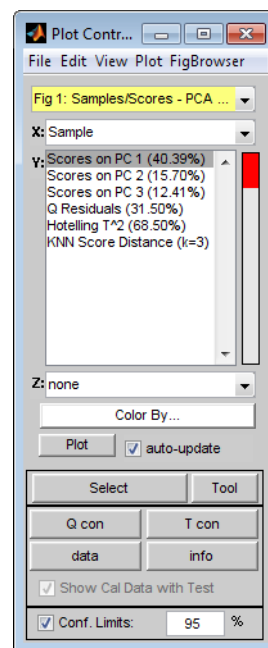
Figure 10-1: Comparison of Plot Controls window



Plot Controls window opened from context menu



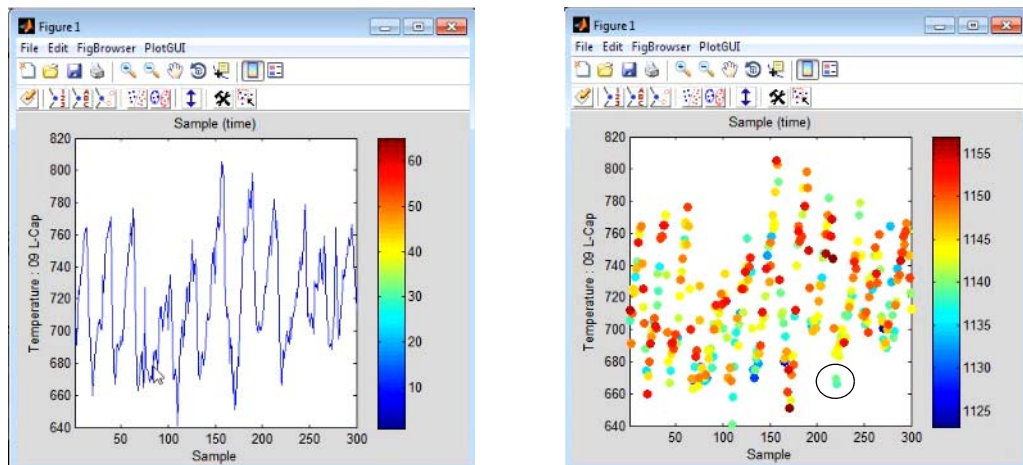
Plot Controls window opened from Plot Loads toolbar icon after building a PCA model with three PCs



Plot Controls window opened from Plot Scores toolbar icon after building a PCA model with three PCs

Regardless of how you open the Plot Controls window, two options are common to the window— auto-update and Color. Auto-update is selected by default. With this option selected, a plot in a Plot window is automatically updated after you make a change to a plot. If you clear this option, you must click Plot to manually update a plot after you make a change to it. You use the Color By option to implicitly superimpose the response of one variable onto the plot of another. For example, in [Figure 10-2](#) below, the figure shown on the left is the plot of the response of variable 9 for all of the data samples. The figure shown on the right is the plot of also the plot of the response of variable 9 for all of the data samples; however, the color of the data points is based on the response of variable 15 for all of the data samples. As you can see in the figure on the right, a selected sample has a value of approximately 660 for variable 9, but it has a value in the 1100s for variable 15.

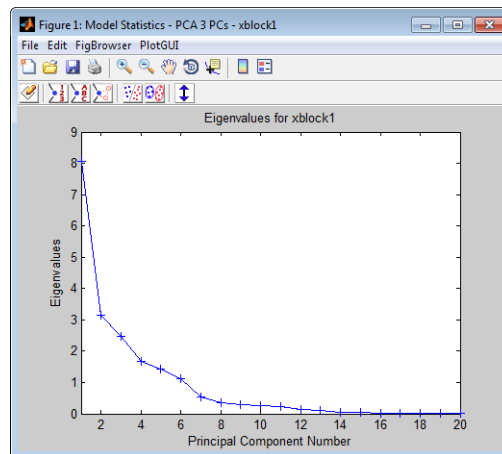
Figure 10-2: Color By example



Plot window

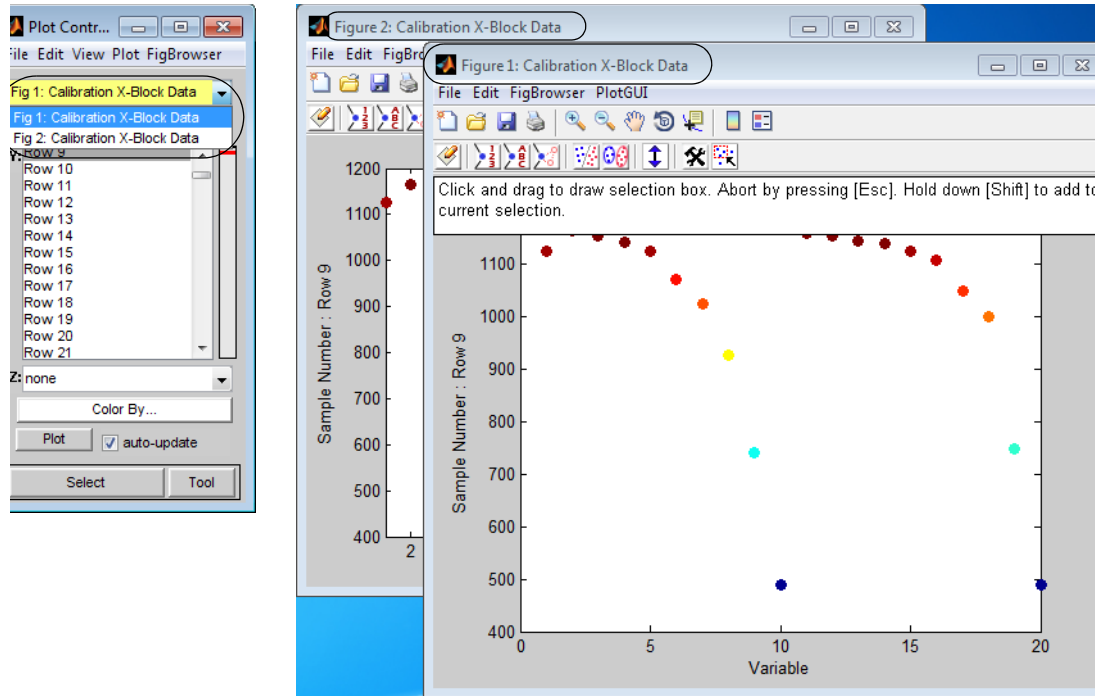
Most plots created in Solo are contained in a Plot window. For example, [Figure 10-3](#) shows an Eigenvalues plot in a Plot window.

Figure 10-3: Eigenvalues plot in a Plot window



You can create multiple plots during an analysis session, and the plots are automatically numbered as they are created. The number is displayed in the Title bar of the Plot window. Most of the plots that you create during an analysis session are under the control of a single Plot Controls window. The Figure Selector dropdown list in the Plot Controls window contains a list all of the plots created during an analysis session and that are under control of the window. At any time, you can select a figure from this list to make it the active plot for the session.

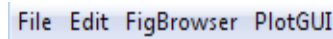
Figure 10-4: Example of multiple plots automatically numbered



All Plot windows, regardless of the plot that they contain, have the same main menu and the same toolbars.

- Main menu—The main menu is set up in a standard Windows menu format with commands grouped into menus (File, Edit, FigBrowser, and PlotGUI) across the menu bar. This menu contains variety of options for working with the plot in the Plot window and modifying the appearance of the plot in the Plot window. For example, the Edit menu contains options for editing the plot axes and the plot font settings.

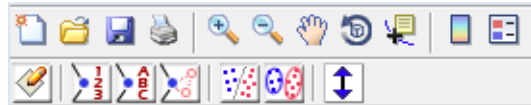
Figure 10-5: Plot window main menu



Chapter 10
Plot Controls Window

- **Toolbar**—The Plot window has two toolbars. The top toolbar contains the buttons for common Matlab plotting tools. The bottom toolbar contains an option for viewing and opening the data in the DataSet editor as well as options for changing the appearance of the plot. Place your mouse pointer over a toolbar button to open tooltip text for the button.

Figure 10-6: Plot window toolbar



Note: For assistance with using any of the common Matlab plotting tools, see:

- Data cursor - http://www.mathworks.com/access/helpdesk/help/techdoc/creating_plots/f4-44221.html
- Zooming - http://www.mathworks.com/access/helpdesk/help/techdoc/creating_plots/f4-44425.html
- Panning - http://www.mathworks.com/access/helpdesk/help/techdoc/creating_plots/f4-44519.html
- Rotating - http://www.mathworks.com/access/helpdesk/help/techdoc/creating_plots/f4-44601.html
- Annotating - http://www.mathworks.com/access/helpdesk/help/techdoc/creating_plots/f0-37626.html

Data plotting options

If you open the Plot Controls window from the Plot option on the context menu, the Plot Controls window opens with the same set of options for plotting summary statistics (mean, standard deviation, and mean \pm standard deviation) for all of your data samples. The default plot that is generated is the mean response across all of your samples for all of your variables. You can, however, plot your data anyway that you want.

- To generate an overlay plot of all of your samples for all of your variables, click Data.
- To generate a plot of the standard deviation across all of your samples for all of your variables, click StdDev.
- To generate a plot of the mean response across all of your samples for all of your variables plus or minus one standard deviation, click Mean+StdDev.
- To generate a plot that shows the number of missing observations for each variable, click Number Missing.
- To select any combination of samples for plotting against all variables, on the Plot Controls window main menu, click Plot > Rows. CTRL-click to select multiple and/or non-contiguous samples.
- To select any combination of variables for plotting against all samples, on the Plot Controls window main menu, click Plot > Columns. CTRL-click to select multiple variables. (This plot is useful for viewing trends for one or more variables for all samples.)
- To change the scale of a plot, on the Plot Controls window menu, click View > Auto Y-Scale and select the option that makes the most sense for your given set of data.
- To review the co-linearity of variables, plot a variable on the X axis as the function of another variable on the Y axis.

Data selection and editing options

The Plot Controls window and the toolbar on the Plot window contain a number of options for selecting data and for modifying the data that you are viewing. If you choose to plot columns of data, then the resulting plot is for every row (sample) for the selected columns. If you choose to plot rows of data, then the resulting plot is a plot for every column (variable) for the selected rows.

Note: Unless specifically stated otherwise, all menu options discussed below refer to the options on the main menu of the Plot Controls window.

- To select only part of the data that you are viewing for plotting, do one of the following:
 - Click Select on the Plot Controls window, and then click and drag your cursor around the data points to select them.
 - On the Plot Controls dialog box, click Tools, and select your tool of choice, and then on the Plot Controls dialog box, click Make Selection, and then click and drag your cursor around the data points to select them.

The color of the selected data points is changed, not only in the currently active plot, but also, in any other open plots that contain the samples.

- To select one or more of the classes that are displayed in the current plot, click Edit > Select Class.
- To select all items that are currently displayed in the plot, click Edit > Select All.
Note: Items that are excluded are not displayed in a plot, and therefore, are not selected.
- To select only excluded items, click Edit > Select Excluded.

After you have made a selection, a variety of options are available for working with the selected data.

Note: These options are enabled only if the data is editable.

- To exclude (mark as “Do Not Use”) the selected items from the data, click Edit > Exclude Selection.
- To include *only the selected data items*, click Edit > Include Only Selection.
- To include items again which were previously excluded and are now selected, click Edit > Include Selection.

Note: You might need to click View > Excluded Data to select the excluded items, or click Edit > Select Excluded.

- To change the class of the selected data, click Edit > Set Class of Selection.
- To make the selected data “Missing,” (replace the values of the data with Not-a-Number values), click Edit > Make Selection Missing.

Note: This is a permanent action. You cannot undo it.

Other options

In addition to the options for plotting data, and for selecting and editing data, the Plot Controls window contains options for displaying information about other DataSet fields.

Note: Unless specifically stated otherwise, all menu options discussed below refer to the options on the main menu of the Plot Controls window.

- To view the data for the currently selected/active plot in a table, click View > Table. The table is displayed in a Plotted Data window. You can click Edit > Copy to copy the data table, and then use standard menu or keyboard commands to paste the copied table into a word processing or presentation application.

Note: The Plotted Data window is an independent window that is not linked to the Plot Controls window.

- To view data (samples or variables) that you have excluded from analysis in the currently selected/active plot, click View > Excluded Data. The excluded data is displayed on the active plot with the same symbol as the included data, but in a color that is several shades lighter to distinguish it from the included data.
- To generate a probability plot for data, right-click on the plot and on the context menu that opens, click Probability Plot > Best Fit (automatic). Solo automatically determines the best distribution plot for the data and displays the plot in a Figure window. To manually

generate a different distribution plot, click Distribution on the Figure window's main menu, and then select a different plot type.

Note: The Figure window is an independent window that is not linked to the Plot Controls window.

- To open a text box that displays critical parameters for a regression model including the RMSEC and the RMSECV, right-click on a plot for the regression model. Click Show on Figure to display this text box on the plot in the Plot window.
- If you have multiple samples or variables plotted in a single Plot window, in lieu of adding a legend to the plot, you can add text that identifies each plot individually. Right-click on each plot in the Plot window, and on the context menu that opens, click Identify Curve.
- To exclude raw data (samples or variables) before analysis, right-click on the plot in the Plot window and on the context menu that opens, click Exclude Curve. The excluded sample or variable is marked with a double arrow in the Plot Controls window.
- To open the plotted data in the DataSet Editor, click File > Edit Data. (See [“DataSet Editor Window” on page 45.](#))
- To duplicate the currently selected/active plot, click View > Duplicate Figure. The duplicated plot is linked to the original plot—whatever samples or variables that you select in the original plot are automatically selected in the duplicate plot.
- To generate a separate, standalone view of the currently selected/active plot, click View > Spawn Static View.

Note: The Static View function is useful for creating snapshots of all your plots during data analysis for before and after comparison purposes. The static plot is a copy of the currently selected/active plot that is contained in an independent Figure window that is not linked to the Plot Controls window.

- To export the currently active/selected plot to Microsoft PowerPoint or Microsoft Word, click Export Figure, and then click To Microsoft Power Point or To Microsoft Word as appropriate. If PowerPoint or Word is not open, Solo opens the application and places the figure in a slide in a new PowerPoint presentation or in a new blank Word document. If PowerPoint or Word is open, Solo places the figure as the next slide in the currently active PowerPoint presentation or at the insertion point in the currently active Word document.

Note: This option is available only for a Windows operating system.

- To copy the currently selected/active plot to your computer's clipboard, click Edit > Copy Figure. Use standard menu or keyboard commands to paste the copied figure into an application that allows for the pasting of graphics.
- To copy the data for the currently selected/active plot in a tabular format to your computer's clipboard, click Edit > Copy Plotted Data. Use standard menu or keyboard commands to paste the copied figure into an application that allows for the pasting of tabular data.

Note: The Edit > Copy Plotted Data option is essentially a single step shortcut for the two step approach of View > Table and then Edit > Copy.

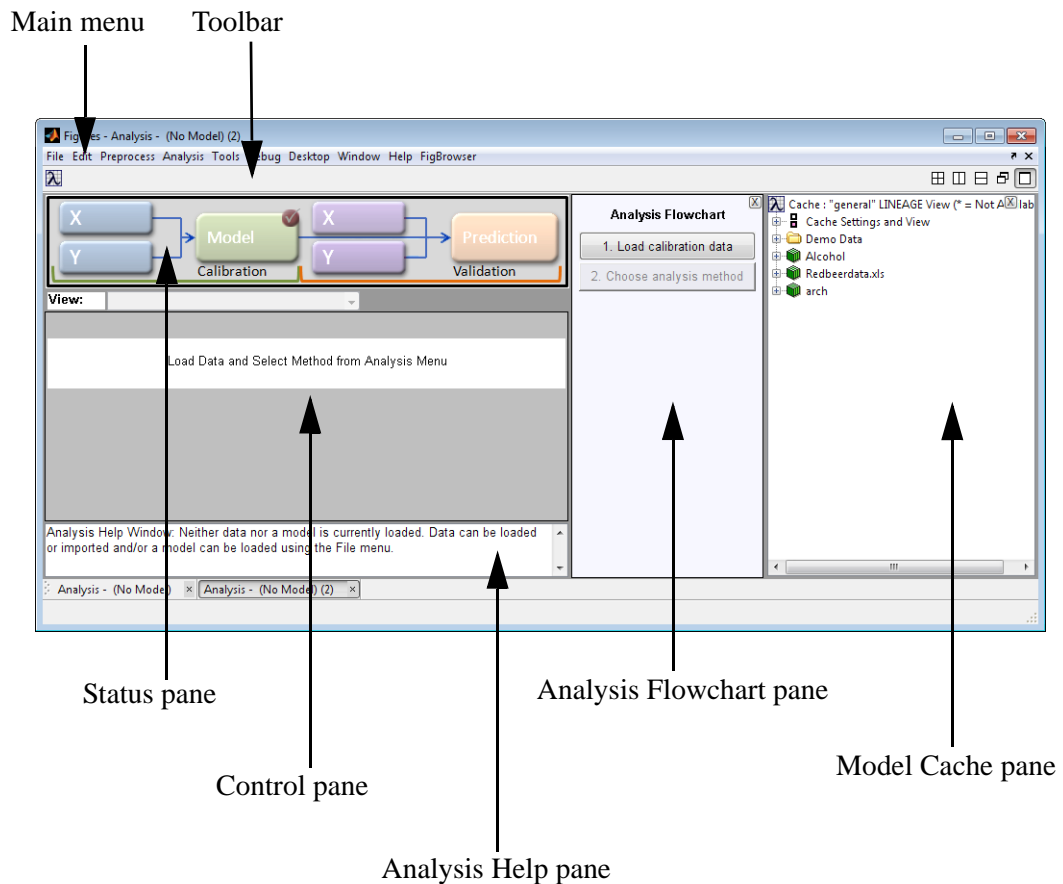
Chapter 10
Plot Controls Window

Chapter 11: Analysis Window

The Analysis window serves as the core interface to the Solo data modeling and analysis functions. You create your models in an Analysis window, apply models in this window, and also analyze and explore the models in this window. The Analysis window has seven major components. These components are the:

- “Analysis window main menu” on page 61.
- “Analysis window toolbar” on page 63.
- “Analysis window Status pane” on page 65.
- “Analysis window Control pane” on page 67.
- “Analysis window Help pane” on page 69.
- “Analysis window Flowchart pane” on page 71.
- “Analysis window Model Cache pane” on page 73.

Figure 11-1: Analysis window layout

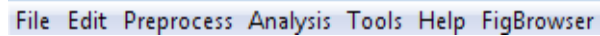


Chapter 11
Analysis Window

Analysis window main menu

The Analysis window main menu is set up in a standard Windows menu format with commands grouped into menus (File, Edit, View, Analyze, Help and FigBrowser) across the menu bar. Some of these commands are available in other areas of the application.

Figure 11-2: Analysis window main menu



File Edit Preprocess Analysis Tools Help FigBrowser

Edit menu

The Edit > Options menu contains the following options:

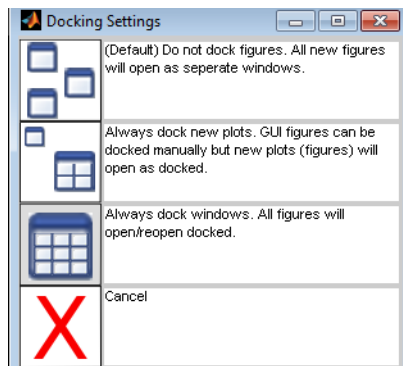
- Method Options - Opens the Options dialog box which lists options for controlling the settings that are specific to the currently selected analysis method such as confidence limits settings, algorithm selections, and constraints. Some analysis methods have more setting options than others. See [“Options dialog box” on page 21](#).
- Analysis GUI Options - Opens the Options dialog box which lists options for the currently opened Analysis window, including the maximum number of factors displayed, the font size for the Control pane, and the display in the Analysis Flow Chart pane. See [“Options dialog box” on page 21](#).
- Model Cache Settings - Allows you to turn model caching on and off and adjust various other aspects of the model cache. See [“Analysis window Model Cache pane” on page 73](#).
- Default Plots - Opens the Default Plots dialog box in which you can adjusting default Scores plots for given methods.
- Window Docking Settings - Opens the Window Docking Settings dialog box which you use to adjust the docking behavior of your Solo windows. See [“To specify the Window docking settings” on page 62](#).
- Preferences (Expert) - Opens a Preferences (Expert) window which you use to override all values for all Solo options. Because there is no checking for validity of settings when you use this tool, you must be very cautious using it. If problems occur, try resetting the default values using “Factory Default.”

To specify the Window docking settings

By default, when you first open Solo, every data figure and interface in Solo is a *floating* window, which is a window that you can drag to any position on your desktop. You can also resize a floating window. You can select different Window docking settings to change the floating behavior of data figures, interfaces, or both.

1. On an Analysis window main menu, click Edit > Options > Window Docking Settings.
The Docking Settings dialog box opens. The first docking setting (All data figures and interfaces open as separate windows) is selected by default.

Figure 11-3: Docking Settings dialog box

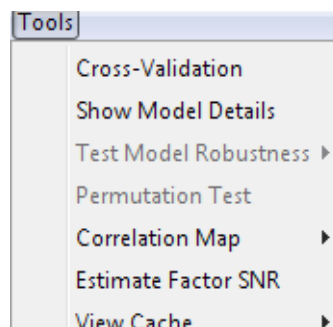


2. To select a different Window docking setting, click the setting.
The Docking Settings dialog box closes and you return to the Analysis window. The new docking setting is effective after you close and re-open a window.

Tools menu

The Tools menu contains a number of tools which you can use to investigate and gather information about the currently loaded data and to assess a model's performance.

Figure 11-4: Tools menu






See:

- [“Cross-Validation Tool”](#) on page 107.
- [“Model Robustness Tool”](#) on page 111.
- [“Correlation Map Tool”](#) on page 113.

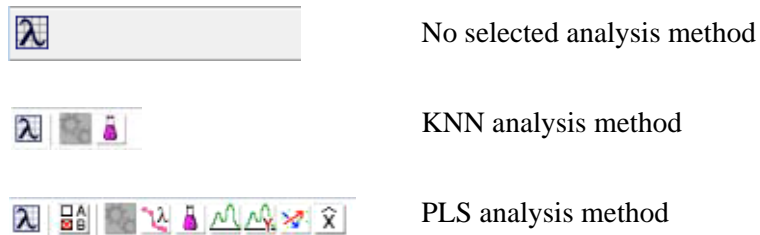
Analysis window toolbar

For any analysis method, the Analysis window toolbar always contains the following toolbar buttons:

- A Workspace Browser button .
- A Plot scores and sample statistics button .
- An Edit Analysis Methods Options button . This button opens the Options dialog box, which lists options for controlling the settings that are specific to the currently selected analysis method such as confidence limits settings, algorithm selections, and constraints. Some analysis methods have more options than others. See [“Options dialog box” on page 21](#).

The Analysis window toolbar is updated dynamically with other toolbar buttons based on the selected analysis mode. These toolbar buttons carry out actions or open tools that produce plots and other visual aids that assist you in examining a model. Place your mouse pointer over a toolbar button to open tooltip text for the button.

Figure 11-5: Example of various Analysis windows toolbars

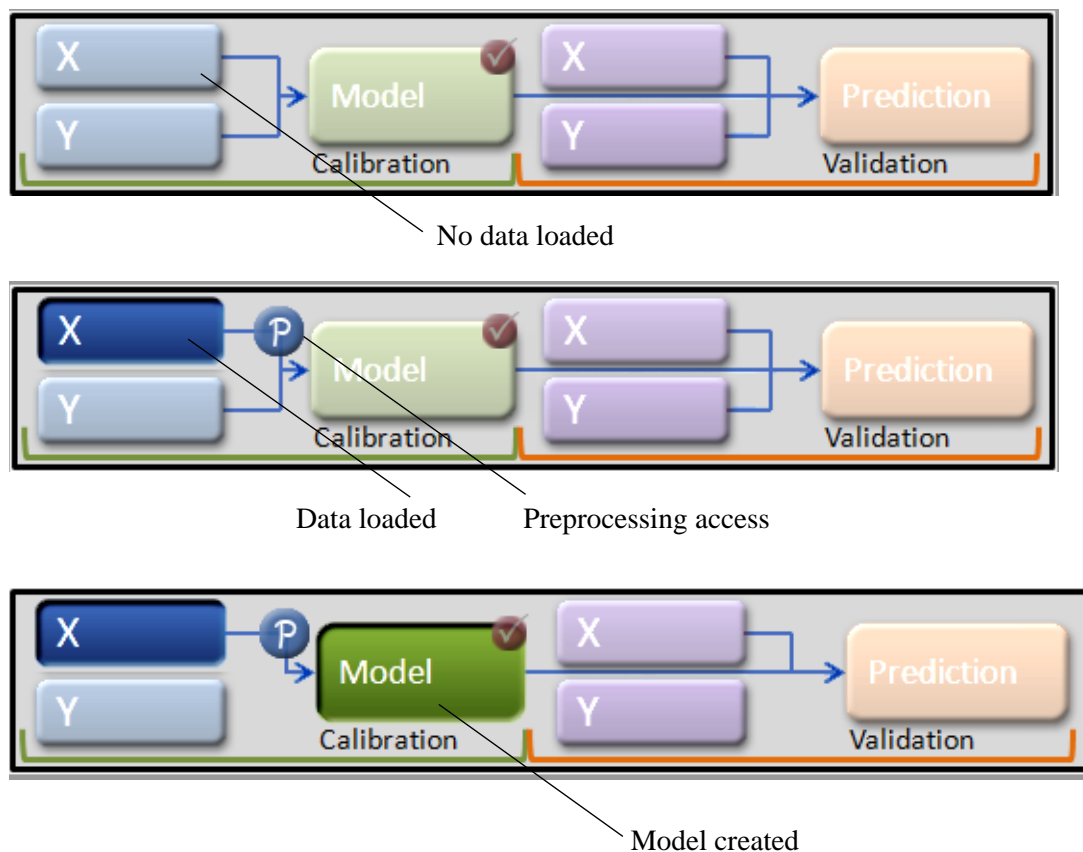


Chapter 11
Analysis Window

Analysis window Status pane

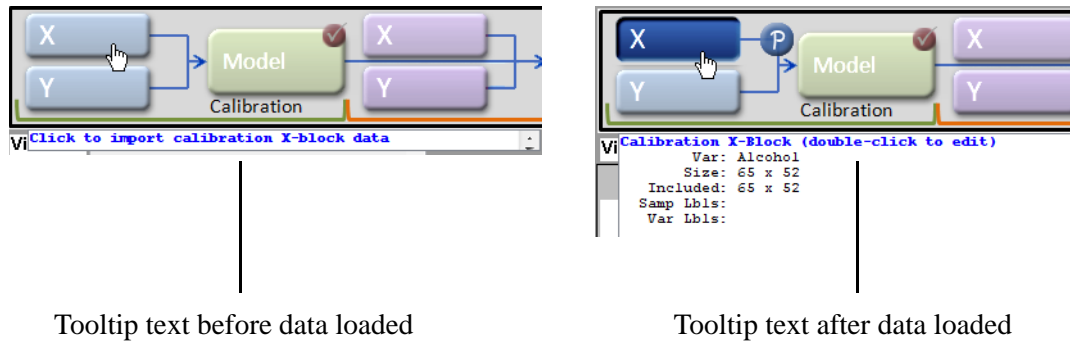
The Status pane is both informational and interactive. The appearance of the pane visually indicates the status of the analysis. It has an initial appearance before you load data into the Analysis window, it has a different appearance after you load data into the Analysis window, and it has yet another appearance after you create an item from the loaded data. For example, as shown in Figure 11-6, the X calibration control (the primary location for loading data in an Analysis window) has one appearance before you load calibration data and it has another appearance after you load calibration data. In addition, a “P” appears after the X control and before the Model control, which you can use to access preprocessing methods for the loaded data. After you create a model based on the loaded data, the Status pane has yet another appearance.

Figure 11-6: Different appearances of the Status pane



In addition, just as the appearance of the pane varies based on the analysis status, the information that is displayed for a control varies based on the analysis status. For example, before you start an analysis, place your mouse pointer on the X calibration control to view instructions about working with the control. After you load data, place your mouse pointer on the control again to view not only information about the loaded data, but also, different instructions about working with the control.

Figure 11-7: Information that is displayed in the Status pane

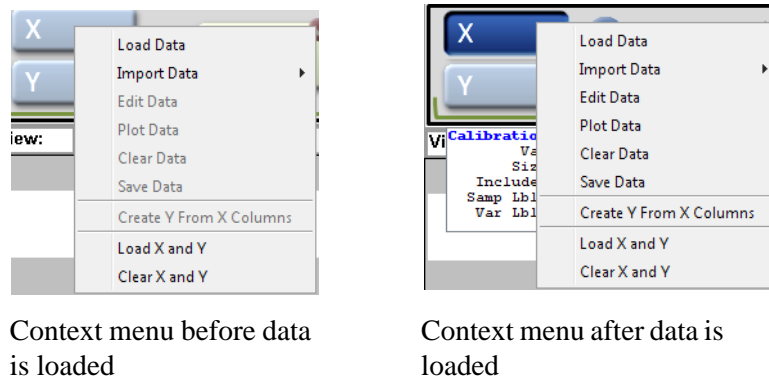


Tooltip text before data loaded

Tooltip text after data loaded

Every control in the Status pane is both left-clickable and right-clickable. What happens after you click on a control depends not only on the control itself (is it an X or Y control or is it a Model or Prediction control), but also, on the type of click (left or right) and the analysis status. For example, if you left click an empty X calibration control, then the Import dialog box opens, and you must select a file to import. If you left click or right click the X calibration control after you have loaded data, then a menu opens with options for manipulating (loading, importing, editing, plotting, and so on) the data. Again, as shown in Figure 11-8, the options that are enabled on the menu depend on the analysis status.

Figure 11-8: Status pane context menu



Context menu before data is loaded

Context menu after data is loaded

Note: All of the actions that are available in the Status pane are also accessible through the Analysis window menus and toolbar, but it is typically easier to work directly in the Status pane.

Analysis window Control pane

The Control pane contains two distinct parts—a row of tabs and the Control pane itself. The Control pane contains controls that are used in manipulating the model settings such as number of factors in a model or other parameters. In many cases, the Control pane also displays the Percent Variance Captured or other statistical information for the model. For some analysis methods (such as MCR or PARAFAC), the Control pane displays the variance that was captured for only the included components. For other analysis methods (such as PCA, PLS, or PCR), the Control panel displays the variance that was captured for both the included and excluded components. The row of tabs is located directly above the Control pane. The tabs in the tab row are analysis-specific controls that determine what is displayed in the Control pane.

Figure 11-9: Analysis window Control pane

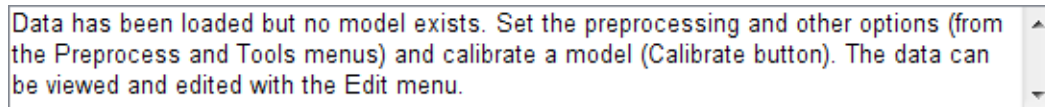
Percent Variance Captured by PCA Model			
Principal Component	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	8.08e+000	40.39	40.39
2	3.14e+000	15.70	56.09
3	2.48e+000	12.41	68.50 <- Suggested
4	1.66e+000	8.32	76.82
5	1.44e+000	7.20	84.02
6	1.12e+000	5.62	89.64
7	5.46e-001	2.73	92.37
8	3.48e-001	1.74	94.11
9	2.82e-001	1.41	95.52
10	2.68e-001	1.34	96.86
11	2.27e-001	1.13	97.99
12	1.48e-001	0.74	98.73
13	1.11e-001	0.55	99.28
14	5.88e-002	0.29	99.58
15	3.64e-002	0.18	99.76
16	1.73e-002	0.09	99.84

Chapter 11
Analysis Window

Analysis window Help pane

The Analysis Help pane is a text description of the current analysis status. It also provides general guidelines for what your next analysis steps should be based on the current analysis status. The Analysis pane can also provide warnings when the settings you have selected or the model results are unusual and require your attention. The pane is colored in red or yellow depending on the severity of the warning. You can dismiss a warning by clicking on the warning text and indicating whether to dismiss the warning once, or always (whenever the warning happens).

Figure 11-10: Analysis window Help pane

A screenshot of a software window's Help pane. The text is displayed in a light blue font on a white background. The text reads: "Data has been loaded but no model exists. Set the preprocessing and other options (from the Preprocess and Tools menus) and calibrate a model (Calibrate button). The data can be viewed and edited with the Edit menu." To the right of the text is a vertical scrollbar with a small upward-pointing arrow at the top and a downward-pointing arrow at the bottom.

Data has been loaded but no model exists. Set the preprocessing and other options (from the Preprocess and Tools menus) and calibrate a model (Calibrate button). The data can be viewed and edited with the Edit menu.

Chapter 11
Analysis Window

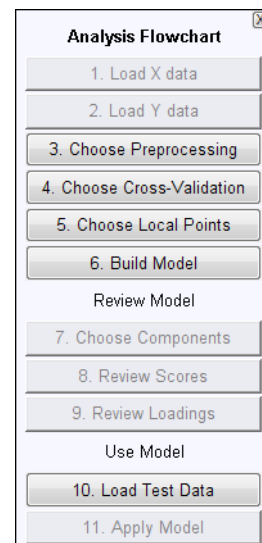
Analysis window Flowchart pane

While the Status pane provides a graphical cue for the sequence of steps in the modeling process, the Analysis window Flowchart pane provides step by step instructions for using a particular analysis method. Each step is a button that you can click to interactively carry out the analysis according to the listed steps. The available steps in this pane are dynamically updated based on the last completed step. For example, as shown in [Figure 11-11](#), the Analysis Flowchart pane for a PLS regression initially contains only two steps—Load X data and Load Y data. If you were to click Load X data, then the Import dialog box would open and on this dialog box, you would select the X data for importing. You would then click Load Y data and on the Import dialog box, select the Y data to load. After you load the Y data, the Analysis Flowchart pane is automatically updated with additional steps for continuing with the analysis.

Figure 11-11: Analysis window Flowchart pane



Analysis Flowchart pane before
X and Y data are loaded



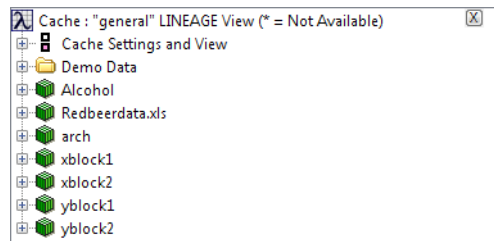
Analysis Flowchart pane after
X and Y data are loaded

Chapter 11
Analysis Window

Analysis window Model Cache pane

Data that is used to create models and predictions during data analysis as well as all of the models and predictions created from this data is automatically cached. After a model is created, the model and the data used to create it are added to the cache. If new data is loaded and a prediction is made, this new data and the prediction are added to the cache. Every time you edit data and create a new model or prediction from the data, each version of the modified data and the corresponding model or prediction are added to the cache. The Model Cache pane displays this *complete* history of all your cached items—all of the data that was used to create models and predictions during data analysis as well as all of the models and predictions that were created from this data—in a hierarchical view.

Figure 11-12: Analysis window Model Cache pane



Note: By default, the Model Cache pane is turned on when Solo first opens. You can close the Model Cache pane by clicking the Close (“X”) button in the upper right hand corner of the pane. To turn the Model Cache on again, on the Tools menu on the Analysis window, click View Cache, and then select a viewing option. Models and data are cached even if the Model Cache pane is not open.

Multiple options are available to you for changing the Model Cache view, managing your cache, and manipulating cache items. See:

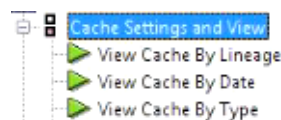
- [“Model Cache pane view.”](#)
- [“Cache management” on page 75.](#)
- [“Manipulating cached items” on page 76.](#)

Model Cache pane view

The options to change the Model Cache pane view are found in two locations:

- In the Model Cache pane itself under Cache Settings and View
- On the Tools menu (the View Cache option) for the Analysis windows

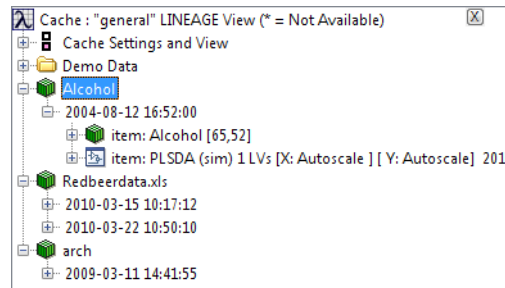
Figure 11-13: Options to change the Model Cache pane view



You can select from three different organizational views:

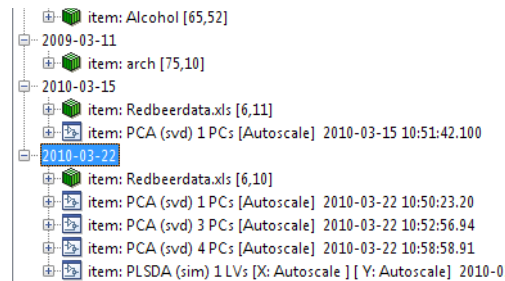
- View Cache by Lineage—In the View Cache by Lineage view, the cache items are sorted by the *parent data item*. Under the parent data item, the child items (modified data, model, and predictions) are sorted by the timestamps at each modification point.

Figure 11-14: View Cache by Lineage view



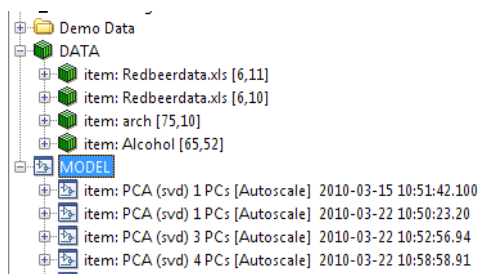
- View Cache by Date—In the View Cache by Date view, all cached items are simply sorted by creation date in chronological order.

Figure 11-15: View Cache by Date view



- View Cache by Type—In the View Cache by Type view, all cached items are simply sorted by object type—data, model, or prediction.

Figure 11-16: View Cache by Type view



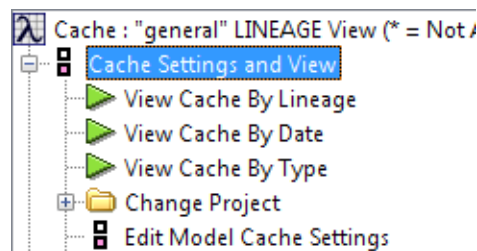
Cache management

Several options are available to you for managing your cache. You can modify all of these options in the Options dialog box.

Note: For a detailed discussion about the Options dialog box, see “Options dialog box” on page 21.

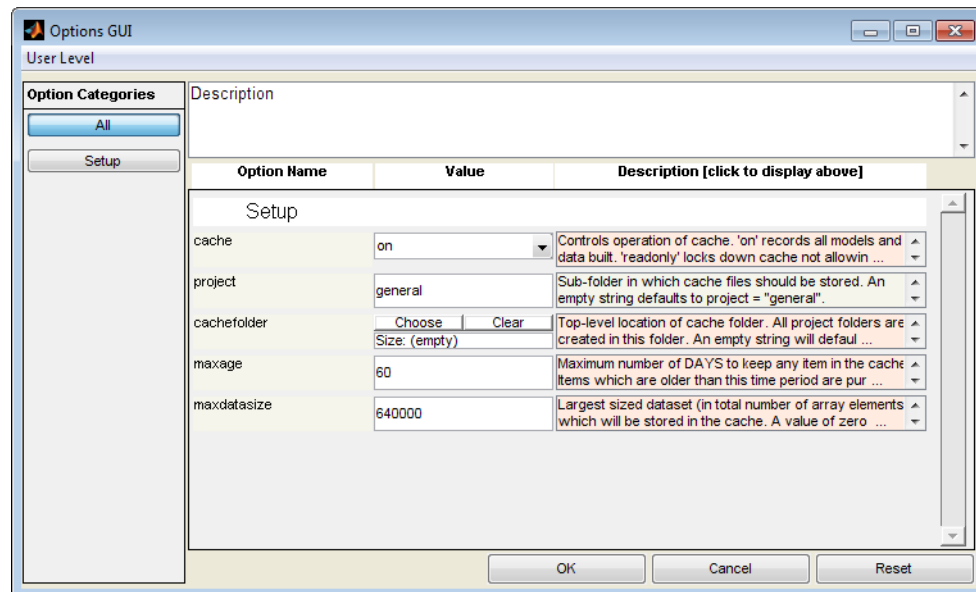
1. Do one of the following:
 - In the Model Cache pane, expand the Cache Settings and View option, and under this option, click Edit Model Cache Settings.

Figure 11-17: Cache Settings and View options



- On the Analysis window menu, click Edit > Options > Model Cache Settings. The Options dialog box for the Model Cache pane opens.

Figure 11-18: Options dialog box



2. Modify the value for any option, and then click OK.

Note: See the table on the next page for a list of the available options and their descriptions.

The Options dialog box closes and you return to the Model Cache pane. Any changes that you made are effective immediately.

Option	Description
cache	Controls the operation of the cache. “On” records all data, models, and predictions. “readonly” locks down the cache and prevents recording to the cache and clearing of the cache.
cachefolder	Specifies the folder in which all of your cached information is stored. By default, the cached items that do not exceed the maxdatasize value (see maxdatasize below) are stored automatically in your system’s TEMP folder and remain in this folder as long as they do not exceed the maxage value. (See maxage below.) For all operating systems other than Macintosh (a “Mac”), this means that even if you close Solo (or crash and reboot your computer, or just reboot your computer), the cached items are immediately available when you restart Solo or your computer. Note: If you are running Solo on a Mac, the first time that you cache an item, you are prompted to choose a folder other than the TEMP folder. Because the TEMP folder is sometimes deleted when you restart a Mac, if you are running Solo on a Mac, you should specify a different setting for your cachefolder.
project	Specifies an optional subfolder within the cachefolder for holding the cached items. The default name for project is “general.” You can use the project setting to separate cached items among different work projects. Note: No option is available in Solo to delete a project folder after you create it. You can delete old project folders only by manually locating the project folder on your computer and deleting the folder. Note: You can also change value for the project setting using the Change Project option under Cache Settings and View in the Model Cache pane.
maxage	The maximum number of days an item is saved in the cache. An item that is older than this value is permanently deleted from the cache. Note: See “Manipulating cached items” below for information about saving cached items to a location other than the cachefolder.
maxdatasize	Specifies the maximum size of a data item (in total number of table elements) that can be stored in the cache.

Manipulating cached items

You can right-click on an entry for a cached item Model Cache pane to open a context menu that contains options for manipulating the item.

Menu Option	Description
Load Item	Loads the selected cached item into the Analysis window. By default, if no data is loaded into the Analysis window, then the cached item is loaded into the X calibration component; otherwise, you are prompted to override the loaded data or load the data to another location. Note: You can also double-click a cached item to load it into the Analysis window.
Show Item	Opens and displays the selected item in a separate window—either the DataSet Editor window or the Model Reader window.
Save Cached Item As	Save the cached item to the workspace or to a file.
Open in New Window	Opens the cached item in a new Analysis window.
Rename	Rename the cached item.

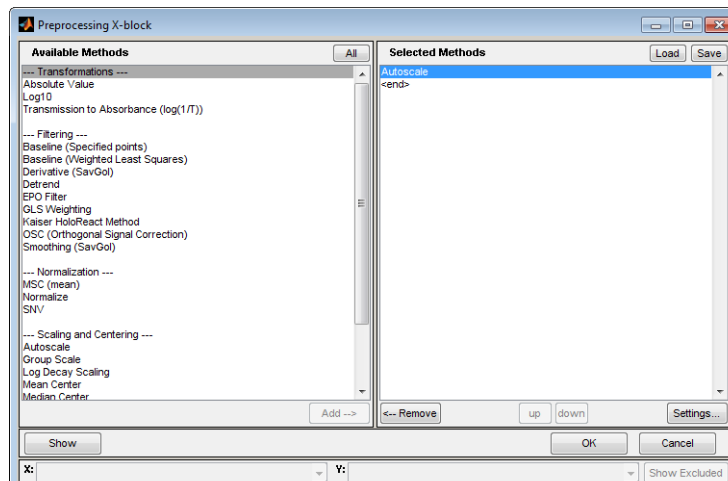
Chapter 12: Preprocessing Methods



Data preprocessing describes any type of processing procedures that are performed on raw data to prepare it for *another* processing procedure and ultimately, analysis. Preprocessing linearizes the relationships among the variables in your DataSet and removes extraneous sources of variation that are of no interest to the analysis. A variety of preprocessing methods are available in Solo. This section describes the basic steps for setting up preprocessing rules for an analysis and verifying that the rules that you have set up are as you want them. Specific information about the different preprocessing methods that are available (such as the purpose of each method and when and how to use a specific method) is beyond the scope of this section.

Preprocessing window

You use the Preprocessing window to specify the preprocessing methods that you want to carry out for your data and in what order. The available methods are grouped by type (Filtering, Normalization, and so on) in the Available Methods (left) pane of the window. The methods that you select for preprocessing your data are displayed in the Selected Methods (right) pane. The methods are carried out in the order in which they are listed in the pane. The default selected method is Autoscale.

Figure 12-1: Preprocessing window



1. To open this window, do one of the following:
 - In the Analysis window Flowchart pane, click Choose Preprocessing.
 - In the Status pane, click the Preprocessing icon .
 - In the Status pane, right-click the Preprocessing icon , and then click Preprocessing > Custom.
 - On the Analysis window main menu, click Preprocess > X-block or Y-block > Custom.

Chapter 12 Preprocessing Methods

2. In the Available Methods pane, select the method by which you want to preprocess your data, and then click Add.

Note: Some preprocessing methods, for example, the Savitzky-Golay method, require you to specify values for method-specific parameters before you can add the method to the Selected Methods pane.

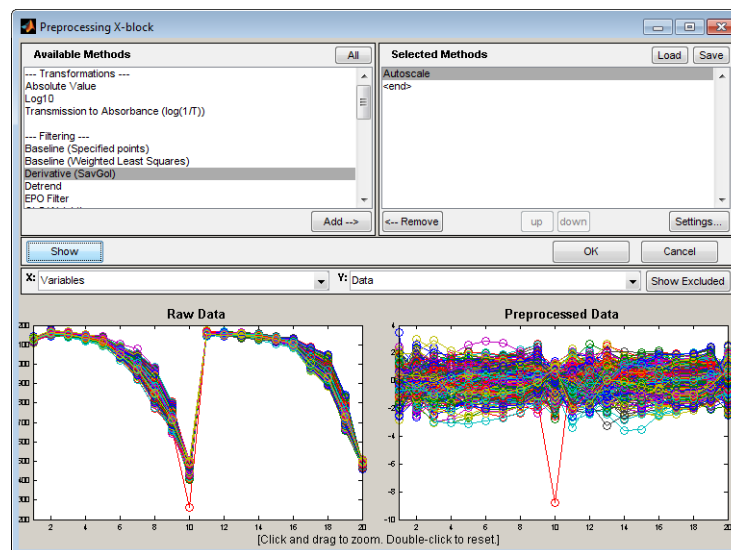
3. Repeat Step 2 until you have selected all of the necessary preprocessing methods.


Note: Regardless of the other preprocessing steps selected, Autoscale or Mean Centering should typically be the final preprocessing step in the series.

4. Optionally, after you have selected all of the necessary preprocessing methods, you can do one or more of the following:
 - To change the order in which the selected preprocessing methods are to be carried out, select a method, and then click Up or Down as needed.
 - To remove a preprocessing method from the Selected Methods list, select the method, and then click Remove.
 - To change the parameters values for a method, select the method, and then click Settings to open a dialog box in which you modify the settings.
 - To show the effect of preprocessing on your data, click Show. The Preprocessing window is updated with two panes. The Raw Data pane (the lower left pane) shows your raw data plotted as a function of all of the samples against all of the variables. The Preprocessed Data pane (the lower right pane) shows the effect of the preprocessing on the raw data.

Note: Figure 12-2 shows preprocessing using *default preprocessing methods* on a 200 x 30 DataSet for a simple PCA model.

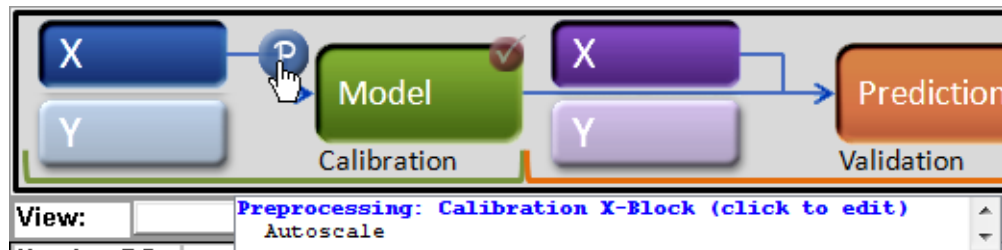
Figure 12-2: Preprocessing using default preprocessing methods on a 200 x 30 DataSet for a simple PCA model



- To zoom in on a region of the raw data, place your mouse pointer in the Raw Data pane. The cursor changes to a Zoom In icon . Drag your cursor around the region of interest. A box is formed around the area that being reduced for viewing. The x axes on the two panes are linked, so as you change the focus in the Raw Data pane, the focus is changed to the same region in the Preprocessed Data pane.
- To reset the view to the original view, double-click in the Raw Data pane, or right-click in the Raw Data pane to open a context menu with options for Zooming Out, Resetting to the Original View, and other options.

After you have selected preprocessing methods for your data, you can point your mouse cursor on the Preprocessing icon for a control in which you have loaded data. Tooltip text opens with preprocessing information.

Figure 12-3: Tooltip text with preprocessing information



Chapter 12
Preprocessing Methods

Chapter 13: Analysis Phases

The Analysis window serves as the core interface to the Solo modeling and analysis functions. You create your models in an Analysis window, apply models in this window, and also analyze and explore the models in this window. Three phases are required to completely carry out modeling and analysis in the Analysis window—the [Calibration phase](#), the [Test and Validation phase](#), and the [Model Application phase](#).

Calibration phase

The Calibration phase consists of model building and exploratory analysis. In this phase, which affects only the Calibration side of the Status pane, you must load data into the X calibration control. This data is referred to as *x block data*, and it is a set of multivariate measurements on your data samples. Some analysis methods also require you to load data into the Y calibration control. This data is referred to as *y block data* and it is a set of secondary or reference measurements on the same data samples. During analysis, you identify any patterns or trends in the data, and any other information that you consider relevant, for example, any relationships that might exist between the x data and the y data, and use this information to build a model. See [“Building the Model in the Calibration Phase” on page 83](#).

Test and Validation phase

The Test and Validation phase consists of applying the model that you built in the Calibration phase to your *validation* data, which is data with known physical and/or chemical characteristics. In this phase, which affects the Validation side of the Status pane, you must load data into to the X validation control, and if applicable, the Y validation control. As is the case in the Calibration phase, the data that you load into the X control is referred to as *x block data*, and it is a set of multivariate measurements on your data samples. Likewise, the data that you load into the Y control is referred to as *y block data* and it is a set of secondary or reference measurements on the same data samples. You use this validation data to confirm that the model that you built captures valid patterns and trends in the data. You test and validate the model by applying it to the validation data and verifying that the test results are acceptable. For example, PCA analysis is typically used for pattern recognition. A correctly built PCA model, therefore, can identify the instances for which this pattern has been broken, such as a failure in material that does not meet specifications. During the Test and Validation phase of a PCA model, some of the validation data samples should meet specifications and some of the validation data samples should be “out of spec.” A well-built PCA model will identify or flag these “out of spec” samples. If the test results are acceptable, you can continue to the next phase, the Model Application phase. If the test results are not acceptable, you must return to the Calibration phase. See [“Applying the Model in the Test and Validation Phase” on page 103](#).

Model Application phase

The Model Application phase consists of applying the tested and verified model to *new* data, which is data with unknown characteristics, and therefore, the results of applying the model cannot be known in advance. If your test results, however, were acceptable in the Test and Validation phase, then the results from the Model Application phase are also likely accurate. For example, a correctly built PCA model that was successfully tested and validated in the Test and Validation phase should identify “out of spec” samples during the Model Application phase.

Chapter 14: Building the Model in the Calibration Phase

Regardless of the analysis method, building a model in the Calibration phase consists of a series of the same general steps, with the second and third steps being iterative, until you are satisfied with your model. These steps are:

1. Loading the calibration data and building the initial model. See [“Loading the calibration data and building the initial model”](#) on page 83.
2. Changing the number of components or factors that are to be retained in the model and recalculating the model. See [“Changing the number of components”](#) on page 85.
3. Examining the model and refining the model by excluding certain samples and/or variables to enhance the model performance. See [“Examining and refining the model”](#) on page 86.
4. After you are satisfied with the model, you can then do one of the following:
 - Save the model and use it at a later date.
 - Load validation and test data and apply the model immediately.

Note: Decomposition and Clustering analysis methods require only x block data for model building in the Calibration phase. Regression analysis methods require both x block data and y block data. Classification analysis methods require x block data with classes in either X or Y. For simplicity and brevity, this section describes model building during the Calibration phase using *default preprocessing methods* for a simple PCA model; however, all of the *general* information in this section is applicable for all analysis methods.

Note: Although this section describes model building using default preprocessing methods, remember, for most analyses, it is critical to select the appropriate preprocessing methods for the data that is being analyzed. To review detailed information about preprocessing, see [“Preprocessing Methods”](#) on page 77.

Note: To review a detailed description of the Calibration phase, see [Chapter 13, “Analysis Phases,”](#) on page 81.

Loading the calibration data and building the initial model

You have a variety of options for opening an Analysis window and loading data. Because these methods have been discussed in detail in other areas of the documentation, they are not repeated here. Instead, a brief summary is provided with a cross-reference to the detailed information. Simply choose the method that best fits your working needs.

- To open an Analysis window:
 - In the Workspace Browser, click the shortcut icon for the specific analysis that you are carrying out.
 - In the Workspace Browser, click Other Analysis to open an Analysis window, and on the Analysis menu, select the specific analysis method that you are carrying out.

Chapter 14
Building the Model in the Calibration Phase

- In the Workspace Browser, drag a data icon to a shortcut icon to open the Analysis window and load the data in a single step.

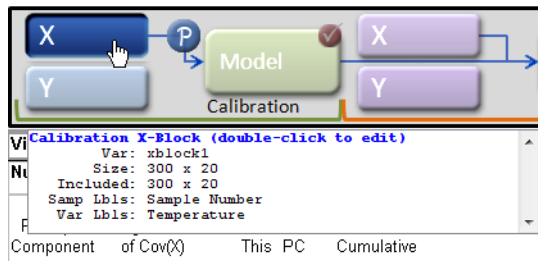
Note: For information about working with icons in the Workspace Browser, see [“Icons in the Workspace Browser” on page 37](#).

- To load data into an open Analysis window:
 - Click File on the Analysis window main menu to open a menu with options for loading and importing calibration data.
 - Click the appropriate calibration control to open the Import dialog box and select a file type to import.
 - Right-click the appropriate calibration control to open a context menu with options for loading and importing data.
 - Right-click on an entry for a cached item Model Cache pane to open a context menu that contains options for loading the selected cached item into the Analysis window.

Note: For information about the data manipulation options on the context menu, see [“Icons in the Workspace Browser” on page 37](#) or [“Importing Data into the Workspace Browser” on page 31](#). For information about loading items from the Model Cache pane, see [“Analysis window Model Cache pane” on page 73](#).

Also, remember that after you load data into a calibration control, you can place your mouse pointer on the control to view not only information about the loaded data, but also, different instructions about working with the control. In [Figure 14-1](#), data has been loaded into the X calibration control for a PCA analysis.

Figure 14-1: Example of loaded data in the X calibration control for a PCA analysis



After you have opened the Analysis window and loaded the calibration data, you then calculate the initial model. To calculate the initial calibration model, you can do one of the following:


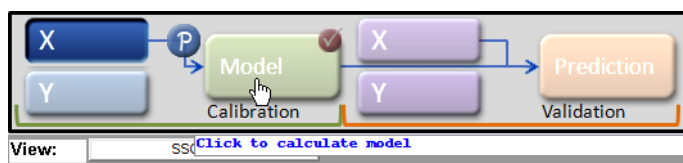
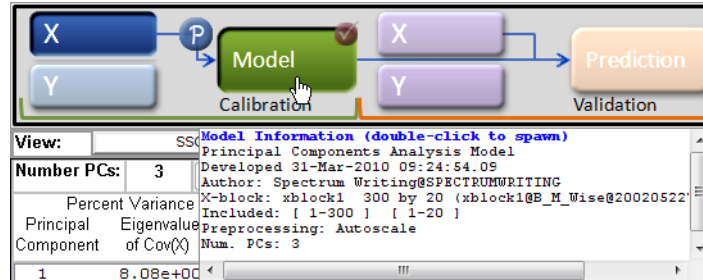
- On the Analysis window toolbar, click the Calculate/Apply model icon .
- Click the Model control.

Figure 14-2: Clicking the Model control in the Analysis window




After the initial model is calculated, you can place your mouse pointer on the Model control to view general information about the model. To view detailed information the model, right-click on the Model control and on the context menu that opens, select Show Model Details.

Figure 14-3: Showing model details in the Analysis window



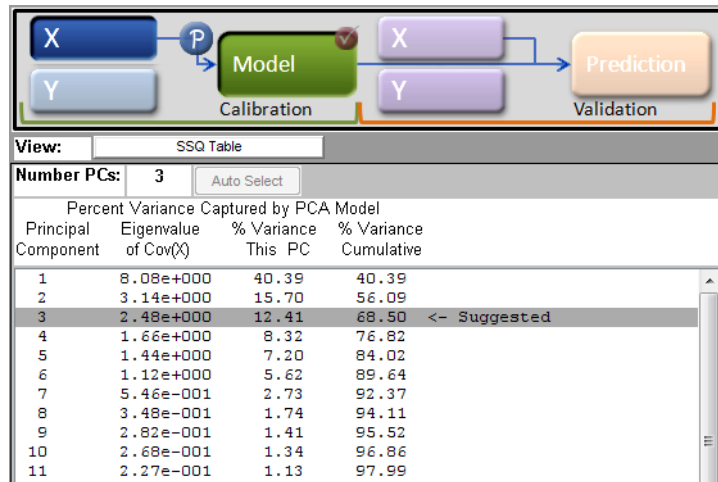
Changing the number of components

For analysis methods which use factors or principal components, you can choose a different number of components or factors to retain in the model and then recalculate the model. To choose a different number of components or factors:

1. Click on the appropriate row in the Control panel.
2. Recalculate the model by doing one of the following:
 - On the Analysis window toolbar, click the Calculate/Apply model icon .
 - Click the Model control.

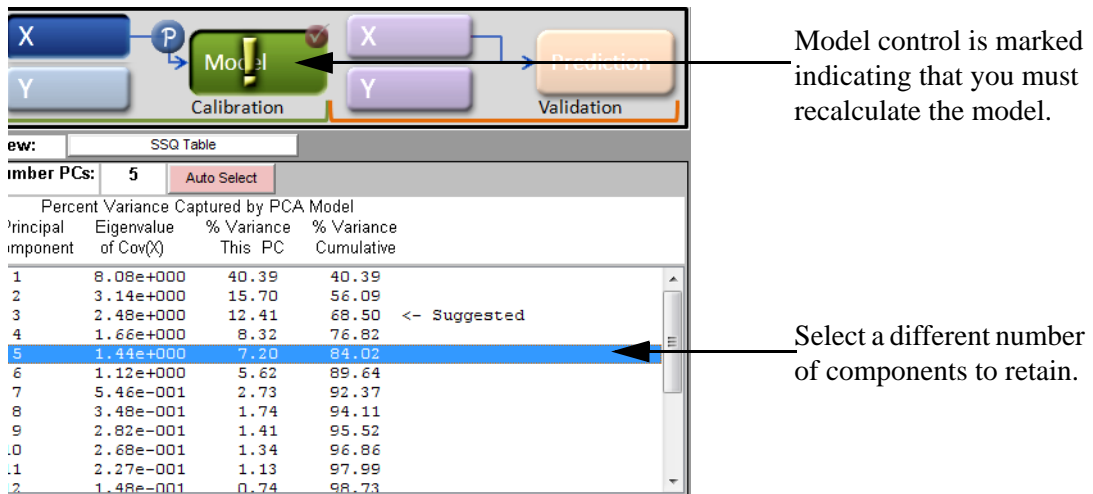
Note: By default, the maximum number of principal components or factors that you can retain in a model is 20. You can change this value in the Analysis options settings for the Edit menu. For example, Figure 14-4 shows an initial model calculated for a PCA analysis with the suggested value for the number of components to retain set to three.

Figure 14-4: Initial model calculated for a PCA analysis with number of suggested components = 3



After you select a different number of components or factors to retain, the Model control is marked with an Exclamation icon indicating that you must recalculate the model.

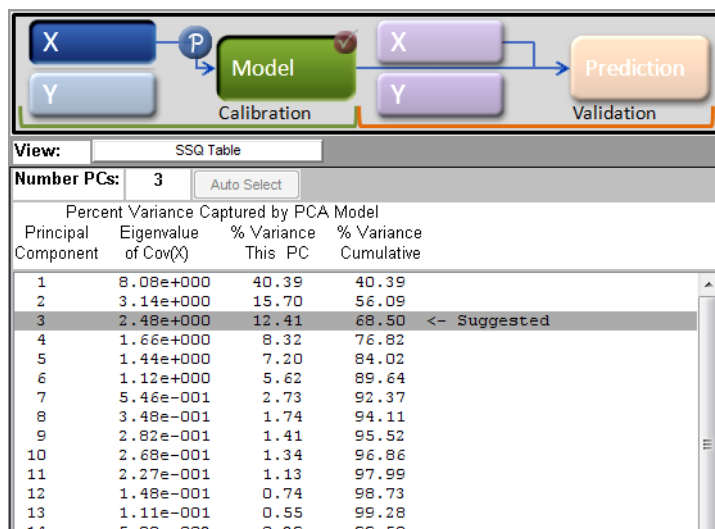
Figure 14-5: Model marked for recalculation






Examining and refining the model

After the model is calculated, the Control pane displays the percent variance captured and other statistical information for the model. For certain analyses, the application provides a suggested number of components or factors to retain for the model based on internal tests. For example, Figure 14-6 shows an initial model calculated for a PCA analysis with the suggested value for the number of components to retain set to three.

Figure 14-6: Initial model calculated for a PCA analysis with number of suggested components = 3




The Analysis window toolbar is updated dynamically with other toolbar buttons based on the selected analysis method. All of these toolbar buttons create plots and other visual aids that assist you in examining and refining the model by excluding certain samples and/or variables to enhance the model performance. Common toolbar buttons include the following:

- The Plot Eigenvalues button . See “Plotting Eigenvalues for a Calibration Model” on page 89.
- The Plot scores and sample statistics button . See “Plotting Scores and Statistical Values for a Calibration Model” on page 93.
- The Plot loads and variable statistics button . See “Plotting Loads and Variable Statistics for a Calibration Model” on page 97.

Note: All other Analysis window toolbar buttons are specific to an analysis method and therefore, are not discussed in this guide.

Chapter 14
Building the Model in the Calibration Phase

Chapter 15: Plotting Eigenvalues for a Calibration Model

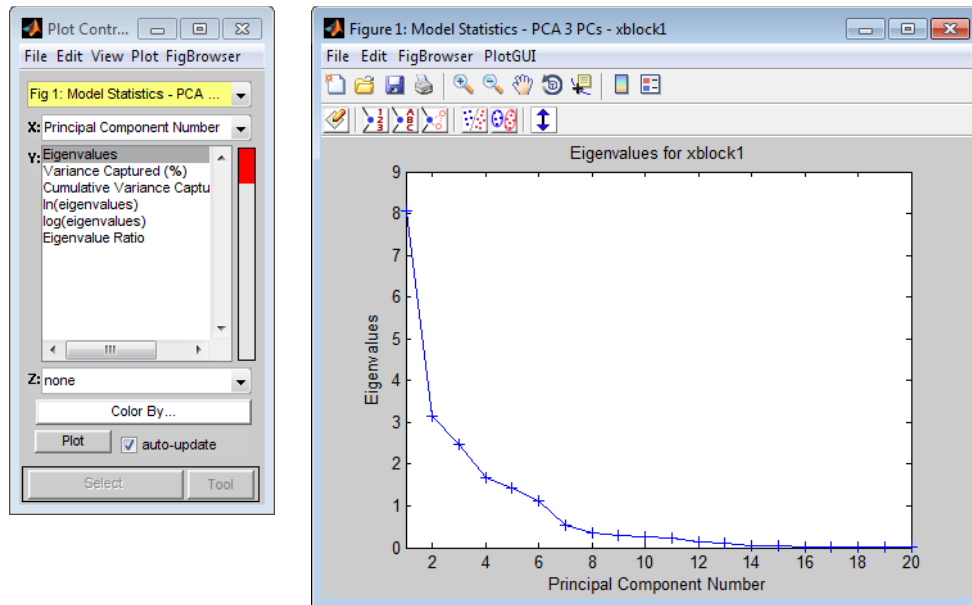
For most analysis methods, the Analysis window toolbar contains a Plot Eigenvalues button . You use the Plot Eigenvalues option to plot a series of univariate metrics as a function of the number of principal components or factors retained in the model. These values assist you in determining the number of principal components or factors to retain the model and often include the following:

- Eigenvalues.
- Variance Captured (%)—The amount of variance captured for each principal component or factor.
- Cumulative Variance Captured (%)—The Cumulative Variance Captured (%) value tracks to the % Variance Cumulative column (the last column) in the Variance Captured data table in the Control pane. This plot shows that with an increasing number of principal components or factors, the cumulative variance asymptotically approaches 100%.
- The natural log of the eigenvalues.
- The log of the eigenvalues.
- The ratio of the eigenvalues.
- The results from any cross-validation that was carried out.

[Figure 15-1 on page 90](#) shows the plot of eigenvalues as a function of the number of principal components retained in the model for a PCA analysis in which twenty variables were measured and three principal components were retained. [Figure 15-2 on page 90](#) shows the plot of the cumulative variance captured as a function of the number of principal components retained in the model for a PCA analysis in which twenty variables were measured and three principal components were retained.

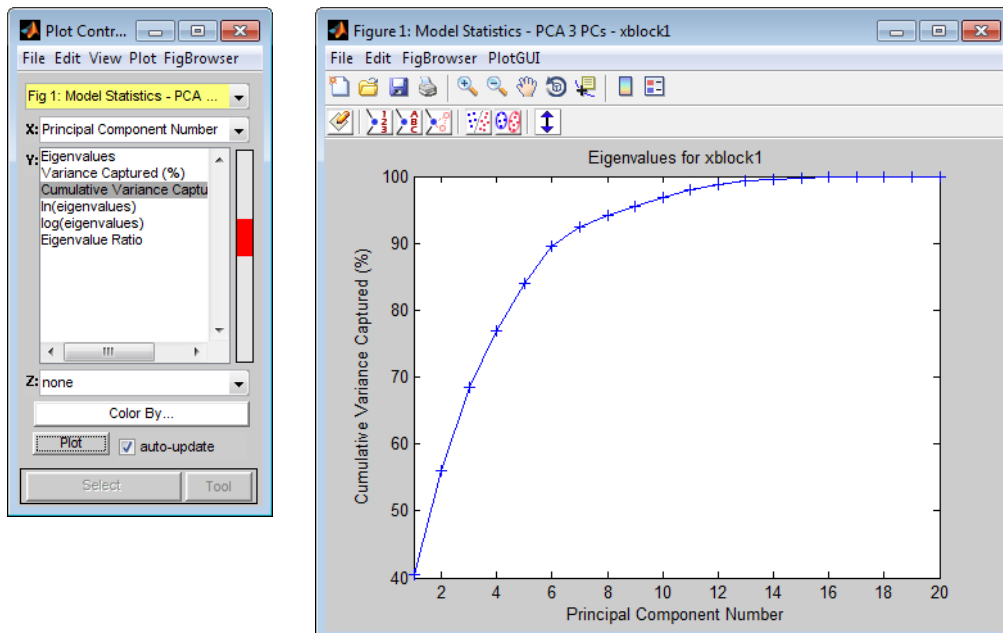
Chapter 15
Plotting Eigenvalues for a Calibration Model

Figure 15-1: Plot of eigenvalues as a function of the number of principal components retained in the model for a PCA analysis



Note: For information about the Plot Controls window and Plot window, see “Plot Controls Window” on page 51.

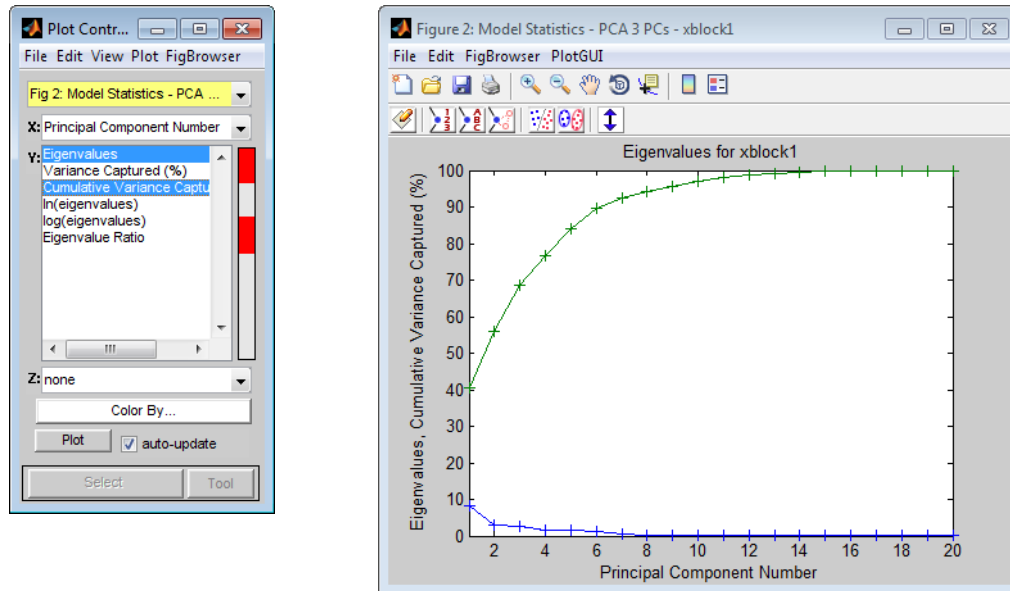
Figure 15-2: Plot of the cumulative variance captured as a function of the number of principal components retained in the model for a PCA analysis



Eigenvalues plot options

You can select multiple Y metrics in the Plot Controls window to overlay these metrics in the Eigenvalues plot. For example, you can CTRL-click Eigenvalues and Cumulative Variance Captured (%) to overlay these values in the Eigenvalues plot.

Figure 15-3: Example of Eigenvalues plot with different plot options



Chapter 15
Plotting Eigenvalues for a Calibration Model

Chapter 16: Plotting Scores and Statistical Values for a Calibration Model


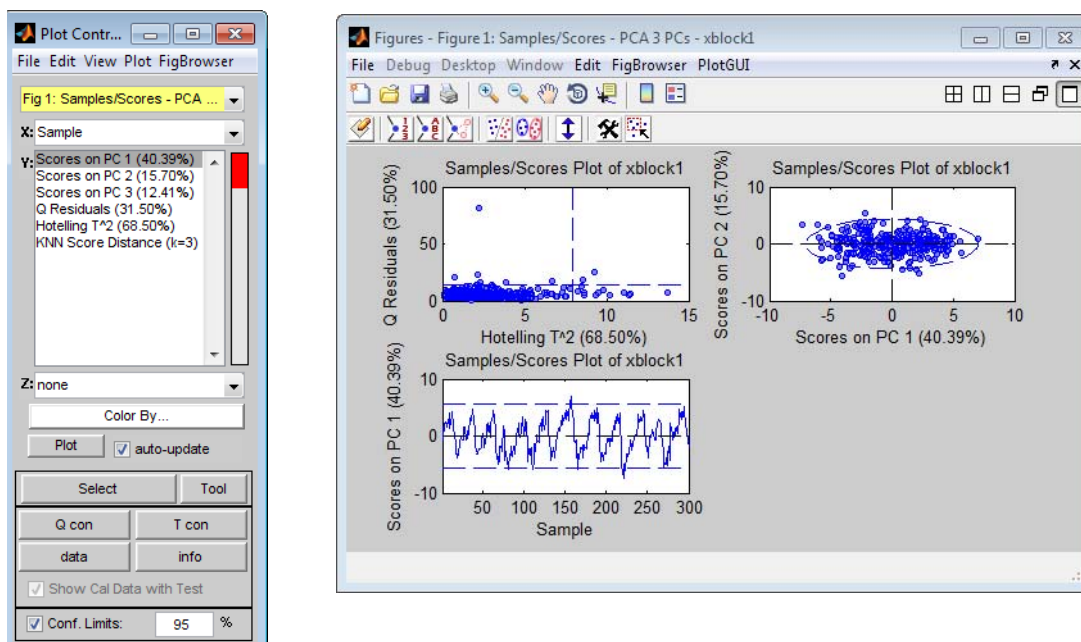
For most analysis methods, the Analysis window toolbar contains a Plot scores and sample statistics button . Scores are the coordinates of the samples in the new principal component or factor coordinate system. A Scores plot shows the relationship among the samples in plots that are displayed in a multiplot Figure window. The Plot Controls window lists the options that can be plotted. The options that are available for plotting depend on the analysis that was carried out. To change the information that is plotted, select a different X value, a different Y value, or both in the Plot Controls window. Figure 16-1 shows some of the possible Scores plots for a PCA analysis in which three principal components were retained.

Figure 16-1: Possible Scores plots for a PCA analysis in which three principal components were retained



Note: For information about the Plot Controls window and Plot window, see “Plot Controls Window” on page 51.

Options are available for changing the plot display and for examining and refining the model by excluding certain samples and/or variables to enhance the model performance. See:

- “Changing the plot display” on page 94.
- “Refining the model by excluding samples” on page 94.

Changing the plot display

Note: The examples listed here are not meant to be an exhaustive list of all of the available Plot Controls options for changing a Scores plot display. Instead, it is simply to provide representative examples of some of the more commonly used options when building a model.

- You can select an individual Y metric to plot, or you can CTRL-click to select multiple Y metrics to plot.
- You can double-click on a plot in the multiplot Figure window to open the plot in its own Figure window, or you can select the plot in the multiplot Figure window, and on the Plot Controls window, click View > Subplots.
- You can view the labels or classes that are associated with the samples in your original data. On the Plot Controls window menu, click View > Labels or View > Classes.
- You can overlay Sample IDs/numbers on a plot. On the Plot Controls menu, click View > Numbers.
- You can declutter a plot for easier viewing if you have Sample IDs/numbers displayed on a plot. On the main menu on the Plot Controls dialog box, click View > Declutter Labels, and then select a Declutter level.

Note: The phrase “Decluttered” appears in the lower left hand corner of a decluttered plot.

- You can change the confidence level for a plot. In the Plot Controls dialog box, ensure that Conf. Limits is selected and in the Confidence Limit field, enter a new value for the confidence limit. (The default value is 95%.)


Note: Confidence Limits for PC versus PC plots are indicated by either a straight dashed line or an ellipse around the plotted data. As you change the value for the Confidence Limit, line moves in position or the ellipse expands or contracts in size accordingly. Data points that are inside the line or the ellipse, therefore, fall within the specified confidence limits. Conversely, data points that are outside the line or the ellipse exceed the specified confidence limits.


Refining the model by excluding samples


Typically, when using Scores plots to refine a model, you identify *samples* that you consider to be unusual for the plotted data and then carry out a series of steps to determine whether to include the samples in the model, or to exclude the samples from the model.

Note: Typically, if you want to refine a model by removing *variables*, you use the information in a Loads plot. See [“Plotting Loads and Variable Statistics for a Calibration Model” on page 97.](#)

Note: The examples listed here are not meant to be an exhaustive list of all of the available Plot Controls options for refining a model using a Scores plot. Instead, it is simply to provide representative examples of some of the more commonly used options when building a model.

1. Initially, you can do one or more of the following to review your samples, and determine which samples, if any, require further investigation:
 - On the Plot Controls window, click Info, and then click on a sample in the plot to open a dialog box that displays information about the sample, such as its Q residual value and its T^2 value.
 - On the Plot Controls window, click T con, and then drag your cursor around a sample in the plot to open the T^2 Contributions dialog box. This dialog box shows the contribution of each variable that was measured for the sample to the T^2 value.
 - On the Plot Controls window, click Q con, and then drag your cursor around a sample in the plot to open the Q Residuals Contributions dialog box. This dialog box shows the contribution of each variable that was measured for the sample to the Q Residual value.
 - On the Plot Controls window, click Data to generate a plot that shows the trend or response of the selected samples for all variables.
 - Double-click on a variable in a Q con and or T con plot to generate a separate trend plot for the variable.
2. For samples that you have determined require further investigation, you can do the following:
 - On the Plot Controls dialog box, click Tools, and then select your tool of choice. (You can also click the Choose Selection Tool icon  and then select your tool of choice.)

Note: “Lasso” is the most flexible tool for selecting samples.
 - On the Plot Controls dialog box, click Make Selection, and then click and drag your cursor around the samples to select them. (You can also click the Make Selection icon  and then click and drag your cursor around the samples to select them.)

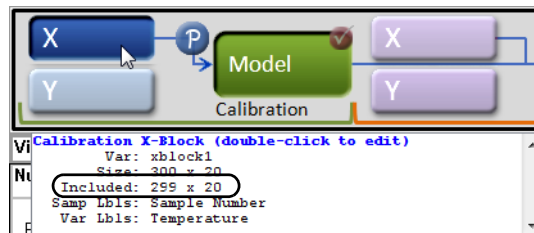
The color of the selected samples is changed, not only in the currently active plot, but also, in any other open plots that contain the samples.
3. With the selected samples now highlighted in the plot, you can do one or more of the following to place the focus on the selected samples and glean further information about the selected samples before deciding to include them or exclude them for the model:
 - To remember what indices correspond to a specific group of selected samples, on the Plot Controls window main menu, click File > Save Selected Indices, and at the prompt, provide a file name for the saved indices. The selected indices are saved as an item  in your Workspace Browser. You can double-click this icon to open a read-only Data Editor window that shows that samples that you selected and the rows from which the samples came. To view the saved indices in a Scores plot, on the Plot Controls window main menu, click File > Load Selected Indices.
 - To display the sample labels or sample numbers/IDs next to the selected samples, on the Plot Controls window main menu, click View > Labels, or click View > Numbers accordingly.

Chapter 16
Plotting Scores and Statistical Values for a Calibration Model

- Plot the Q Residuals for the DataSet versus the T^2 values for the dataset and note where the selected samples fall in the plot. If in this plot you notice additional samples that seem to be unusual (for example, a sample that has a high Q residual value but a low T^2 value), you can select these samples as well. (These additional samples are referred to as the “new” samples in the next two options.)
 - To add the new samples to the currently selected samples, hold down the Shift key, click Select on the Plot Controls window, and then click and drag your cursor around the new samples to select them.
 - To select only the new samples (and deselect the previously selected samples), click Select on the Plot Controls window, and then click and drag your cursor around the new samples to select them. (Do not hold down the Shift key while selecting this new samples.)
4. To exclude all of the selected samples from the model in a single step, on the Plot Controls window main menu, click Edit > Exclude Selection.
- The selected samples are removed from the Scores plots and the Scores plots are updated to reflect this removal.
 - The initial model that was calculated in the Analysis window is removed.

Note: After you exclude samples from a DataSet, point your mouse cursor on a control in which you have loaded data. Tooltip text opens indicating that number of samples that are being included in the model.

Figure 16-2: Tooltip text indicating that number of samples that are being included in the model



- The Plot Controls window indicates that you no longer have Confidence limits because the model has been removed.
- On the Q Residuals versus the T^2 values plot, the boundaries for these values have been removed because you no longer have a model.

At this point, you should iteratively repeat the steps of recalculating the model and then examining the model and refining the model by including or excluding samples until you are satisfied with the model. You can then do one of the following:

- Save the model to the Workspace Browser or to a file and use it at a later date, or export the model to a file or a predictor.

Note: See http://wiki.eigenvector.com/index.php?title=Exporting_Models

- Load validation and test data and apply the model immediately. See “Applying the Model in the Test and Validation Phase” on page 103.

Chapter 17: Plotting Loads and Variable Statistics for a Calibration Model


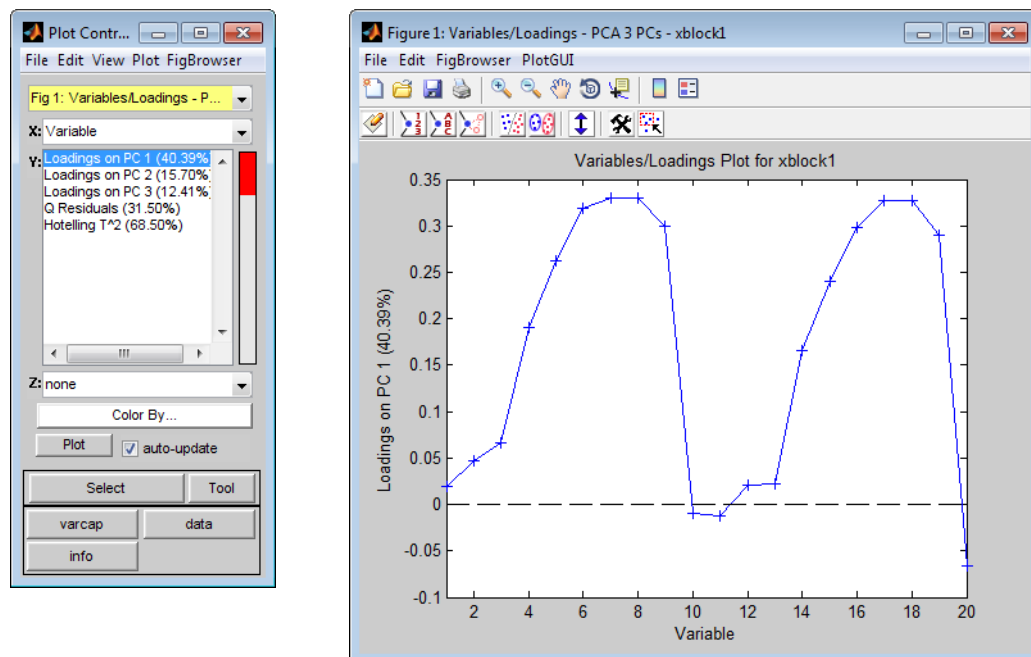
For most analysis methods, the Analysis window toolbar contains a Plot loads and variable statistics button . *Loading* is defined as the contribution of each variable to a principal component or factor. A Loads plot (also known as a Loadings plot) shows, at a minimum, as many different loadings as the number of principal components or factors that were retained in the model. Loads plots help you assess the extent to which the variables contribute to each of the individual principal components or factors. [Figure 17-1](#) shows a Loads plots for a PCA analysis in which 20 variables were measured and three principal components were retained. In the Plot Controls window, three loadings—Loading on PC1, Loading on PC2, and Loading on PC3—are available for plotting.

Figure 17-1: Loads plots for a PCA analysis in which 20 variables were measured and three principal components were retained



Note: For information about the Plot Controls window and Plot window, see [“Plot Controls Window”](#) on page 51.

Options are available for changing the plot display and for examining and refining the model by excluding certain samples and/or variables to enhance the model performance. See:

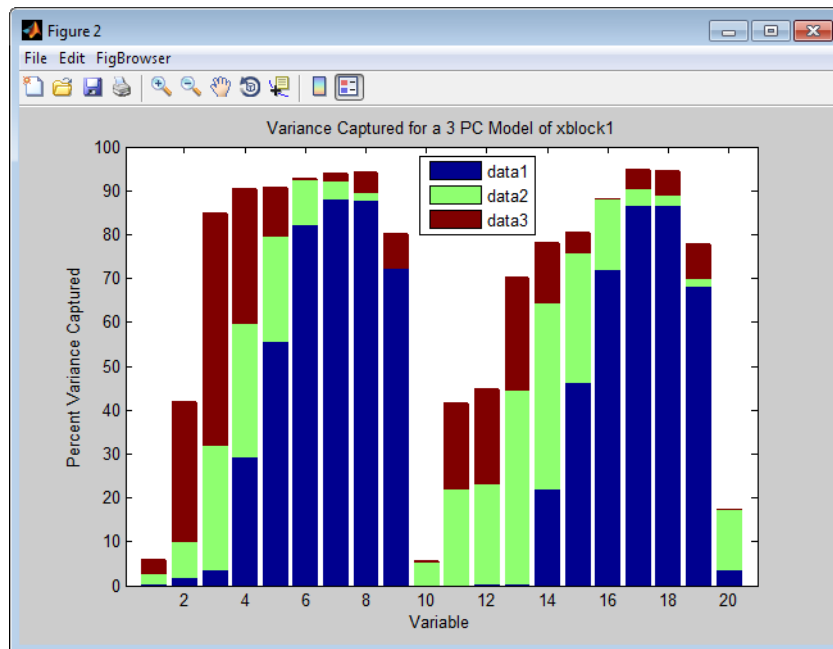
- [“Changing the plot display”](#) on page 98.
- [“Refining the model by removing variables”](#) on page 99.

Changing the plot display

Note: The examples listed here are not meant to be an exhaustive list of all of the available Plot Controls options for changing a Loads plot display. Instead, it is simply to provide representative examples of some of the more commonly used options when building a model.

- Turn on the Data Cursor and then click on a data point in a Loads plot to open an Information dialog box that provides the variable number and the contribution of the variable to the principal component or factor.
- On the Plot Controls window, click varcap to generate a Variance Captured plot. This plot details the percent variance captured for each variable by each principal component or factor. For example, the Variance Captured plot shown in Figure 17-2 shows that the model captures well over 90% of the variance for variables 5 through 8. Likewise, the model captures well over 90% of the variance for variables 17 and 18. The plot also shows that the model captures approximately 40% of the variance in variable 2, with principal component 3 containing the majority of this variance. Likewise, the plot shows that the model captures well over 90% of the variance for variables 6 through 8, with principal component 1 containing the majority of this variance. Finally, the plot also shows that the model captures less than 20% of the variance in variable 20, with principal component 2 containing the majority of the variance and principal component 3 explaining none of the variance in variable 20.

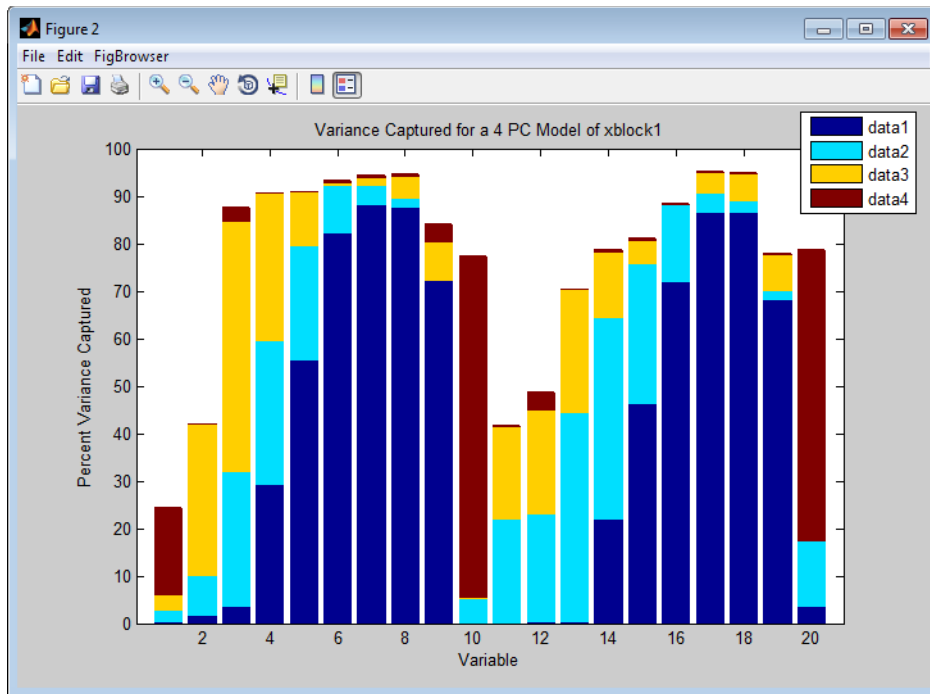
Figure 17-2: Example of a Variance Captured plot



- Choose a different number of components or factors and recalculate the model. The existing varcap plot is automatically updated. Figure 17-3 on page 99 shows the varcap plot for the PCA model with four principal components now retained. This plot shows that as you increase the number of retained principal components or factors, the model

captures more variance. For example, the previous three component model captured less than 20% of the variance in variable 2. A four component model now captures well over 80% of the variance in variable 2, with principal component 4 containing the majority of this variance.

Figure 17-3: Example of a Variance Captured plot



Refining the model by removing variables


Typically, when using Loads plots to refine a model, you select *variables* that you consider to be unusual for the plotted data and then carry out a series of steps to determine whether to include the variables in the model, or to exclude the variables from the model.


Note: Typically, if you want to refine a model by removing *samples*, you use the information in a Scores plot. See [“Plotting Scores and Statistical Values for a Calibration Model” on page 93](#).

Note: The examples listed here are not meant to be an exhaustive list of all of the available Plot Controls options for refining a model using a Loads plot. Instead, it is simply to provide representative examples of some of the more commonly used options when building a model.

1. Initially, you can do the following to review your variables, and determine which variables, if any, require further investigation:
 - On the Plot Controls window, click Data to generate a plot that shows the trend or response of all the variables for all samples.

Chapter 17 Plotting Loads and Variable Statistics for a Calibration Model

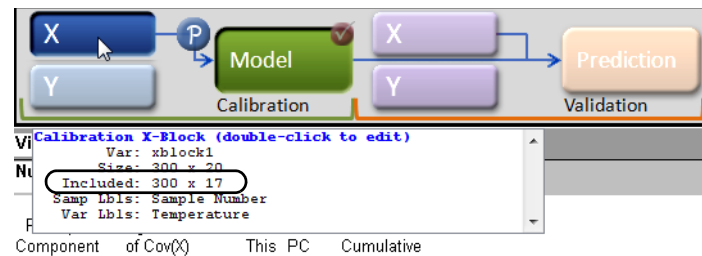
2. For variables that you have determined require further investigation, you can do the following:
 - On the Plot Controls dialog box, click Tools, and then select your tool of choice. (You can also click the Choose Selection Tool icon  and then select your tool of choice.)

Note: “Lasso” is the most flexible tool for selecting samples.
 - On the Plot Controls dialog box, click Make Selection, and then click and drag your cursor around the variables to select them. (You can also click the Make Selection icon  and then click and drag your cursor around the variables to select them.)

The color of the selected variable(s) is not only in the currently active plot, but also, in any other open plots that contain the variable(s).
3. With the selected variables now highlighted in the plot, you can now do one or more of the following to place the focus on the selected variables and glean further information about the selected variables before deciding to include them or exclude them for the model:
 - On the Plot Controls window main menu, click View > Numbers to display the Variable numbers/IDs next to the selected variables.
 - Plot the Q Residuals for the DataSet versus the T² values for the dataset and note where the selected variables fall in the plot. If in this plot you notice additional variable(s) that might not have a large impact on the model (for example, a variable that has a high Q residual value but a low T² value), you can select these variables as well. (These additional variables are referred to as the “new” variables in the next two options.)
 - To add the new variables to the currently selected variables, hold down the Shift key, click Select on the Plot Controls window, and then click and drag your cursor around the new variables to select them.
 - To select only the new variables (and deselect the previously selected samples), click Select on the Plot Controls window, and then click and drag your cursor around the new variables to select them. (Do not hold down the Shift key while selecting this new variables.)
4. To exclude all of the selected variables from the model in a single step, on the Plot Controls window main menu, click Edit > Exclude Selection.
 - The selected variables are removed from the Loads plots and the Loads plots, including the varcap plot, are updated to reflect this removal.
 - The initial model that was calculated in the Analysis window is removed.

Note: As shown in [Figure 17-4 on page 101](#), after you exclude samples from a DataSet, point your mouse cursor on the X calibration control to open tooltip text opens that indicates the number of variables that are being included in the model.

Figure 17-4: Tooltip text that indicates the number of variables that are being included in the model



At this point, you should iteratively repeat the steps of recalculating the model and then examining the model (in particular, by using the varcap plot) and refining the model by including or excluding variables until you are satisfied with the model. You can then do one of the following:

- Save the model to the Workspace Browser or to a file and use it at a later date, or export the model to a file or a predictor.

Note: See http://wiki.eigenvector.com/index.php?title=Exporting_Models

- Load validation and test data and apply the model immediately. See “Applying the Model in the Test and Validation Phase” on page 103.

Chapter 17
Plotting Loads and Variable Statistics for a Calibration Model

Chapter 18: Applying the Model in the Test and Validation Phase

Regardless of the analysis method, the Test and Validation phase consists of the same general series of steps:

1. Loading the validation data and applying the model to the validation data. See [“Loading the validation data and applying the model to the data” on page 103](#).
2. Iteratively examining the model by focusing just on the validation data or on the validation data in the context of the calibration data and refining the model by adjusting the confidence limits and/or reducing the model complexity. See [“Examining and refining the model” on page 105](#).
3. Saving the model.

Note: Decomposition and Clustering analysis methods require only x block data to apply the model in the Test and Validation phase. Regression analysis methods require both x block data and y block data. Classification analysis methods require x block data with classes in either X or Y. For simplicity and brevity, this section describes model application during the Test and Validation phase using a simple PCA model; however, all of the *general* information in this section is applicable for all analysis methods.

Note: For a review of building this example PCA model during the Calibration phase, see [“Building the Model in the Calibration Phase” on page 83](#).

Note: To review a detailed description of the Test and Validation phase, see [Chapter 13, “Analysis Phases,” on page 81](#).

Loading the validation data and applying the model to the data

You have a variety of options for opening an Analysis window and loading data. Because these methods have been discussed in detail in other areas of the documentation, they are not repeated here. Instead, a brief summary is provided with a cross-reference to the detailed information. Simply choose the method that best fits your working needs.

- To open an Analysis window:
 - In the Workspace Browser, click the shortcut icon for the specific analysis that you are carrying out.
 - In the Workspace Browser, click Other Analysis to open an Analysis window, and on the Analysis menu, select the specific analysis method that you are carrying out.
 - In the Workspace Browser, drag a data icon to a shortcut icon to open the Analysis window and load the data in a single step.

Note: For information about working with icons in the Workspace Browser, see [“Icons in the Workspace Browser” on page 37](#).

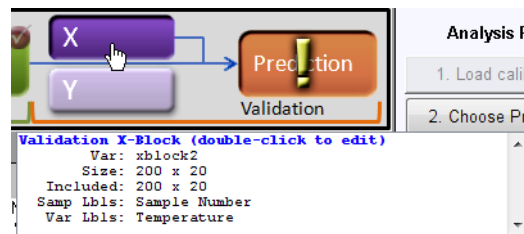
Chapter 18
Applying the Model in the Test and Validation Phase

- To load data into an open Analysis window:
 - Click File on the Analysis window main menu to open a menu with options for loading and importing validation data.
 - Click the appropriate validation control to open the Import dialog box and select a file type to import.
 - Right-click the appropriate validation control to open a context menu with options for loading and importing data.
 - Right-click on an entry for a cached item Model Cache pane to open a context menu that contains options for loading the selected cached item into the Analysis window.

Note: For information about the data manipulation options on the context menu, see “Icons in the Workspace Browser” on page 37 or “Importing Data into the Workspace Browser” on page 31. For information about loading items from the Model Cache pane, see “Analysis window Model Cache pane” on page 73.

Also, remember that after you load data into a validation control, you can place your mouse pointer on the control to view not only information about the loaded data, but also, different instructions about working with the control. In Figure 18-1, data has been loaded into the X validation control for a PCA analysis.

Figure 18-1: Viewing information about loaded data



After you have opened the Analysis window and loaded the validation data, you then apply the model to the validation data. To apply the model to the validation data, you can do one of the following:


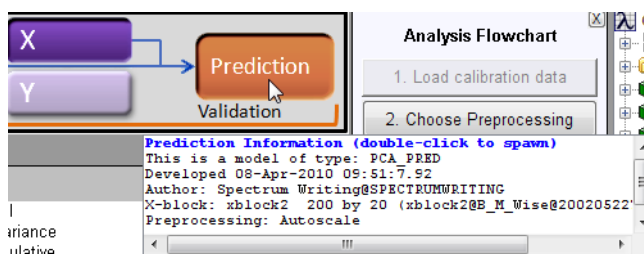
- On the Analysis window toolbar, click the Calculate/Apply model icon .
- Click the Model control.

Figure 18-2: Clicking the model control to apply the model to validation data



After the model is applied to the validation data, you can place your mouse pointer on the Model control to view information about the model.

Figure 18-3: Viewing information about the model



Examining and refining the model

After the model is applied, the most relevant plots to create are Scores plots. (Remember, nothing has changed about the Eigenvalues, nothing has changed about the loads and variable statistics - all you have done is apply the model to validation data.)

Note: The examples listed here are not meant to be an exhaustive list of all of the available Plot Controls options for refining a model using a Scores plot. Instead, it is simply to provide representative examples of some of the more commonly used options when applying a model.


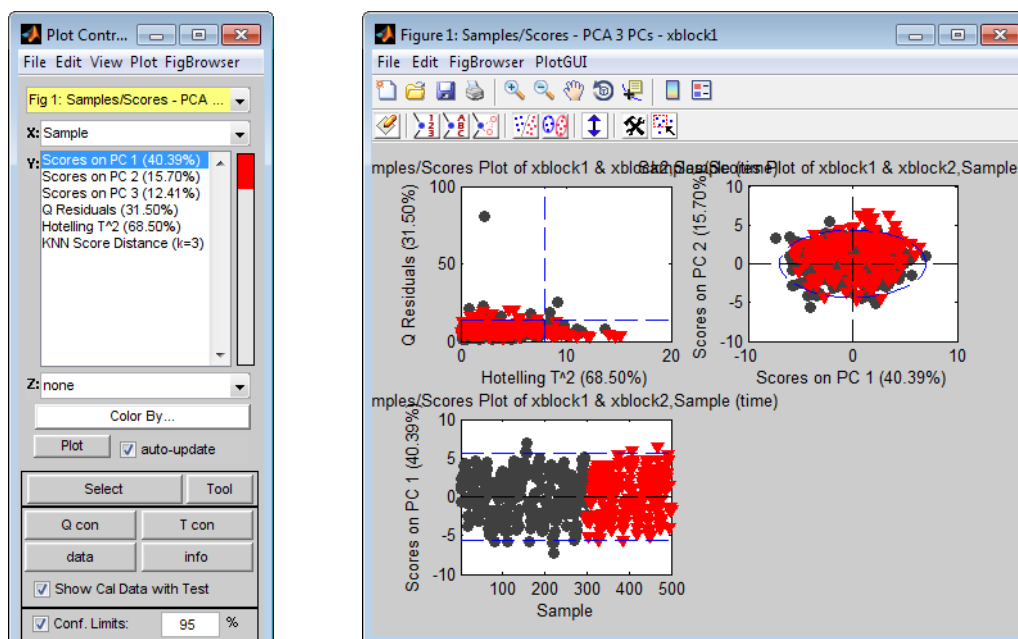
1. On the Analysis window toolbar, click the Plot scores and sample statistics button .

Figure 18-4 shows some of the possible Scores plots for when applying the PCA model that you built during the Calibration phase to validation data. Note that initially, the plots show both the calibration data (▼) and the validation data (●).

Figure 18-4: Scores plot showing both calibration data and validation data



Chapter 18

Applying the Model in the Test and Validation Phase

2. Double-click on a plot of interest in the multiplot Figure window to open the plot in its own Figure window, or you can select the plot in the multiplot Figure window, and on the Plot Controls window, click View > Subplots.
3. With the plot of interest now open in its own Figure window, you can now do one or more of the following to refine the model:
 - To view only the validation data in the plot, clear the Show Cal Data with Test option in the Plot Controls window. To view the validation data in the context of the calibration data, click the Show Cal Data with Test option again.

Note: You should review a variety of plots with just the validation data, and then with both the validation data and calibration data to ensure that your validation data has the same distribution as your calibration data. For example, plot the Q Residuals versus the T^2 values for the validation data alone, then click the Show Cal Data with Test option to add the calibration data to this plot to confirm that your validation samples cover the same “space” (low Q/low T^2) that your calibration samples cover.

- Increase the confidence limits as needed from the default value of 95% to ensure that all of your validation samples are contained within the limits. (Remember, you know the characteristics of your validation data and you know, therefore, that all of the validation data is “good” and must be included in the model.)
 - Reduce the complexity of the model by choosing a different number of components or factors to retain in the model and then recalculate the model. After you recalculate the model, a new prediction is automatically calculated. (See [“Changing the number of components” on page 85](#) for the procedure that describes how to recalculate a model.)
4. After you are satisfied with the model, save the model.

Chapter 19: Cross-Validation Tool

You use the Cross-Validation tool to:

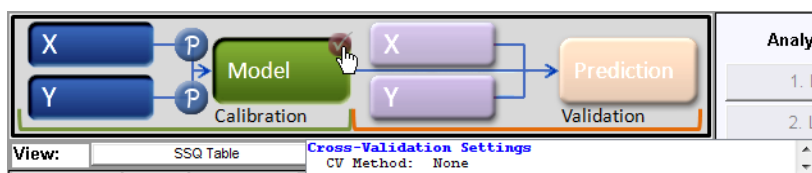
- Assess the optimal complexity of a model (for example, the number of principal components in a PCA or PCR model, or the number of latent variables in a PLS model).
- Estimate the performance of a model when you apply the model to unknown data.

For a given set of data, cross-validation involves a series of steps called *subvalidation steps* in which you remove a subset of objects from a set of data (the *test set*), build a model using the remaining objects in the set of data (the *model building set*), and then apply the resulting model to the removed objects. You note how the errors accumulate as you leave out samples to determine the number of principal components/latent variables/factors to retain in the model. Cross-validation typically involves more than one subvalidation step, each of which in turn involves the selection of different subsets of samples for model building and model testing. In Solo, five different cross-validation methods are available, and these methods vary with respect to *how* the different sample subsets are selected for these subvalidation steps.

1. To open the Cross-Validation tool, do one of the following:
 - On the Analysis window, click Tools > Cross-Validation.
 - Click the Cross-Validation icon in the Analysis window.

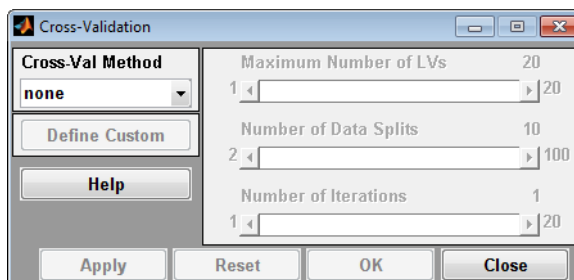
Note: You must load data into the Analysis window before the Cross-Validation icon is available.

Figure 19-1: Cross-validation icon in the Analysis window



- In the Analysis window Flowchart pane, click Choose Cross-Validation.
2. In the Cross-Validation dialog box, select the method of cross-validation that you want to use.

Figure 19-2: Cross-Validation dialog box


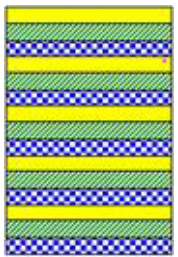




3. Use the slider bars to change the default values for the available parameters.

Note: Not all parameters are relevant for all cross-validation methods. The initial values that are specified for the available parameters are default values that are based on the dimensionality of the data. You can click Reset at any time to reset the parameters to their default settings. For the following descriptions in [Figure 19-3 on page 109](#):

- n is the total number of objects in the set of data.
- s is the number of data splits specified for the cross-validation procedure, which must be less than $n/2$.
- r is the number of iterations.

Figure 19-3: Cross-validation methods compared

	Leave One Out	Venetian Blinds	Contiguous Block	Random Subsets	Custom
Cross-validation method					
Description	The default value. All samples in the set of data are used to build the model.	Each test set is determined by selecting every s th object in the set of data, starting at objects numbered 1 through s.	An alternative to Venetian Blinds. Each test set is determined by selecting contiguous blocks of n/s objects in the set of data, starting at object number 1.	"s" different test sets are determined through random selection of n/s objects in the set of data, such that no single object is in more than one test set. This procedure is repeated "r" times, where "r" is the number of iterations.	You manually define each of the test sets. You can assign specific objects in your set of data in one of three ways: <ul style="list-style-type: none"> • To be in every test set. • To never be in a test set. • To not be used in the cross-validation procedure at all.
Available Parameters	Maximum Number of LVs	<ul style="list-style-type: none"> • Maximum Number of LVs • Number of Data Splits 	<ul style="list-style-type: none"> • Maximum Number of LVs • Number of Data Splits 	<ul style="list-style-type: none"> • Maximum Number of LVs • Number of Data Splits • Number of Iterations 	<ul style="list-style-type: none"> • Number of data splits • Object membership for each split • Total number of objects
# of Subvalidation Steps	n	s	s	(s*r)	s
# of Test Samples per Subvalidation	1	n/s	n/s	n/s	Varies. User-defined.

Chapter 19
Cross-Validation Tool

4. Do one of the following:
 - Click Apply button to apply these settings and keep the Cross Validation dialog box open.
 - Click OK to apply these settings and close the Cross Validation dialog box.

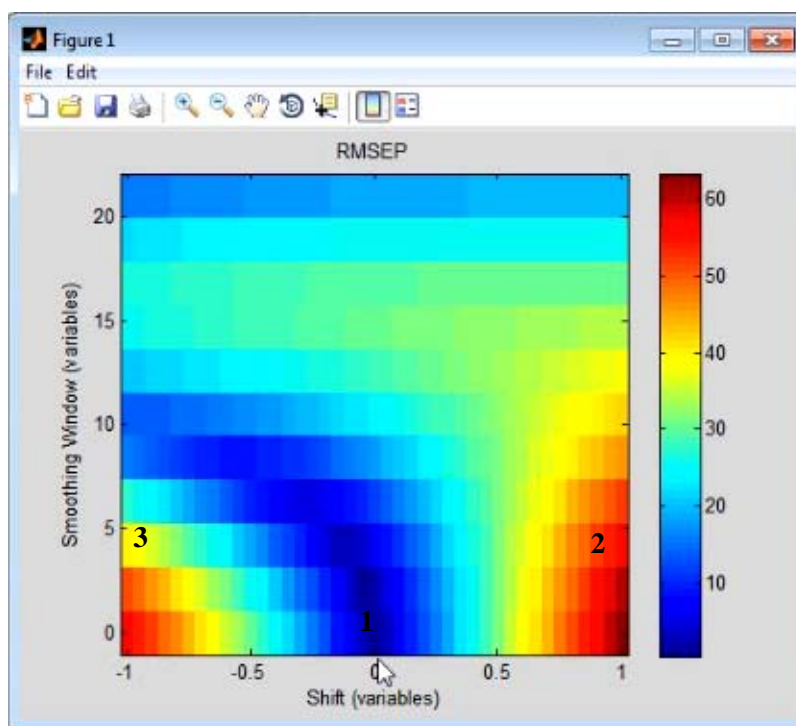
Chapter 20: Model Robustness Tool

You use the Model Robustness tool to measure the sensitivity of a regression model to artifacts in new spectroscopic measurements. To open the Model Robustness tool, on the Analysis window, click Tools > Model Robustness, and then click Shifts or Interferences.

Shifts

The Shifts option measures the sensitivity of a regression model to shifts in x-axis data that are caused by instrument instability—that is, if you have an instrument that is not particularly stable or reproducible over time, what is the impact on predictions using the given model? The Shifts plot is a three-dimensional plot that details the RMSEP (Root Mean Squared Error of Prediction) for the model as a function of shift, where shift is described in terms of the number of variables and the Smoothing window.

Figure 20-1: Example of a Shifts plot



Consider Figure 20-1 above, which shows the model robustness for a regression model with an RMSEC (Root Mean Squared Error of Calibration) of approximately 0.5. As shown in this figure:

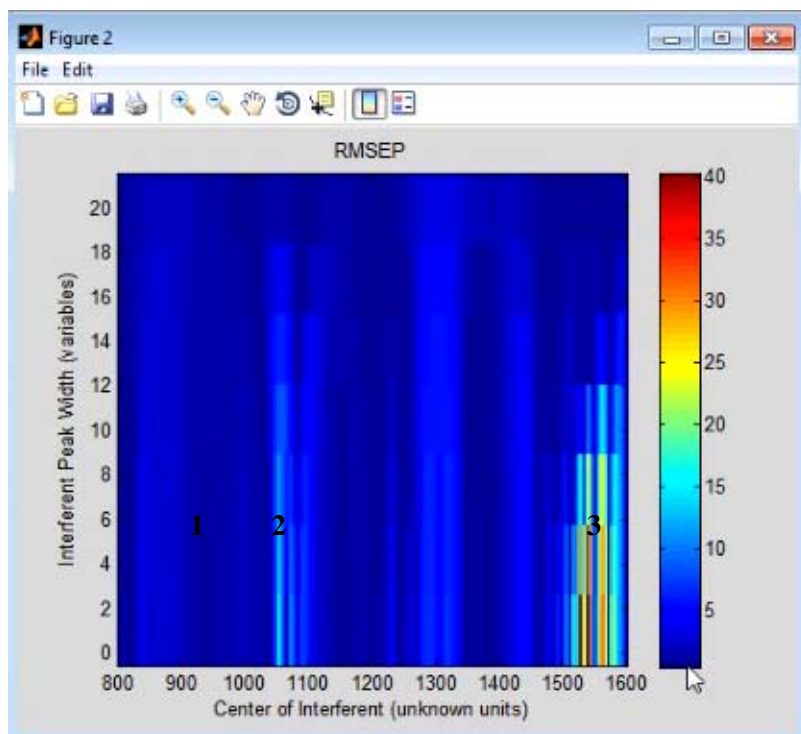
- 1 - Without shift and without smoothing of the variables, the RMSEP indicates that you have test data that is identical to your calibration data and you have performance that is on par for the RMSEC of the model
- 2 - Shifting a spectrum over by simply one variable increases the RMSEP for the model by almost twelve orders of magnitude, from 0.5 to almost 60.

- 3 - With a combination of shifting and smoothing, the impact on the model is lessened somewhat.

Interferences

The Interferences option measures the sensitivity of a regression model to the location and width of a new peak in test data—that is, if you have a chemical entity that is present in the test data but that was not reflected in the calibration data, what is the impact on predictions using the given model? The Interferences plot is a three-dimensional plot that details the RMSEP (Root Mean Squared Error of Prediction) for the model as a function of a new peak, where the peak is described in terms of its width and location.

Figure 20-2: Example of an Interferences plot



Consider [Figure 20-2](#) above, which shows the model robustness for a regression model with an RMSEC (Root Mean Squared Error of Calibration) of approximately 0.5. As shown in this figure, the RMSEP for the model can be impacted in one of three ways:

- 1 - An interferant area where there is virtually no impact on the RMSEP for the model, no matter how wide the interfering peak is.
- 2 - An interferant area where there is a slight impact on the RMSEP for the model, but the impact is lessened as the width of the peak increases.
- 3 - An interference area where there is a significant impact on the RMSEP for the model, but the impact is lessened as the width of the peak increases.

Chapter 21: Correlation Map Tool

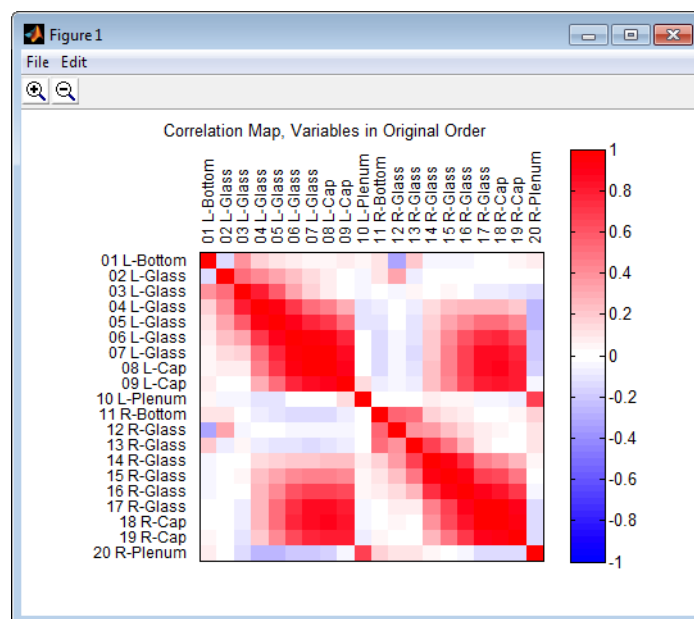
The Correlation Map is a tool that shows the degree of correlation among the variables after you have loaded x block data. You can show the degree of correlation before you preprocess the x block data or after you preprocess the x block data. In addition, you can show the correlation among the variables in one of three ways:

Option	Description
No variable reordering	Variables are plotted in their original order.
Correlation reordering	Adjacent variables will have the highest degree of positive correlation when variables are regrouped by similarity.
Absolute value reordering	Adjacent variables will have the highest degree of positive correlation and/or negative correlation when variables are regrouped by similarity.

To create a correlation map, on the Analysis window main menu, click Tools > Correlation Map, and then select the way in which you want to order the variables.

Figure 21-1 shows a correlation map with the variables in their original order after *default preprocessing methods* were carried out on a 200 x 30 DataSet for a simple PCA model. The color and intensity of the off-diagonal elements indicates the correlation among the variables. For example, in Figure 21-1, variable 18 correlates well with variables 7, 8, and 9 but it does not correlate well with variables 4 and 5.

Figure 21-1: Example of a Correlation map



Chapter 21
Correlation Map Tool



To zoom in on an area for viewing, click the Zoom in icon  and then drag your cursor around the region of interest. A box is formed around the area that being reduced for viewing. To zoom out on area for viewing, click the Zoom Out icon  and then drag your cursor around the region of interest. A box is formed around the area that being enlarged for viewing. [Figure 21-2](#) shows a region that was reduced for viewing. In this figure, the correlation of variable 18 with variables 7 and 8 is much easier to ascertain.

Figure 21-2: Example of zooming in on a Correlation map

