Altair HyperStudy 2021

User Guide

Updated: 01/20/2021

Contents

Intellectual Property Rights Notice

Copyrights, Trademarks, Trade Secrets, Patents & Third Party Software Licenses

Copyrights in the below are held by Altair Engineering, Inc., except where otherwise explicitly stated. This Intellectual Property Rights Notice is exemplary, not exhaustive.

> **Note:** Pre-release versions of Altair software are provided 'as is', without warranty of any kind. Usage of pre-release versions is strictly limited to non-production purposes.

Altair HyperWorks™ - The Platform for Innovation™

Altair AcuConsole™ ©2006-2020

Altair AcuSolve™ ©1997-2020

Altair Activate® ©1989-2020 (formerly solidThinking Activate)

Altair Compose® ©2007-2020 (formerly solidThinking Compose)

Altair ConnectMe™ ©2014-2020

Altair EDEM ©2005-2020 DEM Solutions Ltd, ©2019-2020 Altair Engineering Inc.

Altair ElectroFlo™ ©1992-2020

Altair Embed® ©1989-2020 (formerly solidThinking Embed)

- **Altair Embed SE** ©1989-2020 (formerly solidThinking Embed SE)
- **Altair Embed/Digital Power Designer** ©2012-2020
- **Altair Embed Viewer** ©1996-2020

Altair ESAComp™ ©1992-2020

Altair Feko™ ©1999-2014 Altair Development S.A. (Pty) Ltd., ©2014-2020 Altair Engineering Inc.

Altair Flux™ ©1983-2020

Altair FluxMotor™ ©2017-2020

Altair HyperCrash™ ©2001-2020

Altair HyperGraph™ ©1995-2020

Altair HyperLife™ ©1990-2020

Altair HyperMesh™ ©1990-2020

Altair HyperStudy™ ©1999-2020

Altair HyperView™ ©1999-2020

Altair HyperXtrude™ ©1999-2020

Altair Inspire™ ©2009-2020 including Altair Inspire Motion, Altair Inspire Structures, and Altair Inspire Print3D

Altair Inspire Cast ©2011-2020 (formerly Click2Cast)

Altair Inspire ElectroFlo ©1992-2020

Altair Inspire Extrude Metal ©1996-2020 (formerly Click2Extrude-Metal)

Altair Inspire Extrude Polymer ©1996-2020 (formerly Click2Extrude-Polymer)

Altair Inspire Form ©1998-2020 (formerly Click2Form)

Altair Inspire Friction Stir Welding ©1996-2020

Altair Inspire Mold ©2009-2020

Altair Inspire Play ©2009-2020

Altair Inspire PolyFoam ©2009-2020

Altair Inspire Render ©1993-2016 Solid Iris Technologies Software Development One PLLC, ©2016-2020 Altair Engineering Inc (formerly Thea Studio)

Altair Inspire Resin Transfer Molding ©1990-2020

Altair Inspire Studio ©1993-2020 (formerly 'Evolve')

Altair Manufacturing Solver™ ©2011-2020

Altair Material Data Center ©2019-2020

Altair MotionSolve™ ©2002-2020

Altair MotionView™ ©1993-2020

Altair Multiscale Designer™ ©2011-2020

Altair nanoFluidX™ ©2013-2018 Fluidyna GmbH, ©2018-2020 Altair Engineering Inc.

Altair newFASANT ©2010-2020

Altair OptiStruct™ ©1996-2020

Altair PollEx ©2003-2020

Altair Radioss™ ©1986-2020

Altair Seam™ ©1985-2019 Cambridge Collaborative, Inc., ©2019-2020 Altair Engineering Inc.

Altair SimLab™ ©2004-2020

Altair SimSolid™ ©2015-2020

Altair ultraFluidX™ ©2010-2018 Fluidyna GmbH, ©2018-2020 Altair Engineering Inc.

Altair Virtual Wind Tunnel™ ©2012-2020

Altair WinProp™ ©2000-2020

Altair WRAP ©1998-2020 WRAP International AB, ©2020 Altair Engineering AB

Altair Packaged Solution Offerings (PSOs)

Altair Automated Reporting Director™ ©2008-2020

Altair GeoMechanics Director™ ©2011-2020

Altair Impact Simulation Director™ ©2010-2020

Altair Model Mesher Director™ ©2010-2020

Altair NVH Director™ ©2010-2020

Altair Squeak and Rattle Director™ ©2012-2020

Altair Virtual Gauge Director™ ©2012-2020

Altair Weight Analytics™ ©2013-2020

Altair Weld Certification Director™ ©2014-2020

Altair Multi-Disciplinary Optimization Director™ ©2012-2020

Altair PBSWorks™ - Accelerating Innovation in the Cloud™

Altair PBS Professional® ©1994-2020

Altair Control™ ©2008-2020; (formerly **PBS Control**)

Altair Access™ ©2008-2020; (formerly **PBS Access**)

Altair Accelerator™ ©1995-2020; (formerly **NetworkComputer**)

Altair Accelerator™ Plus©1995-2020; (formerly **WorkloadXelerator**)

Altair FlowTracer™ ©1995-2020; (formerly **FlowTracer**)

Altair Allocator™ ©1995-2020; (formerly **LicenseAllocator**)

Altair Monitor™ ©1995-2020; (formerly **LicenseMonitor**)

Altair Hero™ ©1995-2020; (formerly **HERO**)

Altair Software Asset Optimization (SAO) ©2007-2020

Note:

Compute Manager™ ©2012-2017 is now part of **Altair Access**

Display Manager™ ©2013-2017 is now part of **Altair Access**

PBS Application Services™ ©2008-2017 is now part of **Altair Access**

PBS Analytics™ ©2008-2017 is now part of **Altair Control**

PBS Desktop™ ©2008-2012 is now part of **Altair Access**, specifically **Altair Access desktop**, which also has **Altair Access web** and **Altair Access mobile**

e-Compute™ ©2000-2010 was replaced by **"Compute Manager"** which is now **Altair Access**

Altair KnowledgeWorks™

Altair Knowledge Studio® ©1994-2020 Angoss Software Corporation, ©2020 Altair Engineering Inc.

Altair Knowledge Studio for Apache Spark ©1994-2020 Angoss Software Corporation, ©2020 Altair Engineering Inc.

Altair Knowledge Seeker™ ©1994-2020 Angoss Software Corporation, ©2020 Altair Engineering Inc.

Altair Knowledge Hub™ ©2017-2020 Datawatch Corporation, ©2020 Altair Engineering Inc.

Third party software licenses

AcuConsole contains material licensed from Intelligent Light (www.ilight.com) and used by permission.

Software Security Measures:

Altair Engineering Inc. and its subsidiaries and affiliates reserve the right to embed software security mechanisms in the Software for the purpose of detecting the installation and/or use of illegal copies of the Software. The Software may collect and transmit non-proprietary data about those illegal copies. Data collected will not include any customer data created by or used in connection with the Software and will not be provided to any third party, except as may be required by law or legal process or to enforce our rights with respect to the use of any illegal copies of the Software. By using the Software, each user consents to such detection and collection of data, as well as its transmission and use if an illegal copy of the Software is detected. No steps may be taken to avoid or detect the purpose of any such security mechanisms.

Technical Support

Altair provides comprehensive software support via web FAQs, tutorials, training classes, telephone, and e-mail.

Altair One Customer Portal

Altair One (https://altairone.com/) is Altair's customer portal giving you access to product downloads, a Knowledge Base, and customer support. We strongly recommend that all users create an Altair One account and use it as their primary means of requesting technical support.

Once your customer portal account is set up, you can directly get to your support page via this link: www.altair.com/customer-support/

Altair Training Classes

Altair's in-person, online, and self-paced trainings provide hands-on introduction to our products, focusing on overall functionality. Trainings are conducted at our corporate and regional offices or at your facility.

For more information please visit: https://learn.altair.com/

If you are interested in training at your facility, please contact your account manager for more details. If you do not know who your account manager is, please contact your local support office and they will connect you with your account manager.

Telephone and E-mail

If you are unable to contact Altair support via the customer portal, you may reach out to technical support via phone or e-mail. Use the following table as a reference to locate the support office for your region.

When contacting Altair support, please specify the product and version number you are using along with a detailed description of the problem. It is beneficial for the support engineer to know what type of workstation, operating system, RAM, and graphics board you have, so please include that in your communication.

Location	Telephone	E-mail
Australia	+61 649 413 7981	anzsupport@altair.com
Brazil	+55 113 884 0414	br_support@altair.com
Canada	+1 416 447 6463	support@altairengineering.ca
China	+86 400 619 6186	support@altair.com.cn
France	+33 141 33 0992	francesupport@altair.com
Germany	+49 703 162 0822	hwsupport@altair.de
Greece	+30 231 047 3311	eesupport@altair.com

Location	Telephone	E-mail
India	+91 806 629 4500 +1 800 425 0234 (toll free)	support@india.altair.com
Israel		israelsupport@altair.com
Italy	+39 800 905 595	support@altairengineering.it
Japan	+81 36 225 5830	support@altairjp.co.jp
Malaysia	+60 32 742 7890	aseansupport@altair.com
Mexico	+52 555 658 6808	mx-support@altair.com
New Zealand	+64 9 413 7981	anzsupport@altair.com
South Africa	+27 21 831 1500	support@altair.co.za
South Korea	+82 704 050 9200	support@altair.co.kr
Spain	+34 910 810 080	support-spain@altair.com
Sweden	+46 46 460 2828	support@altair.se
United Kingdom	+44 192 646 8600	support@uk.altair.com
United States	+1 248 614 2425	hwsupport@altair.com

If your company is being serviced by an Altair partner, you can find that information on our web site at https://www.altair.com/PartnerSearch/.

See www.altair.com for complete information on Altair, our team, and our products.

What's New

1

View new features for HyperStudy 2021.

This chapter covers the following:

1.1 Altair HyperStudy 2021 Release Notes

New Features

Sampling Fit

New approach that combines Space-filling DOE and Fit. It samples the design space and builds a fit model at user-defined intervals until desired cross-validation R^2 value is achieved.

Side-by-Side Post-Processing

In Post-Processing, a new capability has been added to simultaneously visualize results of different forms in multiple windows.

Altair Inspire Studio Model

Inspire Studio is a hybrid modeling and rendering environment for 3D conceptual designs. The new connection enables users to perform geometric changes directly within HyperStudy.

Model Conditions

Defining conditions for models is now supported. Users can control the sequence of executions and whether a model is to be executed or not.

Enhancements

Periodic Report Generator

Periodic data extraction during evaluation is now supported.

Multi-Execution for Each Model

Multi-Execution can now be applied to multiple independent models running in parallel.

Registering Python Function Registration via GUI

External python functions can now be registered in the user interface as an alternative to preference file (`*.mvw`) route.

ANOVA Calculation

Previously, ANOVA was a mandatory part of building Least Square Regression models. It is now disintegrated from the model building process and is optional in Post-Processing.

Improved Computational Efficiency of HyperStudy Fit Model

Pyfit file is used in memory for evaluation which provides computational efficiency.

Linked Variables in Trade-Off Panel

Linked variables can now be displayed in Trade-Off panel upon request.

Enhancements in AREA Tool

Two enhancements have been introduced: normalized area difference and logarithmic scaling option for highly dynamic curves.

New Layout for Solver Input Argument Dialog

Arguments are now classified in two categories: Standard and Solver Arguments. Arguments are appended and visualized in stacked form.

Automatic Preview in Expression Builder

Expression field is divided into two windows to accommodate the expression and its preview. Preview update is automatic by default. However, users can opt for manual updates.

Maximize Levels for Discrete Variables in DOE

One-click option has been added to maximize levels for discrete variables in applicable screening DOE methods. Maximum and minimum levels along with values are displayed in table form.

Monitoring HyperStudy in Batch Mode During Evaluation

Monitoring HyperStudy in batch mode is now supported. There are three available options to manipulate the process: exit HyperStudy, stop evaluation, and change multi-execution count.

Resolved Issues

- Evaluation times were incorrectly reported.
- Removing multiple models at the same time resulted in error.
- In Verification Approach, Delta Plots were incorrectly scaled in some cases.
- Adding a goal in approaches other than optimization invalidated the definition.
- Discrete design variables were not supported in Workbench model.

Announcements

- Future versions of HyperStudy will remove support for upgrading studies constructed with version 14.0 and earlier.
- Verification approach can no longer be added as a standalone approach anymore. However, it remains available within Fit and Optimization approaches.
- Korean Language is no longer supported.
- Multi-execution for HyperStudy Fit model is no longer supported.
- In Solver Input Arguments, keyword `filespec` has been replaced with `filepath`. Keywords `filebasename` and `fileroot` are decommissioned.

Get Started

2

Learn the basics and discover the workspace.

This chapter covers the following:

2.1 HyperStudy Overview

A multi-disciplinary design exploration, study and optimization software.

HyperStudy enables you to explore, understand and improve your system's designs using methods such as design-of-experiments and optimization. HyperStudy generates intelligent variations of the parameters of any system model and reveals relationships between these parameters and the system responses. By using HyperStudy, you can make better decisions and optimize the performance, reliability and robustness of your systems.

Figure 1:

HyperStudy provides engineers and designers a user-friendly environment to:

Improve Design Performance and Quality
: HyperStudy includes state-of-the-art, innovative optimization, design of experiments and stochastic methods for rapid assessment and improvement of design performance and quality.

Reduce Development Time and Costs
: HyperStudy helps engineers reduce trial-and-error iterations and hence helps to reduce both the design development and testing time.

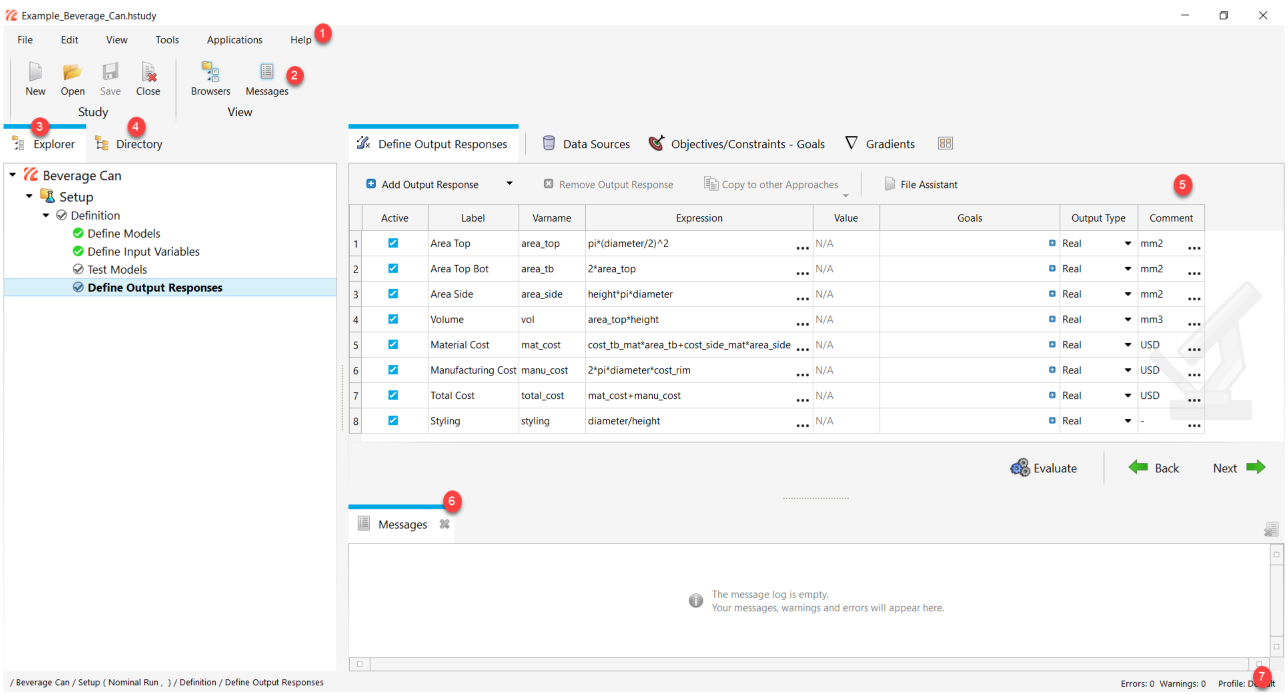
Increase Productivity through Easy-to-use Environment
: HyperStudy's step-by-step process guides the user in setting up and carrying out design studies. Its open architecture allows easy integration with 3rd party solvers.

Powerful Dataset Analyses
: Comprehensive set of post processing and data mining methods simplify and aid an engineer’s job of analyzing and understanding large simulation datasets.

Improve Simulation Correlation
: HyperStudy's optimization capabilities can be applied to improve correlation of analysis models with test results or with other models.

2.2 User Interface

Explore the HyperStudy user interface.



1. Menu Bar
2. Ribbons
3. Explorer
4. Directory
5. Work Area
6. Message Log Window
7. Status Bar

2.2.1 Menu Bar

The menu bar contains pull-down menus that provide access to standard functions such as file management operations, system preferences, and help.

See Also

Set HyperStudy Preferences

2.2.2 Ribbons

The ribbon allows you to quickly access tools and standard functions, and is located along the top of HyperStudy. Click on an icon to open the related tool. Hovering over a group of icons may reveal additional tools.

> **Tip:** Change the icons and controls displayed in the ribbon from the context menu that opens when you right-click on the ribbon.

2.2.3 Explorer

The Explorer displays a hierarchical view of your study setup and approaches, and can be used to guide you through the study setup process.

When you create a study or open an existing study, it will appear in the Explorer. Clicking the different items in the Explorer guides you to the various parts of your study.

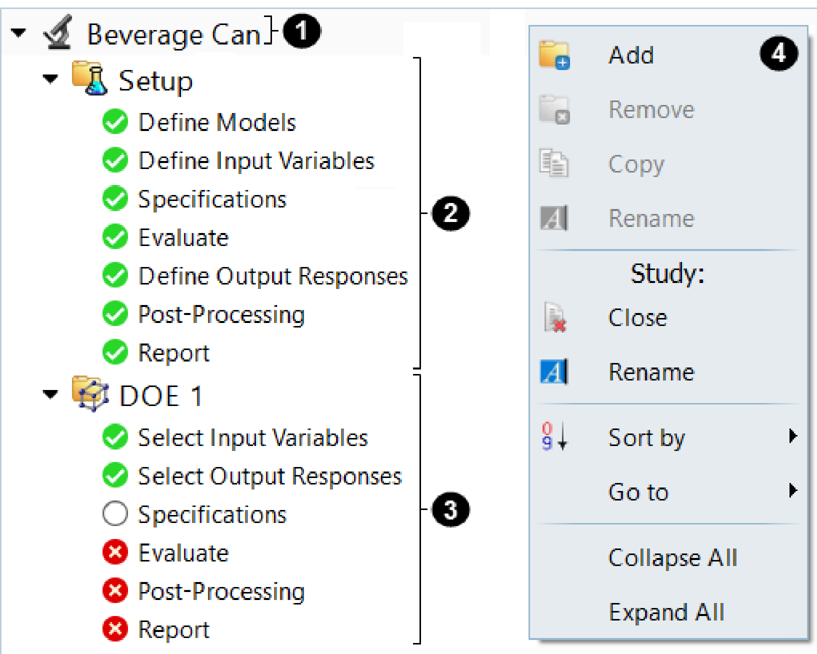


Figure 2: Explorer View

1. Study name; clicking this item guides you to study details and approach details, and enables you to identify batch commands.
2. Steps to setup your study, such as define your model, input variables, and output responses.
3. Steps to define approaches in your study.

   Once the study Setup is complete, an unlimited combination of approaches (DOE, Fit, Optimization and Stochastic) can be added.

4. Additional options are included in the Explorer context menu, such as the options to add, remove, sort, or copy approaches. The Go To option can be used to browse study files, open the study Directory view, or quickly navigate to the study Setup or different approaches.

The steps you must complete while setting up your study and defining approaches display in a hierarchical order. Throughout the study, each steps corresponding checkbox will indicate a status.

	The next step that needs to be completed.
	Proceed to the next step. This step may have an optional work button to perform tasks such as, Evaluate Expressions, Launch Post Processing and Create Report.
	Step has been completed.
	An action in this step is not yet allowed.
	A warning of possible changes in the step since it was first performed.

2.2.4 Directory

The Directory displays all of the files in the current study folder, that is the study in use in the Explorer.

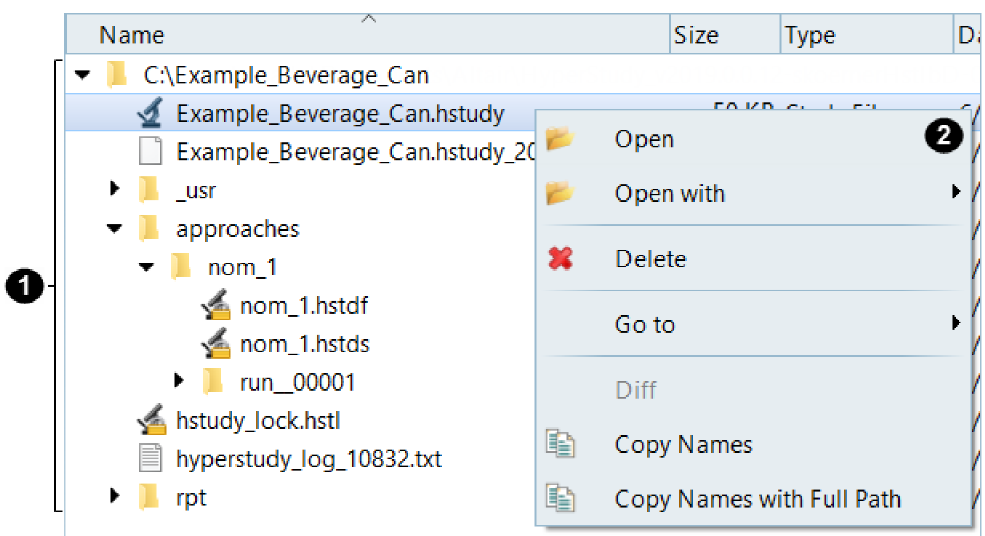


Figure 3: Directory View

1. The study folder contains all of the files in the current study.
   - The study file (`.xml`) at the root level.
   - An `/approaches` folder with one sub-folder per approach (as added in the study Explorer).
   - A report folder (`/rpt`) where all images and HTML files created during the execution of the study are placed.

2. Additional options are included in the Directory context menu, such as options used to open a specific file in the study folder, load output files, and compare to files that can be of different outputs. The Go To option can be used to browse study files or open the study Explorer view.

2.2.5 Work Area

The work area is where you define study models, input variables, output responses, and approach details, and post process results.

> **Note:** The information and options available in the work area vary for each step.

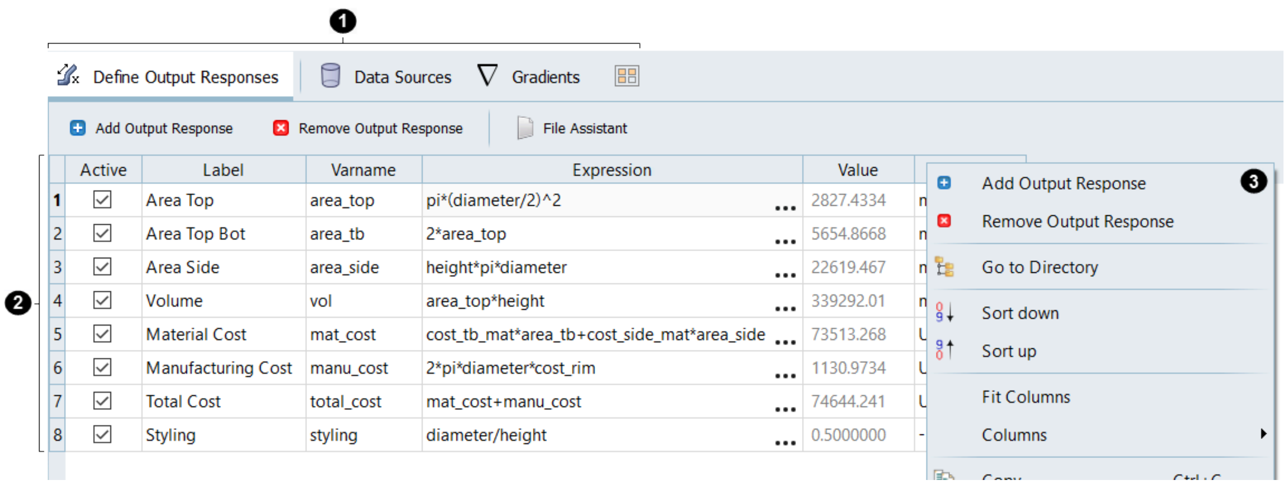


Figure 4: Work Area

1. Tabs separate the different functionality available in each step.

   HyperStudy automatically determines the level of tabs displayed in the work area based on the profile selected from the View menu. You can manually choose which tabs to display by clicking in the tab area and making your selection.
2. Study data related to each step is populated in a table or plot.

   You can edit most of the data displayed in the table. It is possible to copy and paste cells within the application or within an external application (spreadsheet or text editor).
3. Additional options are included in the work area context menu, such as options to sort column data, cut, copy, and paste data into columns, fit table columns, or turn the display of columns on/off.

2.2.6 Message Log Window

The Message Log window, located at the bottom of HyperStudy, logs warnings and errors encountered while you are conducting your investigation.

> **Note:** Messages which are of type Error are always enabled and displayed in the Message Log window. If the Message Log window has been minimized and HyperStudy an error, it will be maximized to alert you of the condition.

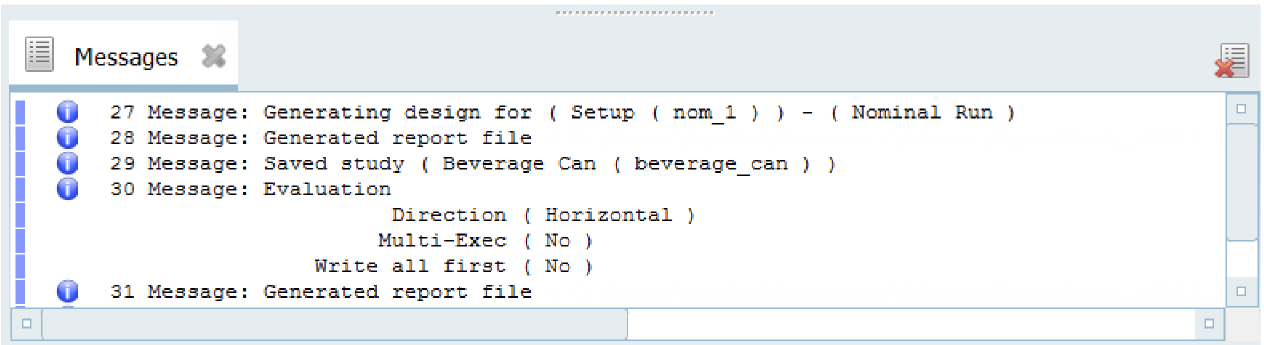


Figure 5:

Message Log Window Settings

Message Log window settings can be accessed from the context menu that opens by right-clicking in the Message Log window.

Setting	Action
Copy	Copy content from the Message Log window, which can then be pasted into another file for your records.
Select All	Select all of the content in the Message Log window.
Clear All	Clear all of the content in the Message Log window.
Show Timestamps	Display the time a message is logged.
Show All	Display of Warning, Info and Verbose class messages, if any or all of these message types have been disabled.
500 Messages	Specify the maximum number of messages to display in the Message Log window at any one time. **Note:** This feature is useful in limiting the amount of memory the Message Log window uses during the session.
Show Warning	Enable or disable the display of warning messages.

Setting	Action
Show Info	Enable or disable the display of information messages.
Show Verbose	Enable or disable the display of verbose messages.
Verbose	Select the level of detail to log for verbose messages. Levels range from zero meaning do not show Verbose to three meaning show as much detail as possible.
Log to File	Display messages in the Message Log window, and log messages in a separate file.
Details	Open the log file or change the file path of the log file.
Info	Display summary information about the current session and message log status. For example: • Time current • Time elapsed • Location of log files in the current HyperStudy session

2.2.7 Status Bar

The status bar, located at the bottom of HyperStudy, displays supplemental information about your study.

The right-side of the status bar displays information about the step you are currently on.

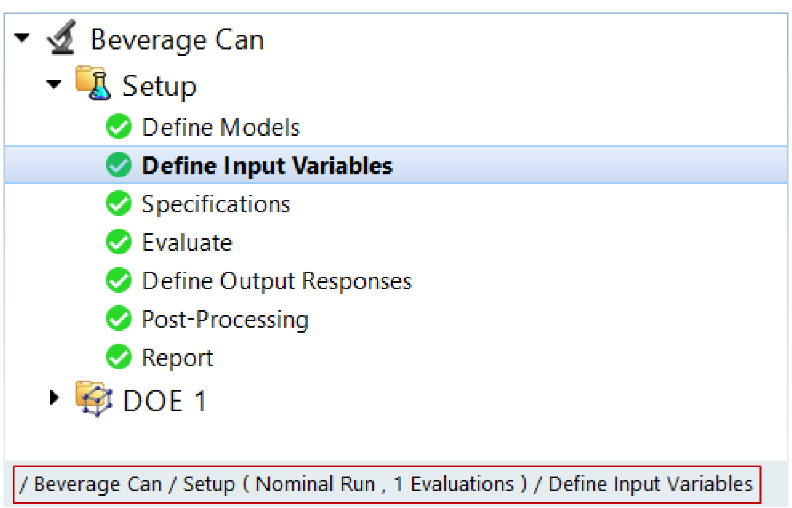


Figure 6: Example: Status

The status reads, `/Beverage Can/Setup (Nominal Run, 1 Steps)/Evaluate` *to indicate that you are currently in the Evaluate step of the Setup in the Beverage Can study and will be performing a nominal run.*

The left-side of the status bar displays the number of errors and warnings in the current study.

> **Tip:** View a detailed list of the warnings and errors by clicking the message in the status bar.

2.2.8 Set HyperStudy Preferences

Configure the view of HyperStudy.

From the menu bar, click **View** to access and set preferences.

Status Bar	Turn the display of the status bar on and off.
Browsers	Turn the display of the Explorer and Directory on and off.
Messages	Turn the display of the Messages pane on and off.
Toolbars	Turn the display of toolbars on and off in the ribbon.
Profile	Choose a profile used to determine the level of tabs displayed in the work area.
Language	Change the language of the user interface.
Font	Change the font family, size, and style.
Display Precision	Change the display precision for numbers in the HyperStudy interface. For example, you can change the display precision from 0.001 to 0.001000.
Color Theme	Choose a predefined color theme that will be applied over the entire user interface.
Reset Prompts to Default	Restore preferences to their default settings.
Full Screen	Display the application in full screen mode.

2.3 Invoke HyperStudy

2.3.1 Run HyperStudy as a Standalone Product

Run in standalone mode for HyperWorks.

To run on	Do this
Windows platforms	From the Start menu, click **Programs** > **Altair HyperWorks** > **HyperStudy**. or Drag and drop a `.hstx` study file into the HyperStudy icon on your desktop to open the study.
PC from the DOS prompt	Run the following executable (where platform is win64): `<install directory>\hst\bin\<platform>\hst.exe -studyfile "<filename>.hstudy" options`
Linux platforms	Invoke via hst in `<install_dir>/scripts/hst`

2.3.2 Run HyperStudy in Batch Mode

Run in batch mode.

To run on	Do this
Windows platforms	Invoke via `hstbatch.exe` in `<install_dir>/hst/bin/winXX`
PC from the DOS prompt	Run the following executable (where platform is win64): `<install directory>\hst\bin\<platform>\hstbatch.exe -studyfile "<filename>.hstudy" options`

To run on	Do this
Linux platforms	Invoke via `hstbatch` in `/Applications/AltairHyperWorks/2021/altair/scripts/hstbatch`.

2.3.3 HyperStudy Start Options

Option	Argument	Description	Supported Platform
-archivefile	<filename>.hstx	Import the specified HyperStudy archive file.	All
-delete	None	Deletes the contents of the run directory before each run.	All
-editfile	<filename>	Launch the HyperStudy Editor without having to open HyperStudy entirely. <filename> is an optional name of the file to open in the Editor.	All
-ex	None	Experimental mode (Ctrl-Alt e).	All
-h or -help or -H	None	Executable documentation.	All Use -H on Linux
-logfile	<filename>	Enable the writing of logfile. If no file specification is provided, a default name will be created (`hyperstudy_log_<pid>.t` in your $HOME area or current working	All

Option	Argument	Description	Supported Platform
		directory if ${HOME} is not defined.	
-multiexec	Number of concurrent jobs	Run HyperStudy batch mode with the multi-execution option.	All
-nobg	None	Run HyperStudy in the foreground and not in the background.	Linux
-o	filename1	Used in conjunction with unit testing.	All
-overwrite	None	Overwrite existing files and directories.	All
-preffile	filename2.mvw	Preferences file for batch _run.	All
-s	None	Display application status information. For diagnostic purposes, –v –s can be used in combination to run in dual mode on Windows. On UNIX, use the –nobg –s option when using the HyperWorks startup script.	All
-studyfile	filename3.hstudy	Load an `.hstudy` file upon opening HyperStudy from the command line.	All
-v	None	Incrementally adjust the Verbose flag (Ctrl-Alt v).	All
-wait	None	Pause the console window.	Windows

2.4 Set HyperStudy Preferences

Configure the view of HyperStudy.

From the menu bar, click **View** to access and set preferences.

Status Bar	Turn the display of the status bar on and off.
Browsers	Turn the display of the Explorer and Directory on and off.
Messages	Turn the display of the Messages pane on and off.
Toolbars	Turn the display of toolbars on and off in the ribbon.
Profile	Choose a profile used to determine the level of tabs displayed in the work area.
Language	Change the language of the user interface.
Font	Change the font family, size, and style.
Display Precision	Change the display precision for numbers in the HyperStudy interface. For example, you can change the display precision from 0.001 to 0.001000.
Color Theme	Choose a predefined color theme that will be applied over the entire user interface.
Reset Prompts to Default	Restore preferences to their default settings.
Full Screen	Display the application in full screen mode.

2.5 Common Use Cases for Setting Up a Study

How you use HyperStudy depends on your model type, your simulation software and your design objectives among other factors. You can use HyperStudy as a standalone software or you can start it from another HyperWorks product.

Study Question or Scenario	Best HyperStudy Option
I have a Radioss model that I want to do a size/ shape optimization study with. I will run the simulations in my PC.	Once you have finished creating your shape variables using the morphing tools in HyperWorks, start HyperStudy from HyperWorks's Applications menu. Your model type is HyperMesh.
I have a MotionSolve model. I will run the simulations in my PC.	Start HyperStudy from MotionView's Applications menu. Your model type is MotionView.
I am using a commercial solver that is not in HyperWorks, but it integrates with HyperWorks (Abaqus, LS-DYNA, DesignLife, and so on.). I would like to set up a size optimization study and will run the simulations in a cluster.	Start HyperStudy in standalone mode. Your model type is Parameterized File. Create the parameterized file, which can be done using the HyperStudy Editor. Register the solver using the Register Solver Script option.
I did my analysis in a spreadsheet. Some of the cells are input variables and some are output responses.	Start HyperStudy in standalone mode. Your model type isSpreadsheet.
I calculated my output responses using analytical equations.	Start HyperStudy in standalone mode. Your model type is Internal Math.

2.6 Common Use Cases for Selecting Approaches

Once you have finished setting up your study, you will need to select one or more study approaches to find the answers to your study questions. The best combination of approaches and the best method to use for each approach depends on your application and objectives.

Study Approach Question or Scenario	Best HyperStudy Approach
Which input variables have a significant effect on my output responses?	Use a parameter screening DOE, such as Fractional Factorial. Once the parameter screening DOE is complete, look at the Linear Effects and Interactions plots.
How can I do quick trade-off studies?	From a space filling DOE, create predictive models in a Fit approach using the Fit Automatically Selected by Training method.
What are the best input variable values to minimize my objective, while meeting my design requirements?	Use a single objective optimization such as Adaptive Response Surface Method or Global Response Search Method (if time permits) to search for an improved solution.
What is the reliability of my design?	Use a Stochastic approach and add your reliability assessment.

Manage Files and Data

3

Learn how to create, open, import and save models.

This chapter covers the following:

3.1 Create Studies

Create a new study.

1. From the ribbon, Study tools, click **New**.

   The **Add** dialog opens.
2. In the Label field, enter a name for the study.
3. In the Location area, navigate to your working directory.
4. Click **OK**.

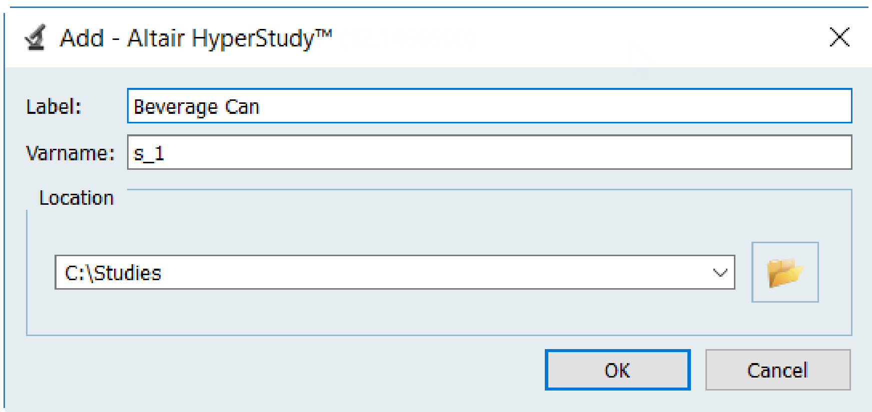


Figure 7:

3.2 Open Studies

Open an existing study.

1. From the ribbon, Study tools, click **Open**.

2. In the HyperStudy **- Open an existing study** dialog, navigate to your working directory and select the study to open.
3. Click **Open**.

> **Tip:** Quickly open a `.hstudy` file by dragging-and-dropping it into the application.

3.3 Save Studies

Save the models, variables, output responses, and approaches defined in your study.

Save your study in one of the following ways:

- Save recent changes made to the current study by clicking **Save** from the ribbon, Study tools.

- Create a copy of the current study by clicking **File** > **Save As** from the menu bar.

3.4 Import Study Archives

Unpackage the study archive (`.hstx`) at a given location.

The `.hstdf` and `.data` files are placed in their respective locations and all of the other files are organized in the study directory. Import Archive also modifies the study directory in the `.hstudy` file to the new location.

1. From the menu bar, click **File** > **Import Archive**.
2. In the **HyperStudy - Open Study Archive** dialog, select the study archive (`.hstx`) and click **Open**.

3.5 Export Study Archives

Packages the study's files into a single `.hstx` file that is easy to handle and share.

The following files are packaged into the `.hstx` file:

- Study `.hstudy` file
- Resource files
- Reference files
- `.hstdf` and `.data` files
- Any other files in study directory

> **Tip:** Archive is the best method for moving studies.

If the resource file and reference files are not in the study directory when Export Archive is selected, then they will be copied over and included in the `.hstx` file. The archived `.hstudy` study file will be updated to point to the new location of the resource file and reference files.

> **Note:** Export Archive does not package run folders in approach directories (such as `approaches/doe_1/run_xxxx`), in order to minimize the archive file size. Files included in resource files that do not reside in the study directory will not be packaged.

1. From the menu bar, click **File** > **Export Archive**.
   The **HyperStudy - Select Location** dialog opens.
2. Navigate to your working directory.
3. In the File name field, enter a name for the archive.
4. Optional: Change the file extension to `.hstxc` to create a custom archive.

   > **Tip:** Custom archives provide additional controls for packaging items such as: model resources, run folders, and data source depots. Custom archives protect sensitive information, such as finite elements models, when sharing the archive. Custom archives also help control the archive size.

5. Click **Save**.

3.6 Package Reports

Package the report directory, `<studyDir>/rpt/…`, in a `.zip` file.

The top level directory in a study is `<studyDir>`. Once an approach is created, a secondary subdirectory, `<studyDir>/rpt`, is created for reports. Each approach then has its own subdirectory within the `rpt` directory.

1. From the menu bar, click **File** > **Package Report**.
   The **HyperStudy - Select Location** dialog opens.
2. Navigate to your working directory.
3. In the File name field, enter a name for the packaged report.
4. Click **Save**.

Perform Studies

4

A study is a self-contained project in which models, variables, output responses, and approaches are defined.

This chapter covers the following:

A study is saved into a XML file, with a `.hstudy` extension. A study file contains a structured list of study statements.

Tip: To generate the structure of study files, pass the argument `-xmltag` to HyperStudy batch. It writes out the `.hstudy` file in the current working directory with all of the structural XML elements.

4.1 Setup the Study

Before you can create approaches you must first setup your Study by defining input variables and output responses.

4.1.1 Edit Files

Use the Editor to create and edit template files, parameterize ASCII files, create variables or shape templates, recognize and edit Templex statements in files, execute parameterized files, and perform advanced editing.

Create, Edit, Remove Parameters

Create, edit, and remove parameters using the HyperStudy Editor.

Before creating and editing parameters, the HyperStudy Editor requires a format to be set. For more information, see Parameter Format Specifiers.

Create Parameters

1. From the menu bar, click **Tools** > **Editor**.
   The Editor opens.
2. In the File field, open the file to be used as the base input template.
3. Highlight the appropriate fields in the text area.

   > **Tip:** To assist in selecting a specified number of fields, press **Ctrl** to activate the Selector, enter the number of fields to select, then click a field in the text area. HyperStudy highlights the number of fields specified.

4. Create a parameter.
   - Click **Create**.
   - Right-click on the highlighted fields and select **Create Parameter** from the context menu.

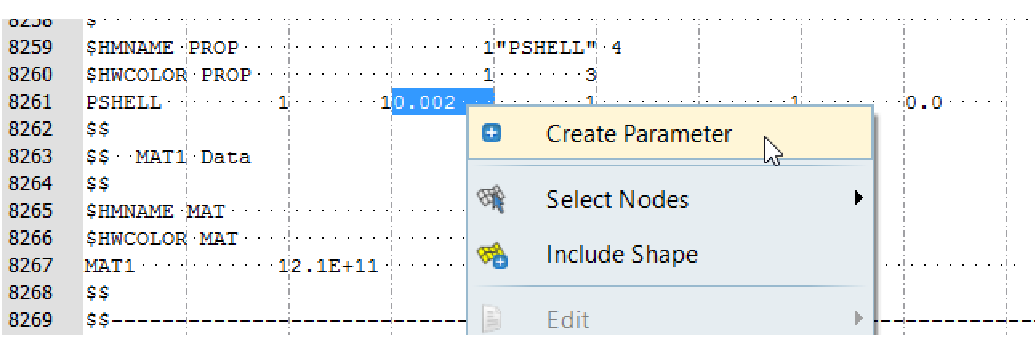


Figure 8:

5. In the **Parameter - varname** dialog, define the new parameter.

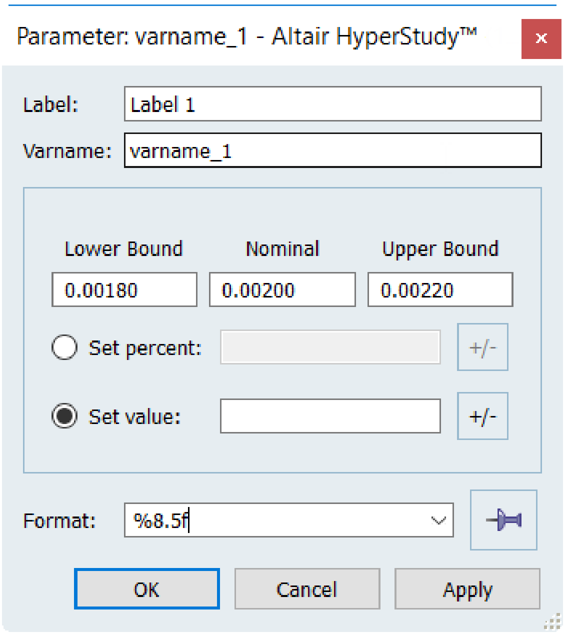


Figure 9:

6. Click **Apply** to accept the input values.
7. Click **OK** to finish creating the parameter.

Edit Parameters

1. From the menu bar, click **Tools** > **Editor**.
   The Editor opens.
2. In the File field, open the file to be used as the base input template.
3. Select the parameter to edit.
   - Highlight the parameter to edit in text area, then right-click and select **Edit** from the context menu.

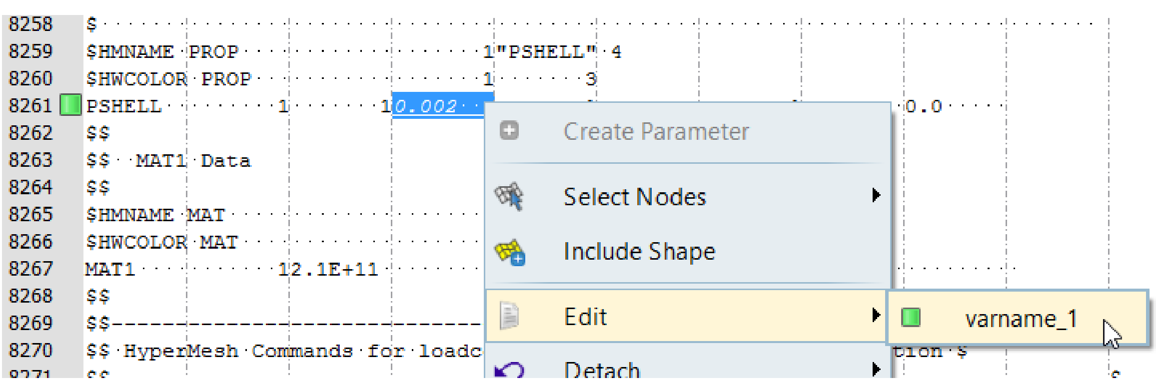


Figure 10:

- In the row that contains the parameter to edit, click and select **Edit** > **varname** from the context menu.

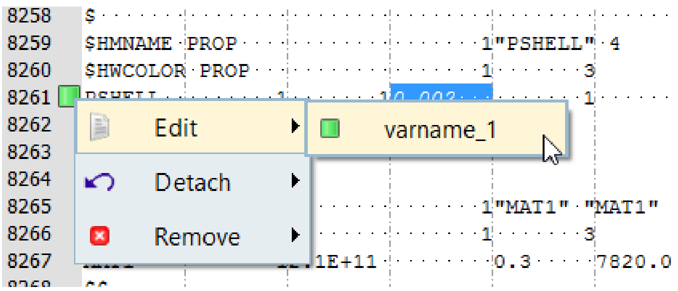


Figure 11:

4. In the **Parameter - varname** dialog, edit the parameter.
5. Click **Apply** to accept the changes.
6. Click **OK** to finish editing the parameter.

Remove Parameters

1. From the menu bar, click **Tools** > **Editor**.
   The Editor opens.
2. In the File field, open the file to be used as the base input template.
3. Remove parameter.
   - Highlight the parameter to remove in text area, then right-click and select **Remove** from the context menu.

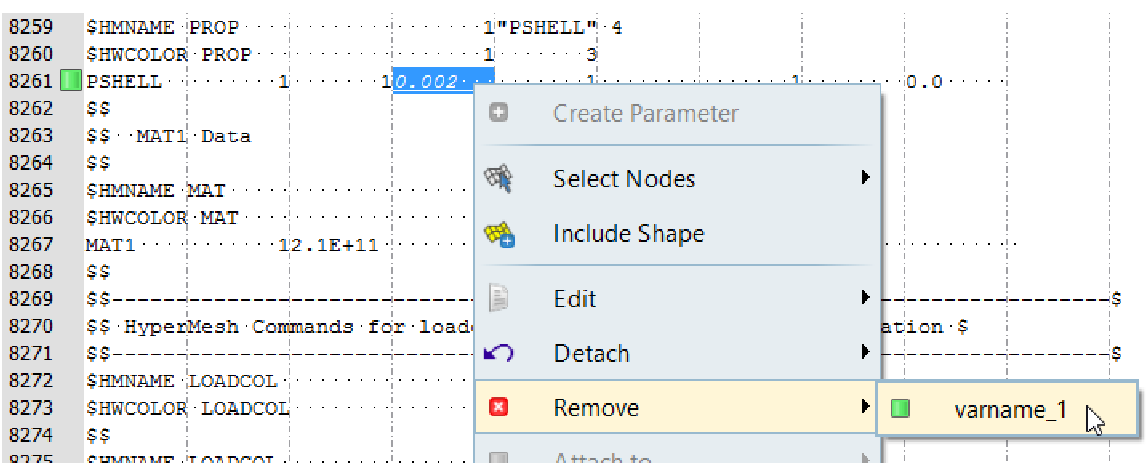


Figure 12:

- In the row that contains the parameter to remove, click and select **Remove** > **varname** from the context menu.

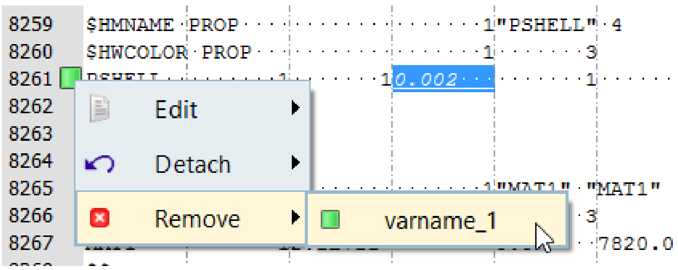


Figure 13:

Attach/Detach Parameters from Existing Parameters

1. From the menu bar, click **Tools** > **Editor**.
   The Editor opens.
2. In the File field, open the file to be used as the base input template.
3. Attach/detach parameters.

Option	Description
Attach parameters to existing parameters	Highlight the appropriate fields in the text area, then right-click and select **Attach to** > **varname** from the context menu.
Detach parameters from existing parameters	Highlight the parameter to detach in text area, then right-click and select **Detach** > **varname** from the context menu. or In the row that contains the parameter to detach, click and select **Detach** > **varname** from the context menu.

Use Shape Variables

Insert Shape Templates

1. From the menu bar, click **Tools** > **Editor**.
   The Editor opens.
2. In the File field, open the file to be used as the base input template.
3. Highlight the fields (usually nodes) to be replaced by a shape template in the text area, then right-click and select **Include Shape** from the context menu.

4. In the **Shape Template** dialog, open the shape template (`.node.tpl` file generated by HyperMesh).

HyperStudy replaces the selected text with an include statement linked to the selected `.node.tpl` file.

```
$
$$
$$ GRID Data
$$
{include "../HS-1560/shapes.dynakey.node.tpl"}
$$
$$ SPOINT Data
$$
$
```

Figure 14:

Select a Group of Nodes

1. From the menu bar, click **Tools** > **Editor**.
   The Editor opens.
2. In the File field, open the file to be used as the base input template.
3. Select all of the GRID, *NODE, or /NODE cards by right-clicking anywhere in the text area and selecting **Select Nodes** > **GRID, *NODE, or /NODE** from the context menu.

HyperStudy selects and highlights all of the nodes for the card you selected.

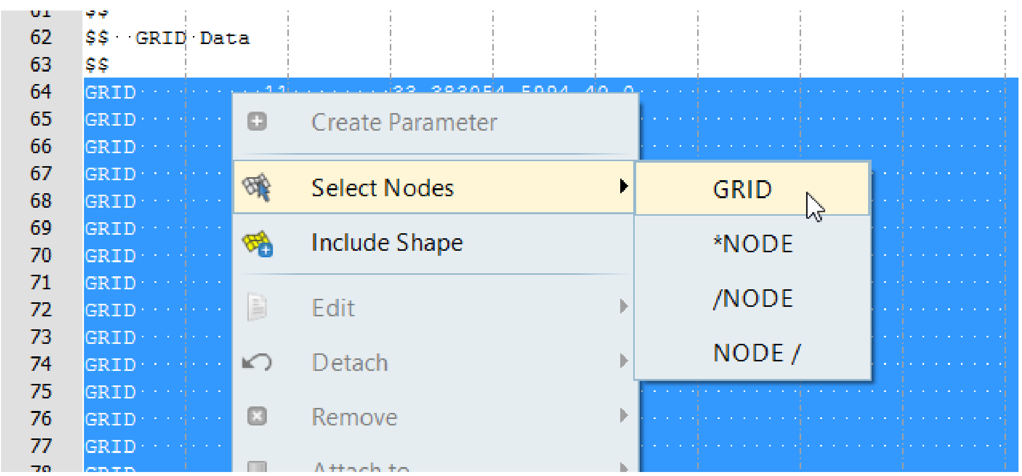


Figure 15:

Parameter Format Specifiers

When text is selected for replacement with a parameter, the HyperStudy editor recognizes the format and size of the highlighted text. This specifier controls how the data is written to the file. The formatting should be adjusted as required.

A specifier begins with a % and ends with an alphabetic character that controls the basic format type.

f	Floating point number
i	Integer
e	Exponential
s	String

Between the % and the type character is a number to define the width of the field.

%8i	An integer printed to be 8 characters wide
%7s	Prints a string 7 characters wide

For numeric data, the number of digits of precision can be controlled explicitly by inserting a period and a number after the format. Consider the number -5:

- Formatting -5 with %8.5f will result in -5.00000, a floating point number 8 characters wide with precision 5.
- Formatting 123.456789 with % 8.2e will result in 1.23e+02.

> **Note:** When additional precision is needed in field formats like OptiStruct and Nastran, consider using free format the individual line.

By convention, the formatting is right justified. In order to force left justification, the – character can be used before the field width.

- 5.1 formatted as %-8.2 will produce "5.10 ". Note that the quotes in this example are not part of the formatting but used to show the whitespace padding.

The format specifiers in HyperStudy conform to conventional syntax. A more detailed explanation on the syntax can be found in references on computer programming.

Editor Settings

Settings to configure the Editor.

Annotations	Displays a summary of all the templex statements in the file at the top of the text area, and displays a templex statement underneath every parameter.
Whitespaces	Displays a black dot for all of the spaces and tabs in the file.
Line numbers	Displays the line number for each line in the file.
Overview and cursor position	Provides an overview of the open file, and displays a range of selection when you select a string of fields in the file.
Column guides	Displays column lines.

4.1.2 Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

Define Models

A model is part of a study. It is the model of the system that is subjected to a study.

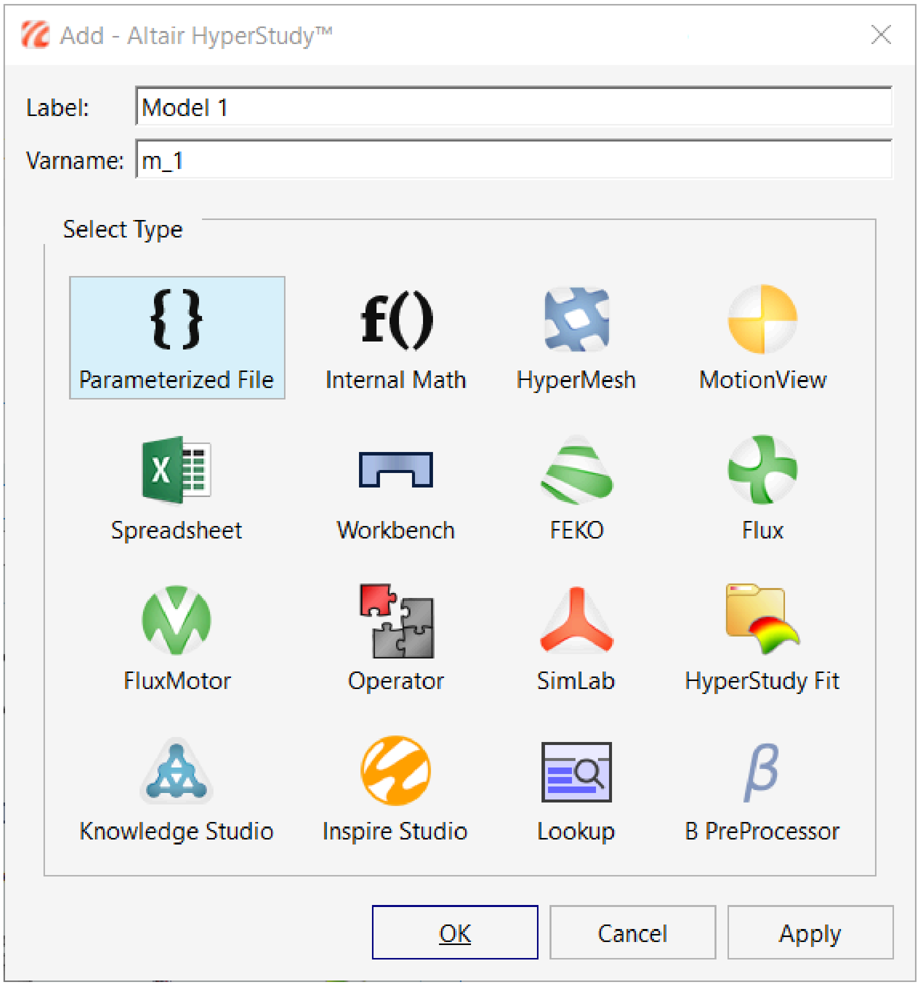
Add and Delete Models

In each study, you can add one or more models.

Add Models

Add a model to your study.

1. In the Define Models step, click **Add Model**.
2. In the **Add - HyperStudy** dialog, define model settings and click **OK**.
   a) In the Label field, enter a name for the model.
   b) Select a model type.



Tip: Quickly add a model by dragging-and-dropping a supported model file directly into the work area of the Define Models step.

Delete Models

Delete models from your study.

1. In the Define Models step, select the model in the work area that you wish to delete
2. Click **Remove Model**.

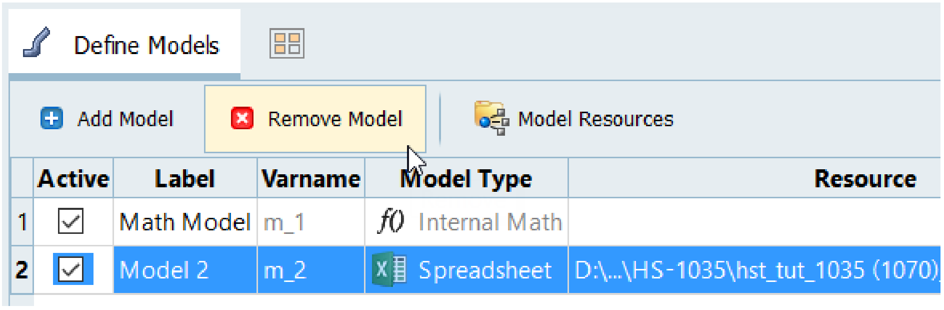


Figure 16:

Model Types

Learn about the different type of models available in HyperStudy.

Beta Preprocessor Model

Use the Beta Preprocessor model to create a parametric connection to a B Preprocessor database enabling you to generate solver input files.

Version

N/A

File Extension

`.ansa`

Multi-Execute Support

Yes

Response Definition Support

No

Registration Steps

Register a solver script that points to an ansa startup.

For example, `...BETA\ansa_v17.1.2\ansa64.bat`.

Usability Characteristics

The FE output path in a `.ansa` file should be relative. If it is not, you will receive a warning when importing input variables.

Once the optimization task is defined, save the current model as `<modelName>.ansa`. Verify the dvfile name follows a naming convention of `<modelName>_dvfile.txt` and is in the same directory as the `.ansa` file. For example, `feasolver_beam.ansa`, `feasolver_beam_dvfile.txt`.

FEKO Model

Use the FEKO model to connect to a Feko project, which enables you to automatically detect input variables and run high frequency electromagnetic simulations.

Version

2017 or above

File Extension

`.cfx`

Multi-Execute Support

Yes

Response Definition Support

No

Registration Steps

Register a solver script that points to `runfeko.exe`.

Usability Characteristics

Output responses from Feko are written to the `hst_output.hstp` file by a custom `.lua` script. If no `.lua` script is present, Feko will generate a nominal script which performs a trivial calculation. This nominal script must be modified to request the specific output response of interest from Feko.

Flux Model

Use the Flux model to connect to any Flux 2D and 3D project, which enables you to automatically detect input variables and run low frequency electromagnetic simulations.

Version

Flux 12.1 or above

File Extensions

`.F2G`
`.F2HST`

Multi-Execute Support

Yes

Response Definition Support

Yes

Registration Steps

Register a solver script that points to `flux.exe`. For more information, refer to Register Solver Scripts.

Usability Characteristics

- Some Flux outputs (such as curve value extractions) require a post-processing python file to be executed after each simulation. If a python command file is required, it should be specified in the `.F2G` or `.F2HST` file. Drag and drop the `.F2G` or `.F2HST` file into HyperStudy to connect the model and identify the inputs and the outputs. Once a connection is established, HyperStudy can automatically run as many Flux simulations as requested by the method (DOE, Optimization).
- For information on how to expose Flux model parameters and outputs to HyperStudy by means of a `.F2G` or `.F2HST` coupling file, refer to the Flux help and HyperStudy dedicated tutorial.

FluxMotor Model

Use the FluxMotor model to connect to a FluxMotor model.

Version
: 2018.1

File Extension
: `.fm2hst`

Multi-Execute Support
: Yes

Response Definition Support
: Yes

Registration Steps
: Register a solver script that points to `FluxMotor_Install_dir/Scripts/win/FluxMotors.exe`.

Usability Characteristics
: Refer to the FluxMotor help for more information on how to expose FluxMotor model parameters to HyperStudy.

HyperStudy Fit Model

Use the HyperStudy Fit model to import the python based report created by a Fit approach from another HyperStudy session.

Version
: N/A

File Extension
: `.pyfit`

Multi-Execute Support
Yes

Response Definition Support
Yes

Registration Steps
N/A

HyperMesh Model

Use the HyperMesh model to directly access the HyperMesh database of the model, which enables you to perform easy design parameterization of solver input files.

Version
Current HyperWorks installation

File Extension
`.hm`

Multi-Execute Support
Yes

Response Definition Support
No

Registration Steps
HyperMesh is automatically registered in HyperStudy.

Usability Characteristics

- Input variables are limited to solver interface specific presets and general parameter entities.
- Can be used with any solver script.

Internal Math Model

Use the Internal Math model to create input variables that are contained entirely within HyperStudy and do not map to any external applications.

f()

Version
N/A

File Extension
N/A

Multi-Execute Support
Yes

Response Definition Support

No

Registration Steps

N/A

Usability Characteristics

- Used to link variables between models.
- Used for variables used directly in response expressions.

Knowledge Studio Model

Use a Knowledge Studio model to import the predictive model from a Knowledge Studio session.

Version

2020.1 or above

File Extensions

`.kdm`

Multi-Execute Support

Yes

Response Definition Support

Yes

Registration Steps

Register a solver script that points to `<...>/Altair Knowledge Works/KS Workstation 2020.1.0/HyperStudy/runKS.bat` for Windows and `<...>/Altair Knowledge Works/KS Workstation 2020.1.0/HyperStudy/runKS.sh` for Linux.

Usability Characteristics

- Verify that the path to the Python executable is valid in the Knowledge Studio installation configuration file, `KS Workstation 2020.1.0/HyperStudy/runKS.cfg`.

Lookup Model

Use the Lookup model to extract values from a file with known input/output pairs.

Version

N/A

File Extension

`.csv`

Multi-Execute Support
Yes

Response Definition Support
Yes

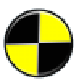
Registration Steps
N/A

Usability Characteristics

- The `.csv` file must be structured so that all of the first N columns are imported as input variables, and all remaining columns are imported as output responses.
- Evaluation of the model with input variables that have no matching row will result in an execution failure.

MotionView Model

Use the MotionView model to directly access the HyperView database of the model, which enables you to perform easy design parameterization of solver input files.



Version
Current HyperWorks installation

File Extension
`.mdl`

Multi-Execute Support
Yes

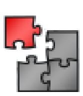
Response Definition Support
No

Registration Steps
MotionView is automatically registered in HyperStudy.

Usability Characteristics
Can be used with any solver script or the have MotionView run the solver defined inside itself.

Operator Model

Use the Operator model to execute a process but not have an associated write operation.



Version
N/A

File Extension
: N/A

Multi-Execute Support
: Yes

Response Definition Support
: No

Registration Steps
: N/A

Usability Characteristics
: Frequently used in conjunction with linked model resources in order to define the steps of a process flow. A process can be defined as a sequence of HyperStudy models, including Operators, rather than a single model with a single detailed and non-reusable solver script.

Parameterized File Model

A Parameterized File model is a general ASCII text file, which has search and replace substitution for input variables.

{ }

Version
: N/A

File Extension
: `.tpl`

Multi-Execute Support
: Yes

Response Definition Support
: No

Registration Steps
: N/A

Usability Characteristics

- A Templex language syntax is used inside of the parameterized file.
- The formatting in the parameterized file is independent of the internal storage.

SimLab Model

Use the SimLab model to parametrically drive SimLab macros, which can be used to automatically mesh and solve FEA problems.

Version

2019.1 minimum

File Extension

.js

Multi-Execute Support

No

Response Definition Support

Yes

Registration Steps

Register a solver script that points to bin/<platform>/SimLab.bat.

Usability Characteristics

- The FEA solver is run as part of .js script.
- The HyperStudy directory (.hstudy) and the SimLab project directory cannot be the same.

Spreadsheet Model

Use the Spreadsheet model to perform calculations.

Version

Microsoft Office

File Extension

.xls

.xlsx

.xlsm

Multi-Execute Support

No

Response Definition Support

Yes

Registration Steps

N/A

Usability Characteristics

- Input variable and output response cells are identified by the user.
- When creating a Spreadsheet model an input variable's value and label can be formatted in two consecutive rows or two consecutive columns. Input variable labels should only contain English characters, or a combination of English characters and numbers. If you do not create a label for a variable, HyperStudy will assign one by default.

- On a Windows operating system, Spreadsheet models will be brought to the front. HyperStudy will prevent you from accessing any spreadsheets that you had opened prior to or after (by double-clicking) starting your current HyperStudy session.

  To access an Excel spreadsheet while running a study with Spreadsheet models, you must explicitly start a new session of Excel.
- Windows platform only
- Support for VB macros

Workbench Model

Use the Workbench model to connect to an ANSYS workbench project.

Version
: 17.1

File Extension
: `.wbpj`

Multi-Execute Support
: No

Response Definition Support
: Yes

Registration Steps
: Register a solver script that points to `runwb.exe`.

Usability Characteristics

- Windows support only
- Input variables are automatically detected from Workbench input parameters.
- Output responses are automatically detected from Workbench direct outputs.

Load Resource Files

A resource file establishes the basis of the connection to HyperStudy. Information about the input variables and/or output responses are passed through the file.

If you are defining an Interanl Math model, dragged-and-dropped a model file into the application, or invoked HyperStudy from another HyperWorks Desktop application, you will not be required to manually load a resource file.

1. In the Define Models step, Resource column of the model you are defining, click .
2. In the **HyperStudy - Load model resource** dialog, navigate to your working directory and open the resource file.

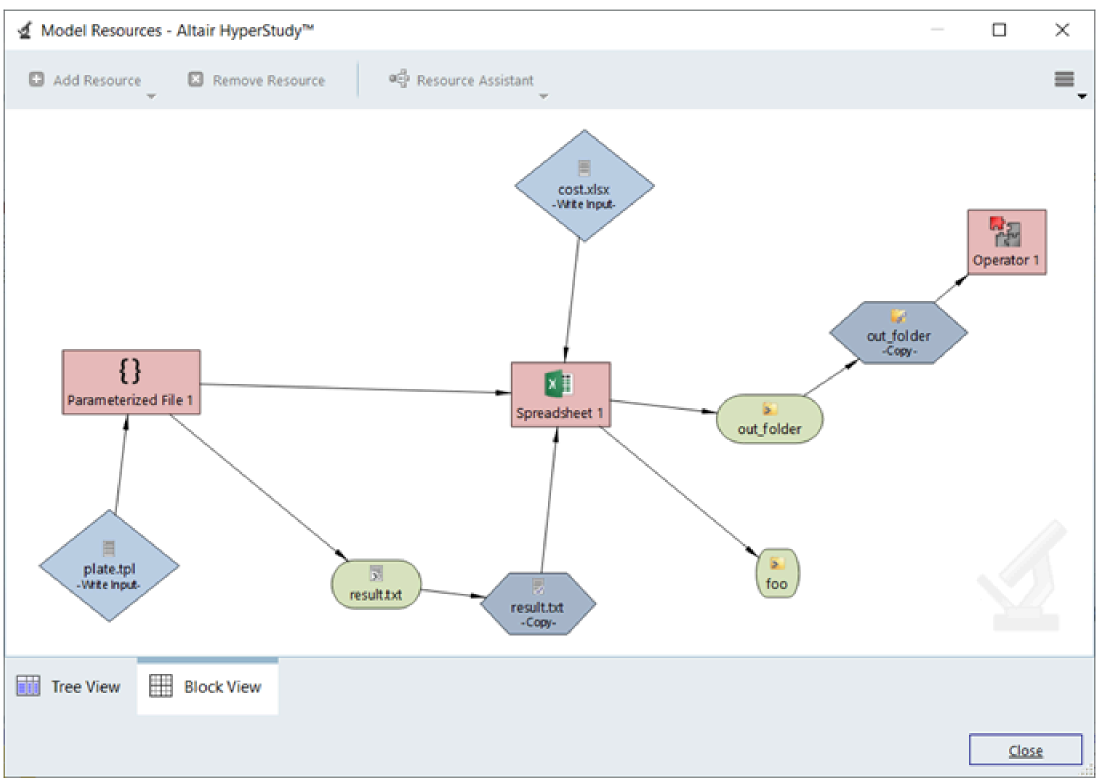
> **Note:** If you load a file for a Parameterized File model that does not contain parameters, a dialog will appear asking if you would like to parameterize the file. If you select **Yes**, the file will open in the **Editor** as a `*.tpl` file. When you close the **Editor**, HyperStudy automatically loads the`*.tpl` in the Resource column, and inserts the file name in the Solver input file column.

Define Model Dependencies

When defining a model, model resources identify additional files required for the model to evaluate successfully, for example reference files called by script or linked result files that were generated by other models.

Varnames can be referenced in the Solver Input Arguments field, for example $file $filebase, and so on. The proper syntax is `${varname}`, for example `${m_1.file_2}`.

1. In the Define Models step, click **Model Resources**.
   The **Model Resources** dialog opens.
2. Optional: Graphically view the model.
   a) In the **Model Resources** dialog, click **Block View**.
   b) Click and drag items to rearrange the view of the model.
   c) Use the options menu to determine which items display in the window.



3. Select the model to which the resource will be assigned.

4. Click **Add Resource** and select the type of resource to add.
   - **Add Input File** adds a file that does not change.
   - **Add Output File** tags a file that can be referenced by a link resource.
   - **Add Input Folder** adds a folder that does not change.
   - **Add Output Folder** tags a folder that can be referenced by a link resource.
   - **Add Link** adds an Output File or Output Folder that changes with each run.
5. In the **Select** dialog, navigate to your working directory and open the file/folder/resource to reference.

   > **Tip:** To correct mistakes, right-click the resource file and select **Modify Resource** from the context menu. The resource is maintained and the operation is performed on a new file.

6. In the Operation column, select an option.
   - **None** leaves the file as is. The file is able to be referenced from its existing location.
   - **Copy** copies the resource into the run directory.
   - **Move** moves the resource into the run directory (this is only available for Link Resources).
   - **Copy Content** copies all resources from the selected folder into the run directory.
   - **Move Content** moves all resource from the selected folder into the run directory.
   - **Write Input** is a special case for the resource file for the model (for example, the HyperMesh database for a HyperMesh model type).
7. Click **Close**.

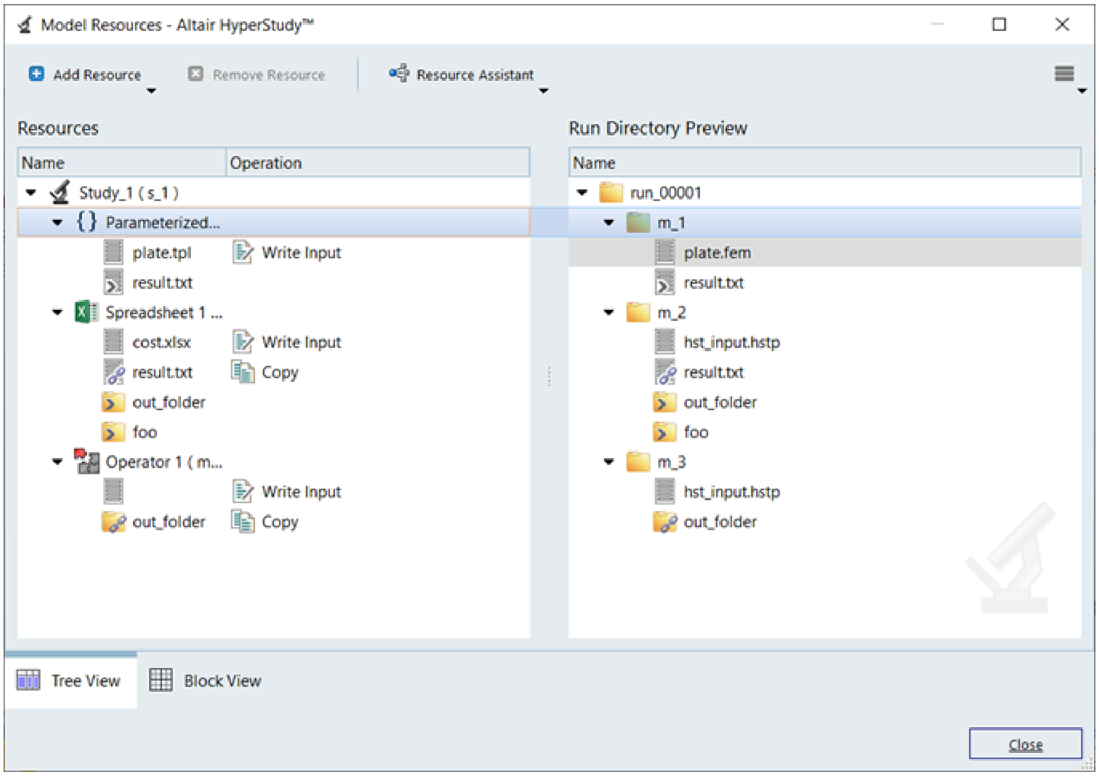


Figure 17:

> **Tip:**
>
> - View Include files and links between models in the run directory by enabling **Run Directory Preview**.
> - For guided help defining a model resource, click **Resource Assistant**.

Specify Solver Input Files

A solver input file is the name of the input file that the model will write during the evaluation.

In HyperStudy, a solver input file is required in order to run the solver. In the Define Models step, you can specify what solver input file you would like to run the solver with.

In the Define Models step, Solver Input File column of the model you are defining, enter the solver input file name.

Select Solver Execution Scripts

Select the solver execution script you would like HyperStudy to use when executing the model.

A set of default solver codes are registered in HyperStudy. You will need to register solver scripts if the solver is not a HyperWorks solver and hence is not registered by default, or in cases where you need to perform a series of actions such as copying files, running one or more solvers, and extracting data.

In the Define Models step, Solver Execution Script column of the model you are defining, select a solver script.

Specify Solver Input Arguments

A solver input argument is any argument to be passed to the solver.

The default argument is ${file}, which means that the qualified solver input file name is passed to the solver script.

In the Define Models step, Solver Input Arguments column of the model you are defining, enter solver input arguments.

> **Note:**
> If the path to the file contains spaces, wrap the argument with quotes "${file}" or remove the path using ${filebasename}.

> **Tip:** Access a list of HyperStudy specific variables and solver specific options from the Solver Input Arguments field by clicking .
> Depending on the mode type selected, a default list of arguments is provided for specific Altair solvers. You can optionally attach your own help for any registered solver.

Example: Solver Input Arguments

Examples solver input arguments for various solvers and Linux.

In these examples, ${file} refers to the solver file that is passed to the solver script.

Radioss	${file} -both
OptiStruct	${file} -scr C:\temp
Compose/OML	-f ${file}
Excel	My_Vb_script1
LS-DYNA	i=${file} MEMORY=5000000
Nastran	Batch=no
Flux	-env_FLUX_NCORES=1 -env_MEMSIZN3=...
MADYMO	-fg <filename>.xml
Matlab	To run a script called test.m: matlab -nosplash -noFigureWindows -wait -r "try; run('test.m'); catch; end; quit"

	`matlab -nosplash -noFigureWindows -wait -r "try; run('${filebasename}'); catch; end; quit"`
Abaqus	`job=<filename>.inp memory=200Mb interactive` You may need to edit the Abaqus environment file (ex: `<ABAQUS INSTALL>\v6.11\6.11-1\site\abaqus v6.env`) file to include `ask_delete=OFF` or comment the line `ask delete=on` if any. This is needed as Abaqus prompts the user if they want to overwrite the old files when rerunning the analysis. In order to eliminate the need for user interaction we need to command Abaqus not to ask this question and overwrite.
Solver Scripts Running on Linux	`-nobg`

Example: Specify Multiple Solver Input Arguments

Solver input arguments are passed to the solver script as separate arguments. The script then treats them as %1 (first argument) , %2 (second argument), and so on.

In Figure 18, `file1.txt` is the model file updated by HyperStudy, and `file2.txt` and `file3.txt` are additional files needed for the solver (that is include files).

	Active	Label	Varname	Model Type	Resource	Solver Input File	Solver Execution Script	Solver Input Arguments
1	☑	Model 1	m_1	{} Parameterized File	C:\model1.tpl	model1.txt	mySolverScript (scr_16)	${file} file2.txt file3.txt

Figure 18:

Having these three files listed in the Solver Input Arguments field allows `file1.txt` to be updated and ensures that all three files are copied into the appropriate folder together. The solver input arguments ensures that `${file}` (= `file1.txt`), `file2.txt` and `file3.txt` are submitted to the solver script.

Figure 19 shows a sample solver script (`.bat` file) for this example.

```
echo %1
echo %2
echo %3
```

Figure 19:

Which results in:

```
"C:\TEST\nom_run\m_1\file1.txt"
file2.txt
file3.txt
```

It is possible to improve the solver script using:

`%1`	Full file name passed by HyperStudy to the solver script

%~n1	File name without extension and path
%~x1	File extension

You could also improve the solver script using a HyperStudy environment variables such as %STUDY_DIR_PATH% (to get the current path) or %STUDY_RUN_NUMBER% (to get the current run number).

Example: Excel Connection Solver Input Arguments

This example illustrates a Spreadsheet model type with solver input arguments for an Excel connection.

In Figure 20 the macro `my_VB_script_1` will be executed for each solver run.

	Active	Label	Varname	Model Type	Resource	Solver Input File	Solver Execution Script	Solver Input Arguments
1	☑	Model 1	m_1	Spreadsheet	C:\....	hst_input.hstp	SpreadSheet (SpreadSheet_HST)	my_VB_Script1

Figure 20:

Use Environment Variables as Solver Input Arguments

In the case of the Excel connection, solver input arguments also recognize a number of environment variables that can be used to pass information to the VB macro used in the spreadsheet.

$file, $filespec	Example: `c:\studyfolder\approaches \doe_1\run__00001\plate.fem`
$filebasename	Example: `plate.fem`
$studydir	Example: `c:\studyfolder`

Pass Current Run Folder Name to the Macro

You can pass the name of the current run folder to the macro, for example `my_VB_Script1 $filespec`.

	Active	Label	Varname	Model Type	Resource	Solver Input File	Solver Execution Script	Solver Input Arguments
1	☑	Model 1	m_1	Spreadsheet	C:\....	hst_input.hstp	SpreadSheet (SpreadSheet_HST)	my_VB_Script1 $filespec

Figure 21:

Define Input Variables

An input variable is a system parameter that influences the system performance in the chosen output response. It is an object that is varied by the study based on certain rules.

Input variables can be of different types with respect to their physical and/or numerical nature. Physically they can be dimensions, material properties, locations, and so on. Numerically they can be continuous numbers, discrete, integer, real, strings.

You can associate an input variable to a particular model parameter. Two input variables cannot be associated with the same model parameter. An input variable does not always need to be associated to a model parameter.

Optimization Input Variables

Input variables should be independent of each other. Selection of input variables may not be unique. In Figure 22, either r_i and r_o or r_i and t can be selected as input variables.

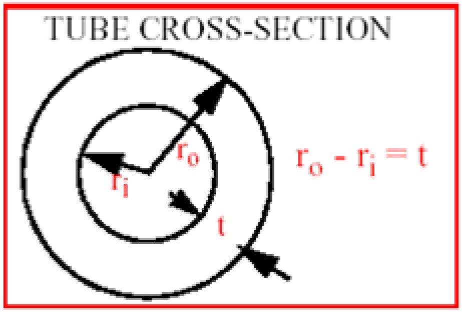


Figure 22: Example: Input Variable Definition

> **Note:** The selection of input variables affects the Optimization process.

For a larger search space and increased flexibility in the solution, you need to use as many independent input variables as possible. For computational affordability, you need to identify and only include the most important input variables.

It is recommended that you use as many independent input variables as possible to promote a large search space and increased flexibility. However, for computational affordability you should only include the most important input variables.

Add and Remove Input Variables

Input variables can be added to your study manually or imported.

Import Input Variables from HyperMesh or MotionView

Quickly add input variables to your model by importing them from a HyperMesh or MotionView session.

1. In the Define Models step, click **Import Variables**.
   The **Model Parameters** dialog opens.
2. Select the parameter(s) to add as input variables from the panel on the left.

   > **Note:** For HyperMesh models, you can edit the upper and lower bounds of the input variables during this step.

3. Click **Add**.
   The parameters you selected for definition as input variables are appended to the panel on the right.
4. Click **OK** (HyperMesh) or **Done** (MotionView).

The model parameters that were appended to the panel on the right are added to the list of input variables in the Define Input Variables step. Each model parameter is associated with an input variable

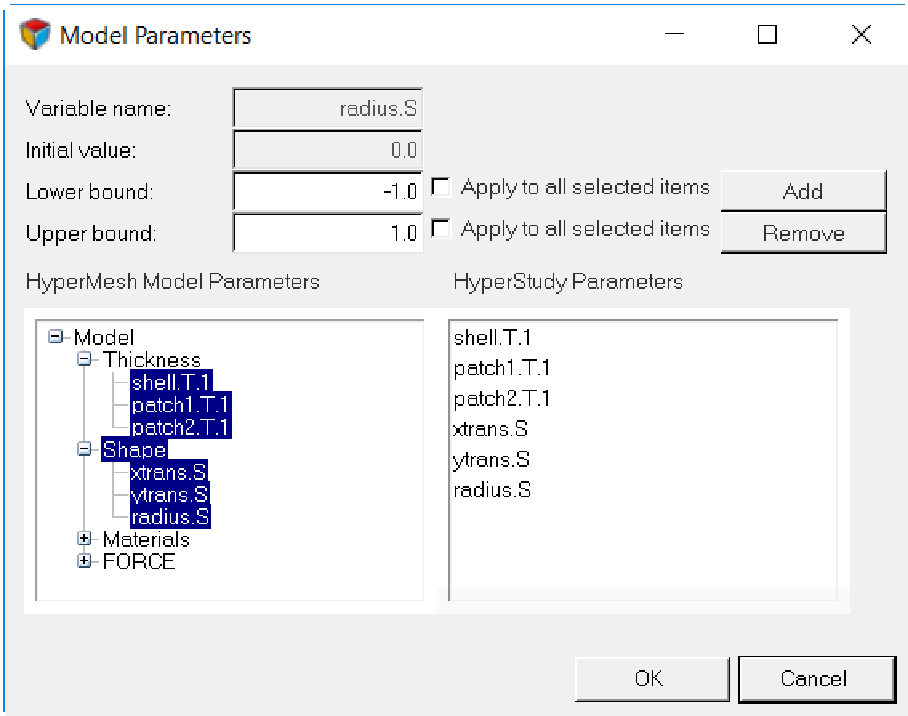


*Figure 23: HyperMesh **Model Parameter** Dialog*

.

Import Input Variables and Output Responses from Excel Spreadsheets

Quickly add input variables to your model by importing them from an Excel spreadsheet.

Before you begin importing input variables from an Excel spreadsheet, it is import that the data in your spreadsheet is formatted correctly. For more information, refer to the Usability Characteristics section in Spreadsheet Model.

1. In the Define Models step, click **Import Variables**.
   The Excel spreadsheet opens.
2. Select input variables.
   a) In the **Excel** dialog, click **Yes** to begin selecting input variables.
      The **Excel - HyperStudy Input Selector** dialog opens.
   b) In the spreadsheet, select the input variable's value and label cells.

      > **Note:** You can select an input variable's value and label in two consecutive rows or two consecutive columns.

      The selected cells display in the **Excel - HyperStudy Input Selector** dialog.
   c) Click **OK** to import the selected input variables.
   d) To continue selecting input variables in other areas of the spreadsheet, repeat steps 2.b and 2.c.

e) Click **Cancel** in the **Excel - HyperStudy Input Selector** dialog to stop selecting input variables.

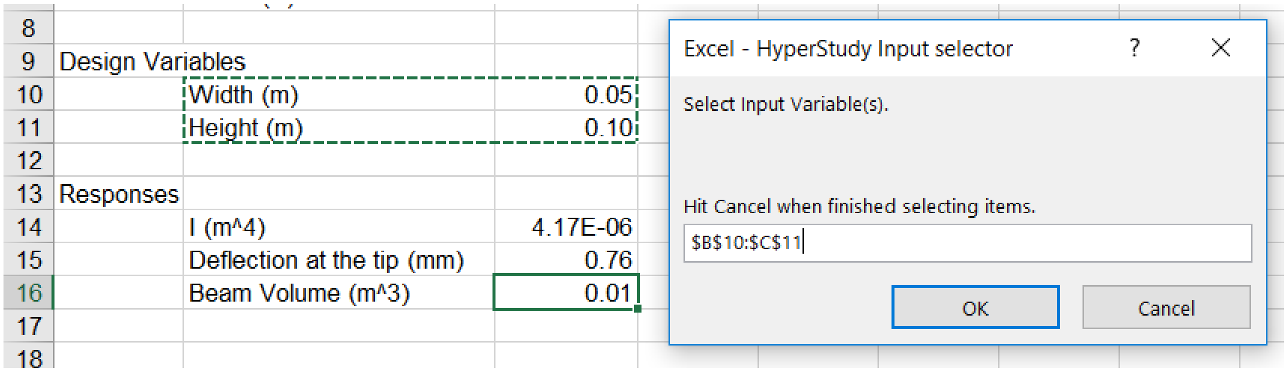


Figure 24:

3. Select output responses.

a) In the **Excel** dialog, click **Yes** to begin selecting output responses.
The **Excel - HyperStudy Input Selector** dialog opens.

b) In the spreadsheet, select the output responses' value and label cells.

> **Note:** You can select an output responses' value and label in two consecutive rows or two consecutive columns.

The selected cells display in the **Excel - HyperStudy Input Selector** dialog.

c) Click **OK** to import the selected output responses.

d) To continue selecting output responses in other areas of the spreadsheet, repeat steps 3.b and 3.c.

e) Click **Cancel** in the **Excel - HyperStudy Input Selector** dialog to stop selecting output responses.

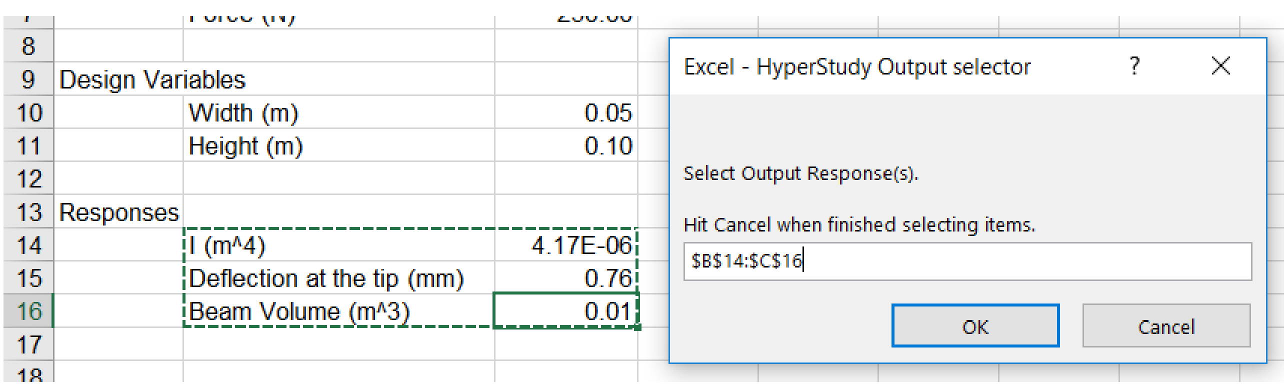


Figure 25:

The imported input variables are added to the list of input variables in the Define Input Variables step, and the imported output responses are added to the list of output responses in the Define Output Responses step.

Create Input Variables

Manually create input variables in HyperStudy.

1. In the Define Input Variables step, click **Add Input Variable**.
   A new input variable is created and added to the list of input variables.
2. Define the input variable by modifying its corresponding cells in the work area.
   a) Enter a label.
   b) Specify a lower, nominal, and upper bound.
   c) Optional: Change the data type, mode, distribution role, and so on from the Modes and Distributions tabs.

	Active	Label	Varname	Lower Bound	Nominal	Upper Bound	Comment
1	☑	Diameter	diameter	30.000000	60.000000	90.000000	mm
2	☑	Height	height	60.000000	120.00000	180.00000	mm
3	☑	Thick Top	thick_top	0.2000000	0.2500000	0.3000000	mm
4	☑	Thick Side	thick_side	0.1000000	0.1200000	0.1400000	mm
5	☑	Cost Top Bot Material	cost_tb_mat	2.0000000	5.0000000	8.0000000	USD/mm2
6	☑	Cost Side Material	cost_side_mat	1.0000000	2.0000000	3.0000000	USD/mm2
7	☑	Cost Rim Manufacturing	cost_rim	1.5000000	3.0000000	4.5000000	USD/mm

Figure 26:

Tip: Create multiple input variables simultaneously by left-clicking and holding the mouse button on **Add Input Variable**. In the pop-up, enter the number of input variables to add and press **Enter**.

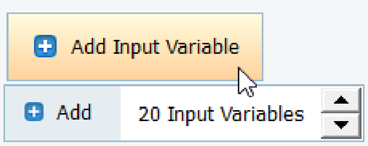


Figure 27:

Remove Input Variables

Remove input variables from your study.

1. In the Define Input Variables step, select the input variable to remove.
2. Click **Remove Input Variable**.

Add Input Variable | Remove Input Variable

	Active	Label	Varname	Lower Bound	Nominal	Upper Bound	Comment
1	☑	Diameter	diameter	30.000000 ...	60.000000 ...	90.000000 ...	mm ...
2	☑	Height	height	60.000000 ...	120.00000 ...	180.00000 ...	mm ...
3	☑	Thick Top	thick_top	0.2000000 ...	0.2500000 ...	0.3000000 ...	mm ...
4	☑	Thick Side	thick_side	0.1000000 ...	0.1200000 ...	0.1400000 ...	mm ...
5	☑	Cost Top Bot Material	cost_tb_mat	2.0000000 ...	5.0000000 ...	8.0000000 ...	USD/mm2 ...
6	☑	Cost Side Material	cost_side_mat	1.0000000	2.0000000	3.0000000	USD/mm2

Figure 28:

Tip: Select multiple input variables to remove by holding Ctrl while left-clicking.

Deactivate Input Variables

By default, all input variables imported or manually added to your study are made active. Make an input variable inactive so that it is no longer an independent variable.

Reducing the number of independent variables can reduce the number of runs required in an Optimization or DOE approach.

In the Define Models step, Active column, clear the input variable's corresponding checkbox.

	Active	Label	Varname	Lower Bound	Nominal	Upper Bound
1	☑	Thickness 1	m_1_varname_1	0.0018000 ...	0.0020000 ...	0.0022000 ...
2	☐	Thickness 2	m_1_varname_2	9.00e-04 ...	0.0010000 ...	0.0011000 ...
3	☑	Thickness 3	m_1_varname_3	0.0045000 ...	0.0050000 ...	0.0055000 ...
4	☐	Thickness 4	m_1_varname_4	0.0018000 ...	0.0020000 ...	0.0022000 ...

Figure 29:

Link Input Variables

Reduce the number of independent variables and/or of different models in multi-model studies in order to have synchronized design updates between models by linking input variables to each other.

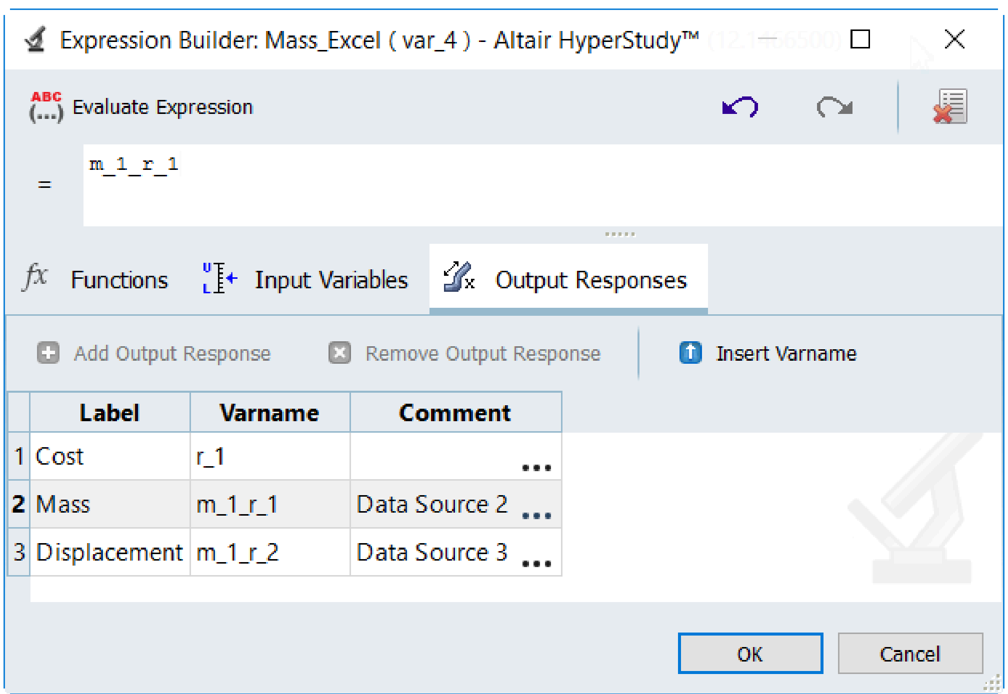
Input variables can be linked within or across models.

The input variable of a model can also be linked to the output responses of other models. If the input of a model is a function of outputs of other models, you can add an input variable for the input of the dependent model and responses for the outputs of the independent models, and link the input variable to the output responses by entering the dependency in the expression field.

1. In the Define Input Variables step, click the **Links** tab.

2. In the Data Type column, select the expected return type of your data.
3. In the Expression column of the input variable you would like to link to another input variable or output response, click ....
4. In the **Expression Builder**, click the **Input Variables/Output Responses** tab.
   Click the **Input Variables** tab to view a list of input variables in your study; click the **Output Responses** tab to view a list of output responses in your study.
5. Select the input variable/output response to which the input variable will be linked.
6. Click **Insert Varname**.
7. Click **OK**.

Figure 30:



The input variable is now linked to the input variable/output responses selected in the **Expression Builder**. A link icon ( ) displays in the Varname column of the input variable to indicate that it is linked to an expression.

	Active	Label	Varname	Data Type	Expression
1	☑	t1	m_1_Variable_01	Real	...
2	☑	t2	m_1_Variable_02	Real	...
3	☑	t3	m_1_Variable_03	Real	...
4	☑	Mass_Excel	var_4	Real	m_1_r_1 ...
5	☑	Cost Coef	var_5	Real	...

Figure 31:

Create Input Variable Constraints

Define an inequality constraint condition that is a function of only the input variables.

Input variable constraints can be evaluated when performing any solver execution, therefore they can be treated in a special way to ensure that no evaluations are generated that violate the condition.

For example, you can fill a space that ensures that no design violates the condition, or an optimizer can avoid generating evaluation runs that would be known to generate a solver failure.

Input variable constraints should primarily be used to preemptively avoid a solver failure. If the solver can run, the optimizer will frequently perform better using normal optimization constraints.

1. In the Define Input Variables step, click the **Constraints** tab.
2. Click **Add Constraint**.
   A new constraint is created and added to the list of constraints.
3. In the Label field, enter a name for the constraint.
4. In the Left Expression and Right Expression fields, specify expressions to compare in the following ways:
   - Manually enter an expression.
   - Click ... to create an expression using the **Expression Builder**.
5. In the Comparison field, select a comparison type.
   - **>=** Greater than or equal to
   - **<=** Less than or equal to

	Active	Label	Varname	Left Expression	Comparison	Right Expression
1	☑	Constraint 1	con_1	m_1_varname_1 ...	>=	m_1_varname_2 ...

Figure 32:

Input Variable Properties

In the Define Input Variables step, various input variable properties can be modified from the Bounds, Modes, and Distributions tabs.

Bounds

Lower bound	Lower limit of the variable range to be studied.
Nominal	Default value of the variable if deactivated; also serves as the initial value in an Optimization.
Upper bound	Upper limit of the variable range to be studied.

Data Type

Real	Variable stored as a real valued floating point number, for example 1.0.

Integer	Variable stored as an integer, for example 1.0.
String	Variable stored as a character without any numeric meaning, for example one.

Mode

Continuous	Input variable that can take any value between the lower and upper bounds, for example 1 < x < 2.
Discrete	Input variable that can take values from an orderable finite list of numeric values, for example x = 0.1, 0.2, 0.3, or 0.4.
Categorical	Input variable that can take values from a non-orderable finite list of values, for example x = red, green, blue.

Distribution Role

Design	Variable that is deterministic and has no uncertainty associated with its value.
Random Parameter	Variable that is probabilistic, but is not controllable by design; for example wind speeds or temperature.
Design with Random	Variable that is probabilistic, but is controllable by design; for example thickness or radius.

> **Note:** In a Stochastic approach, setting Distribution Role to **Design with Random** will use truncation sampling, in which any samples outside the Lower and Upper bounds are ignored.

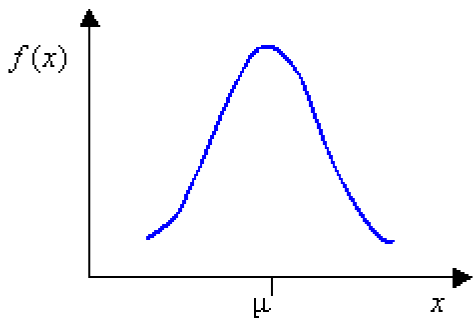
Distribution

Input variables can be characterized statistically using various statistical distributions. An input variable, when used in a statistical sense, is termed as a random variable. In ordinary usage, the term "random variable" indicates that the value this variable will take is unknown, but in a statistical sense, it is precisely known what values this variable will take and the probability associated with that value.

Input variables exhibit different properties depending on the parameter they represent. Some variables may be symmetric about the mean value, while others may be skewed towards either the left or right. Some variables may be bounded on either side or unbounded.

The first categorization of random variables is whether the variable is continuous or discrete. A random variable is considered continuous if it can assume any value in a given interval. A random variable is termed discrete if it can only assume a finite set of values within a given interval.

Normal (CoV or Variance)	Use to approximate many phenomenons in nature.



$$f(x) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and σ is the standard deviation.

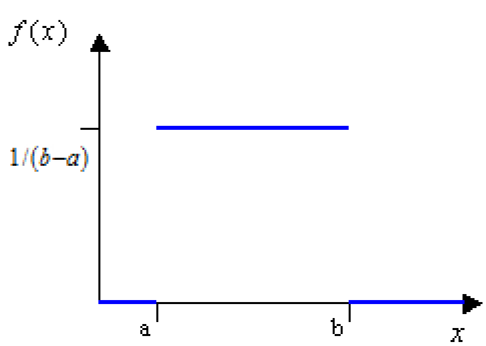
Figure 33:

In HyperStudy a normal distribution can be defined using mean, μ and variance, σ^2 or using mean, μ and coefficient of variance (CoV), σ/μ.

Variance is the second statistical moment and measures the spread of a distribution. CoV measures the relative spread of a distribution. The higher the CoV, the higher the variability.

Uniform

Use when all values between the minimum and maximum are equally likely, such as a number from a random number generator.



$$f(x) = \begin{cases} \frac{1}{b-a} & if\ a \le x \le b \\ 0 & otherwise \end{cases}$$

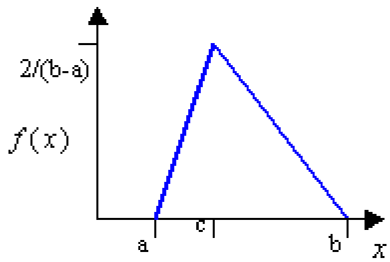
$$F(x) = \begin{cases} 0 & if\ x < a \\ \frac{x-1}{b-a} & if\ a \le x \le b \\ 1 & if\ x > b \end{cases}$$

where a and b are end points.

Figure 34:

Triangular

Use when the only known information is the minimum, the most likely, and the maximum values.



$$f(x)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)} & if\ a\leq x\leq c\\ \dfrac{2(b-x)}{(b-a)(b-c)} & if\ c\leq x\leq b\\ 0 & otherwise\end{cases}$$

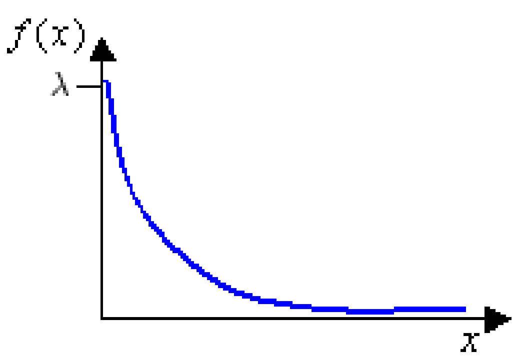
$$F(x)=\begin{cases}0 & if\ x<a\\ \dfrac{(x-a)^1}{(b-a)(c-a)} & if\ a\leq x\leq c\\ 1-\dfrac{(b-x)^1}{(b-a)(b-c)} & if\ c<x\leq b\\ 1 & if\ b<x\end{cases}$$

where a, b, and c are the end points and the mode.

Figure 35:

Exponential

Use to describe the amount of time between occurrences, mean time between failures.



$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\\ 0 & x<0.\end{cases}$$

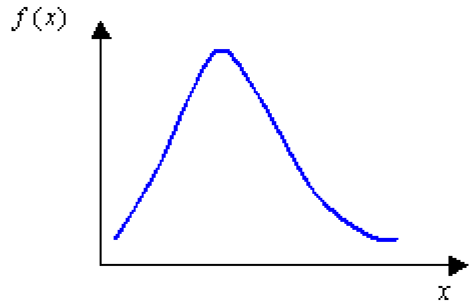
$$F(x)=\begin{cases}1-e^{-\lambda x} & x\geq 0,\\ 0 & x<0.\end{cases}$$

where λ is the scale parameter.

Figure 36:

Weibull

Principal applications are situations involving wear, fatigue and failure, failure rates, life-time expectancies.



$$f(x) = \begin{cases} \alpha\beta^{-\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & if\ x > 0 \\ 0 & otherwise \end{cases}$$

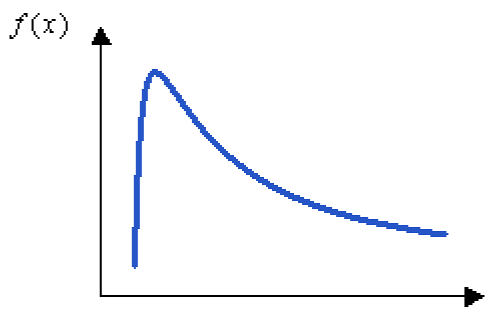
$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & if\ x > 0 \\ 0 & otherwise \end{cases}$$

where α and β are shape and scale parameters which enable it to be adjusted to desired fatigue or reliability laws.

Figure 37:

Log Normal

Use in risk analyses.



$$f(x) = \frac{1}{xs\sqrt{2\pi}} e^{\frac{-(\ln x - m)^2}{2s^2}}$$

where m and s are location and scale.

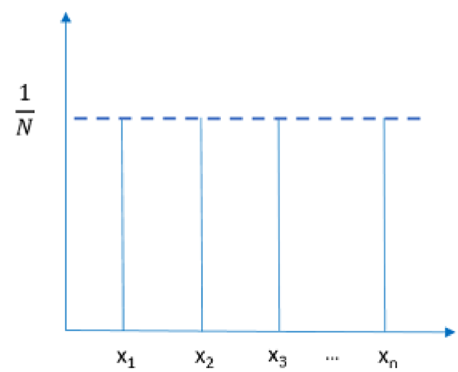
Figure 38:

Uniform Discrete

Use when you have discrete (numeric or string) variables that take values which are equally likely.

Possible numeric values are 1, 2, 3, or 4; each are equally likely.

Possible string variables are "orange", "green", "red", or "blue"; each are equally likely.



$$f(x) = \frac{1}{N} \forall\, x \in \{x_1, x_2, \ldots x_n\}$$

Figure 39:

Define a Categorical Input Variable and Set the Allowable Values

A categorical variable is an input variable that can take values from a non-orderable finite list of values, for example x = red, green, blue.

This example assumes that you already have an input variable created.

1. In the Define Input Variables step, click the **Modes** tab.
2. In the Data Type column for the input variable you are defining, select **String**.
3. In the Mode column, select **Categorical**.

> **Note:** The Values cell turns orange, which indicates you can edit its values.

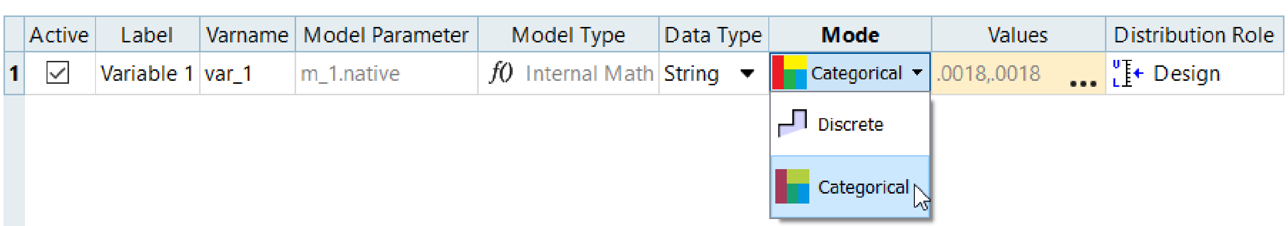


Figure 40:

4. In the Values cell, click ....

5. In the dialog, define the categorical values.

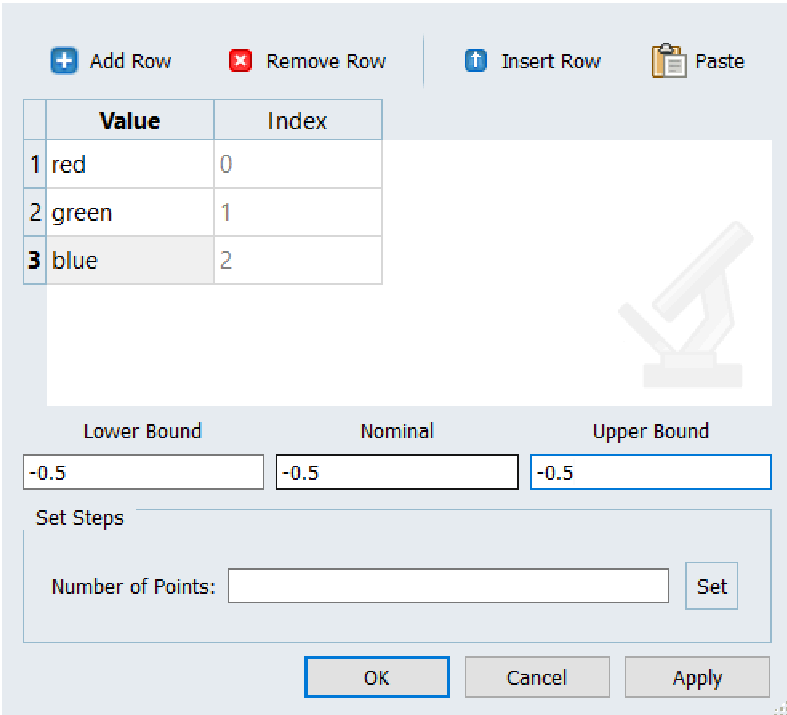


Figure 41:

6. Click **Apply**.
7. Click **OK**.

Define a Discrete Input Variable and Set the Number of Points

A discrete variable is an input variable that can take values from an orderable finite list of numeric values, for example x = 0.1, 0.2, 0.3, or 0.4.

This example assumes that you already have an input variable created.

1. In the Define Input Variables step, click the **Modes** tab.
2. In the Mode column for the input variable you are defining, select **Discrete**.

> **Note:** The Values cell turns orange, which indicates you can edit its values.

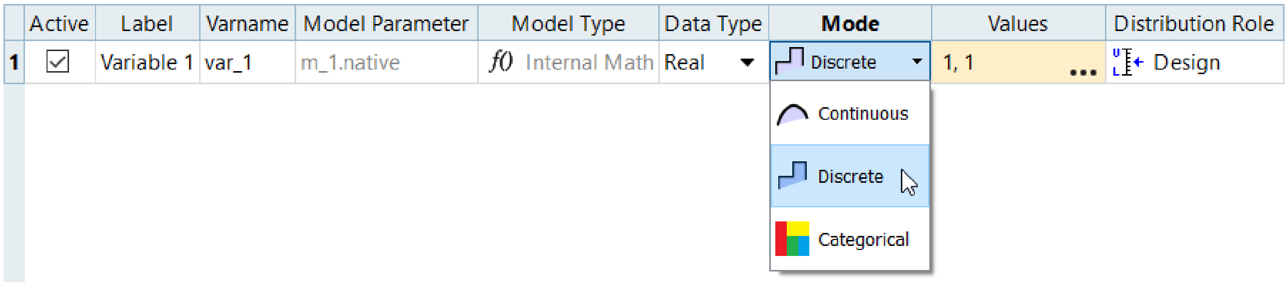


Figure 42:

3. In the Values cell, click [...].
4. In the dialog, edit the discrete value set and then click **OK**.
    - To add content to this set, click **Add Row**; to remove content from this set, click **Remove Row**.
    - To divide the value range into equal segments, change the lower and upper bounds, enter a value in the Step Number field, and then click **Set**.
    - To create a set with equal step sizes, change the lower and upper bounds, enter a value in the Step Size field, and then click **Set**.

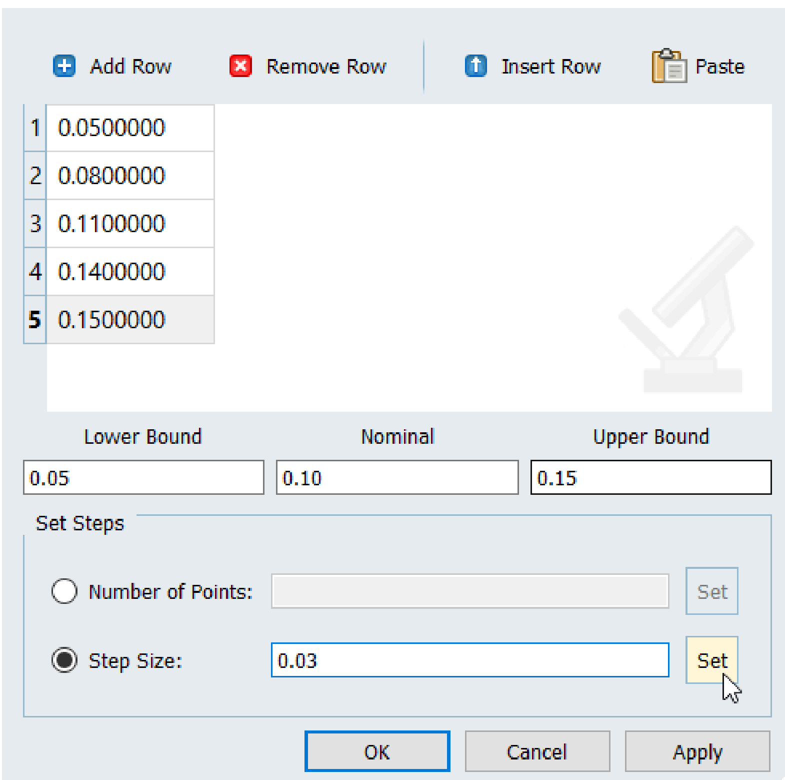


Figure 43:

Test Models

Write input files, execute analysis, and extract output responses for models defined in the study Setup.

1. Go to the **Setup** > **Definition** > **Test Models** step.
2. Define optional settings.
   - To receive a notification when a task is completed, click ≡ and activate **Notify**.
   - To write the solver output in the Message Log window and/or log-file, click ≡ and activate **Log External Output**.
   - To change the number of concurrent jobs to run, click **Multi-Execution** and enter a new number; it does not have to be a static entry. Click ≡ to choose whether Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.
     - Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
     - Horizontal execution mode sequences the write, execute, and extract task for each run independently.
     - Horizontal (Write all First) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

   Set Multi-Execution to 0 to stop the submission of new jobs.
3. Run definition tasks (write input files, execute analysis, and extract output responses) in the following ways:
   - Concurrently run tasks for all models in the definition by clicking **Run Definition**.
   - Concurrently run tasks for an individual model in the definition by clicking **All** in the Test column of the model to run.
   - Select an individual task to perform by clicking **Write**, **Execute**, or **Extract** in the Test column of the model to run.

HyperStudy creates simulation files in the `approaches/setup_1-def/run__00001/m_i` folders, where `m_i` stands for the `ith` model.

Review the run matrix by clicking the **Model Data** tab.

Define Output Responses

An output response is a measurement of system performances, such as weight, volume, displacement, stress, strain, reaction forces, and frequency.

Create and define the output responses that will be used in the studies DOE, Fit, Optimization, and Stochastic approaches.

Output responses may be defined using a combination of results read or extracted from a file and input variable values. These equations may use the functions available or external functions, including Compose/OML and Python functions. HyperStudy uses HyperWorks readers and hence can read the binary output files of most popular CAE solvers. Additional data can be read using custom external readers and import templates.

> **Tip:** For more information about the supported CAE solvers, reference . For more information about the supported file output files, reference .

Add and Remove Output Responses

Output responses can be created manually or automatically using the File Assistant.

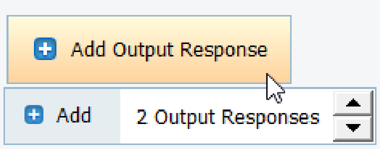
Create Output Responses Manually

Manually create and define output responses.

1. In the Define Output Responses step, Define Output Responses tab, click **Add Output Response**. A new output response is created and added to the list of output responses.
2. In the Label column, enter a name for the output response.
3. Define the output response expression.
   - Manually enter the expression in the Expression column.
   - Build an expression using the Expression Builder, which can be accessed by clicking ... in the Expression column.
4. In the Output Type field, select one of the following:
   - Choose **Real** (default) if the output response is a real number.
   - Choose **String** if the output response is a string.
5. Click **Evaluate** to extract the output response value.

	Active	Label	Varname	Expression	Value	Goals	Output Type
1	☑	Mass	r_1	mass_v[0] ...	2.6219100		Real
2	☑	Displacement at Node 19021	r_2	disp_v[0] ...	0.0026359		Real
3	☑	1st Frequency	r_3	freq_1[0] ...	295.23650		Real
4	☑	File Size	r_4	fileinfo("./m_1/beam.h3d", "size") ...	618211.00		Real

Figure 44:

> **Tip:** Create multiple output responses simultaneously by left-clicking and holding the mouse button on **Add Output Response**. In the pop-up, enter the number of output responses to add and press **Enter**.
>
> 
>
>
> *Figure 45:*

Create Output Responses using the File Assistant

Automatically create and define output responses using the File Assistant's guided set of dialogs to extract data from the output files generated during the evaluation.

1. Go to the Define Output Responses step, Define Output Responses tab.
2. Select the output file that contains the output response data to extract in the following ways:
   - From the Directory, drag-and-drop the output file into the work area. The response file selection is automatically populated in the File Assistant dialog, File field.

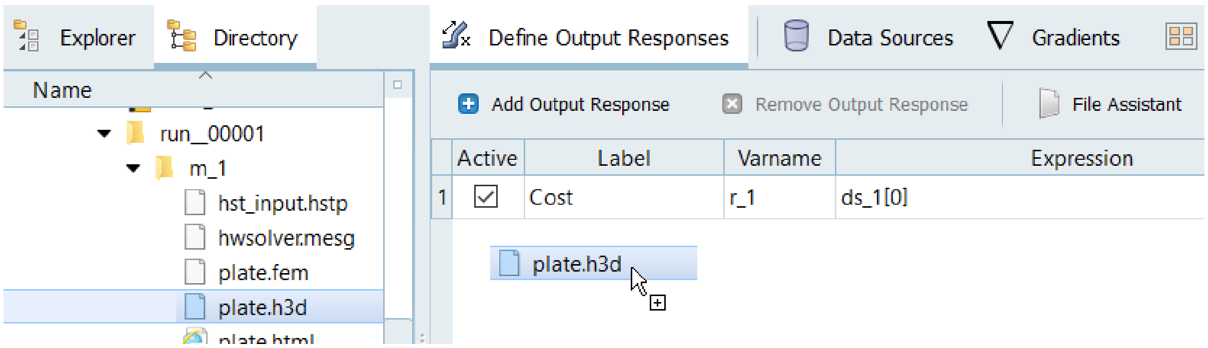


Figure 46:

   - Click **File Assistant**. In the **File Assistant** dialog, File field, navigate to the output file.

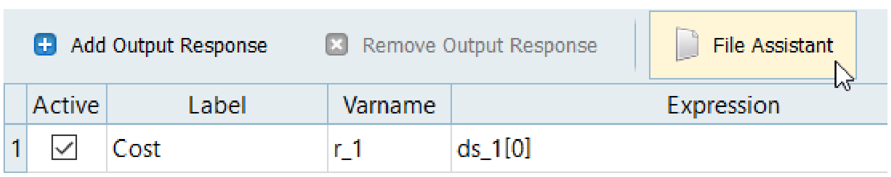


Figure 47:

3. Under Reading Technology, verify the correct reader is selected and click **Next**.
4. Select a data goal and click **Next**.

Option	**Description**
Single Item in a Time Series	Collect scalar information for a specific item, for example the displacement of a specific node.
Multiple Items at Multiple Time Steps	Construct output responses based on a range of items,for example the maximum stress in the entire model.
Mode Tracking	Compare modal result to a known reference to correlate results or avoid mode switching.

5. Define data options (subcase, type, mode, request, and so on) and click **Next**.

> **Note:** The options available will vary according to the data goal selected.

6. Define options for data source and output response creation and click **Finish**.
   a) Under Create a new Data Source, enter a label for the data source.
   b) Optional: To create a new response in which the data source will be used, select **Use Data Source in a new Response**. Enter a label for the output response, and select an expression type.

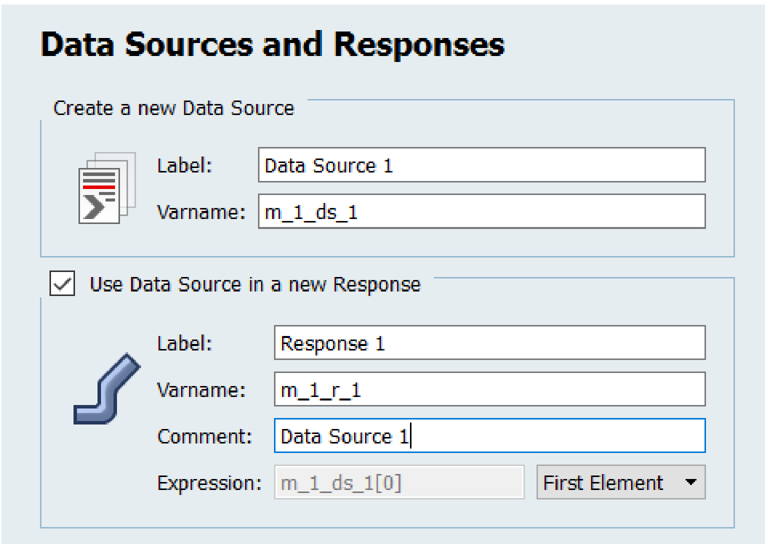


Figure 48:

A new output response is created and added to the list of output responses.

7. In the Output Type field, select one of the following:
   - Choose **Real** (default) if the output response is a real number.
   - Choose **String** if the output response is a string.
8. Click **Evaluate** to extract the output response value.

Remove Output Responses

Remove output responses from your study.

1. In the Define Output Responses step, Define Output Responses tab, select the output responses to remove.
2. Click **Remove Output Response**.

> **Warning:** If the output response is in use in a study approach, a warning message will prompt you for conformation.

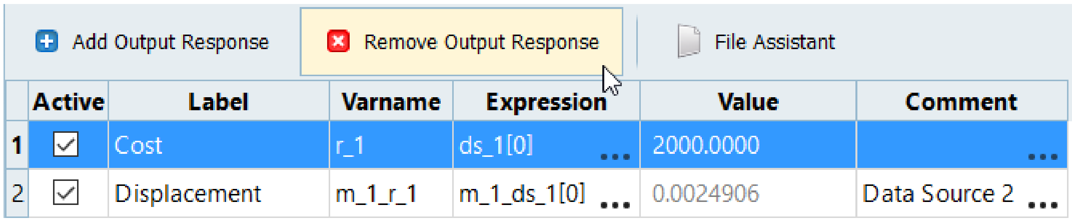


Figure 49:

Tip: Select multiple output responses to remove by holding Ctrl while left-clicking.

Build Output Response Expression using the Expression Builder

Build a mathematical expression that defines the output response using the Expression Builder.

In the **Expression Builder**, functions, input variables, output responses, and data sources are organized in the different tabs.

The Functions tab lists available functions and operators. External functions can be manually added in the preference file or registered using the Register with HyperGraph/HyperStudy option.

The Input Variables, Output Responses, and Data Sources tab lists all of the existing input variables, output responses, and data sources in the study.

1. In the Define Output Responses step, Define Output Responses tab, click ... in the Expression field of the output response to define.
2. In the **Expression Builder**, build the output response expression by appending functions, input variables, output responses, and data sources in the following ways:
   - Manually enter the varname assigned to each function, input variable, output response, or data source in the Evaluate Expression field.

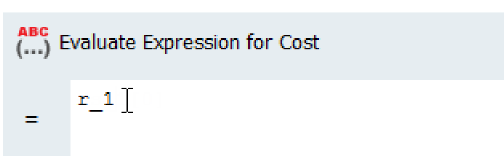
     

     Figure 50:

   - Select existing functions, input variables, output responses, or data sources from their corresponding tabs and click **Insert Varname** to append its corresponding varname in the Evaluate Expression field.

     Note: When inserting a data source, click # to select a common operation to map the data source to a scalar number (for example, maximum or minimum).

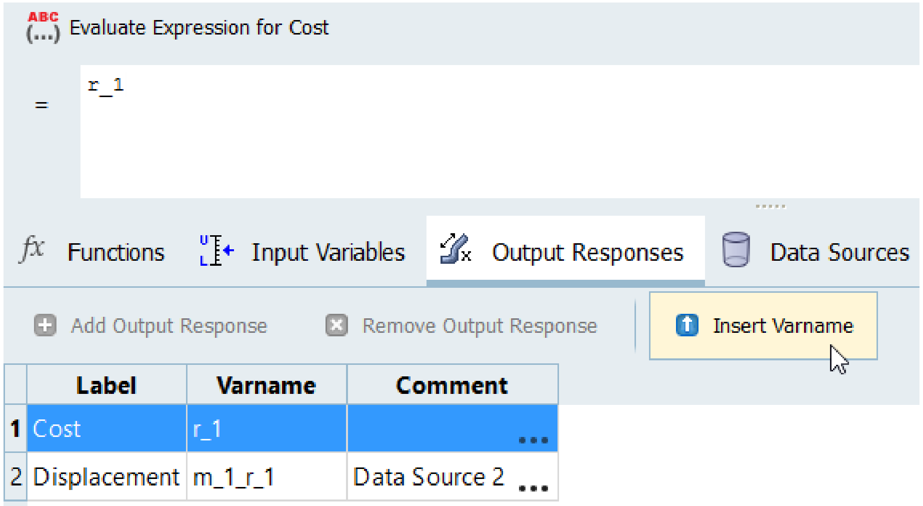


Figure 51:

3. Optional: Click **Evaluate Expression** to extract the output response value.

Define Data Sources

Data sources contain data obtained with various result readers technologies, such as file sources, ascii extract, and readsim.

Data sources are defined for the current study, not just for individual output responses. That is, Data Source 1 is accessible across all output response definitions.

1. In the Define Output Responses step, click the **Data Sources** tab.
2. Click **Add Data Source**.
   A new data source is created and added to the work area.
3. In the Label column, enter a name for the data source.
4. In the File column, click ....
   The **Data Source Builder** opens.
5. In the File field, navigate to the file containing result data.
6. From the Tools drop-down, select a read utility to be used to extract data from a file.

Option	**Description**
File Source	Extract data out of a run file by selecting a subcase, request, and component over all time steps.
Read Simulation	Extract data out of a run file by selecting a combination of multiple requests, components, and timesteps. Based on the readsim templex function.
Modal Assurance Criteria	Compare modal result to a known reference to correlate results or avoid mode switching. Based on the readmac templex function.

Option	Description
Area	Calculate the area between two curves.
Spreadsheet	Extract and store a range of cells in a spreadsheet (readexcelfile).
ASCII Extract	Extract custom data out of ASCII result files. Search for a response in the ASCII file using a Keyword . Highlight the number to define as a response, then right-click and select **Value**. If you select the Keyword option, the position of the value is determined relative to the keyword. This helps to deal with files that vary in length. If no keyword is used, the position of the value is determined with respect to the beginning of the file. You can also define keywords by right-clicking on a highlighted string and selecting **Keyword**.
Templex	Define templex statements or expressions, which can then be referenced in the output response expression.
Python	Execute code in a Python interpreter. The code should contain a return command with the data source.
Hstp Reader	Extract and store data from a HyperStudy parametric data definition file.

7. Define data options (subcase, type, mode, request, and so on).

   > **Note:** The options available will vary according to the Tool selected.

8. Click **OK**.

> **Tip:** If the data source is retained, the array of data is stored for future use. If the data source is not retained, the array of data is discarded after the extractions are complete. It is recommended to retain the information unless disk space is a concern. To retain a data source, select its corresponding checkbox in the Retain column.

Define Goals (Objectives/Constraints)

Goals are used to define objectives and constraints.

Goals can be defined in the following ways:

- Define goals for the active output response.
  a) Go to the **Define Output Responses** step.
  b) Click the **Define Output Responses** tab.
  c) In the Goals column, click .
  A dialog opens.

d) In the Type column, select the type of goal to create.
e) In columns 1 and 2, define additional options as needed.
f) Click **OK**.

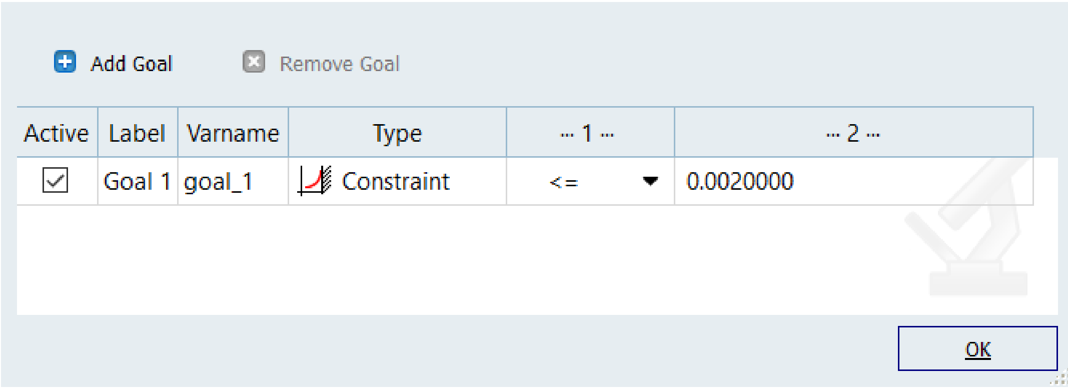
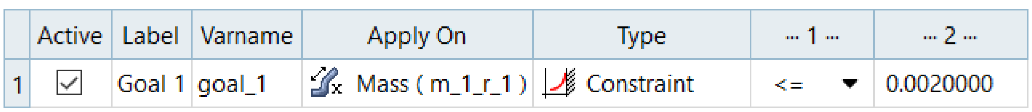


Figure 52:

- Define goals manually.
  a) Go to the **Define Output Responses** step.
  b) Click the **Objectives/Constraints - Goals** tab.
  c) Click **Add Goal**.
  d) In the Apply On column, select the output response to which this goal will be applied.
  e) In the Type column, select the type of goal to create.
  f) In columns 1 and 2, define additional options as needed.



	Active	Label	Varname	Apply On	Type	... 1 ...	... 2 ...
1	☑	Goal 1	goal_1	Mass (m_1_r_1)	Constraint	<=	0.0020000

Figure 53:

Objectives

Objectives are metrics to be minimized or maximized in an optimization exploration. Minimizing mass to find a lightweight design is a common example.

Single and Multi-Objective Optimization Problems

Optimization problems can be grouped as single or multi-objective optimization problems depending on how many objectives are considered. Multi-objective optimization (MOO) objective functions are formulated as:

$$\min f(x) = \{f_1(x), f_2(x), \dots, f_n(x)\}$$

such that $g\,j(x) < 0$

When dealing with multiple objectives ($f_1, f_2, \ldots$) it is unlikely that one design will have minimum objective function values for all objectives. As a result, in MOO applications an optimal Pareto front is searched for instead of an optimal design. Optimal Pareto front is a collection of non-dominated designs. A non-dominated design has a lower objective function value than others with respect to at least one objective.

Usually the computational effort required to solve MOO problems are significantly more compared to single objective optimization problems. In cases where solving MOO problems are prohibitive, these problems have been converted to a single objective problem by summing all the objectives (Weighted Sum Method).

When dealing with probabilistic variables, the objective function also has an associated distribution. When doing robust optimization, instead of the deterministic objective, the objective function is the value of the objective distribution at a specified value of the cumulative distribution function (CDF). A minimization problem might use the 95% value of the CDF (default value), and a maximization problem might use the 5% value of the CDF.

Objective Types

Types of objectives that can be defined when setting up an Optimization in HyperStudy.

Minimize and Maximize

When creating an objective that is of type Minimize or Maximize, you can edit the Weighted sum field to create a weighted sum of all objectives. If the Weighted sum field is edited, only Adaptive Response Surface Method, Method of Feasible Directions, Genetic Algorithm, Sequential Quadratic Programming methods are available in the Specifications step. If multiple Minimize and/or Maximize type objectives are created, but no weights are defined, only Multi - Objective Genetic Algorithm and Global Response Search Method are available.
This weighted objective function is always minimized, therefore it will be positive if the objective is to be minimized and it will be negative if the objective is to be maximized.

$$f x_1 = \sum w_i x_i$$

System Identification

Attempts to minimize the difference between the output response values and the target values of selected objectives. Examples for typical applications are experimental curve fitting or parameter fitting. The objective function is formulated as a least squares formula, where $\tilde{f}_i$ is the target value of the ith output response.

$$\min \sum \left(\frac{f_i - \tilde{f}_i}{\tilde{f}_i} \right)^2$$

If the target value is 0, the denominator will be set to 1.0 to avoid division by 0. In this case, the contribution to the sum is just the raw value itself, unnormalized.
When creating an objective of type System Identification, the Target value field must be edited.
System Identification cannot be used with Sequential Optimization and Reliability Assessment or ARSM-Based Sequential Optimization and Reliability Assessment methods.
System Identification objectives cannot be combined with other types of objectives. If one objective is assigned to this type, then all objectives have to be assigned to the same type.

MinMax and MaxMin

Used to solve problems where the maximum (or minimum) of an output response is minimized (or maximized).

The MinMax problem is formulated as:

$$\min\left\lfloor \max(f_1(x)/\tilde{f}_1, f_2(x)/\tilde{f}_2, \ldots, f_k(x)/\tilde{f}_k) \right\rfloor$$

Subject to:

$$g_j(x) \le 0 \qquad j = 1, \ldots, m$$
$$x_i^L \le x_i \le x_i^U \quad i = 1, \ldots, n$$

These problems are solved using the Beta-method. In this method the problem is transformed into a regular optimization problem by introducing an additional input variable such that:

$$\min\beta$$

Subject to:

$$f_\iota/\tilde{f}_\iota(x) \le \beta \quad \iota = 1, \ldots, k$$
$$g_j(x) \le 0 \qquad j = 1, \ldots, m$$

When creating an objective of type MinMax or MaxMin, the Reference Value field can be edited to normalize the respective function values. MinMax and MaxMin cannot be used with reliability based optimization methods.

MinMax and MaxMin objectives cannot be combined with other types of objectives. If one objective is assigned to this type, then all objectives have to be assigned to the same type.

Constraints

Constraints need to be satisfied for an optimization to be acceptable. Constraints may also be associated with a DOE. While not used in the evaluation of the DOE, constraints can be useful while visualizing DOE results. Limits on displacement or stress are common examples.

Constraint Categories

All constraints in an optimization problem can be placed into the following distinct categories:

Inequality Constraint

One sided condition that must be satisfied.

$$g_j(x) \le 0 \quad j = 1, \ldots, m$$

Equality Constraint

Precise condition that must be satisfied.

$$h_k(x) = 0 \quad k = 1, \ldots, m_h$$

Side Constraint

Bounds on the input variables that limit the region of search for the optimum.

$$x_i^L \leq x_i \leq x_i^U$$

Constraint Types

Constraints can be defined as type Deterministic or Random (probabilistic) when setting up an Optimization in HyperStudy, depending on the design requirements.

Deterministic

Deterministic constraints enable you to manually define a Bound Type, Bound Value, and evaluation source for the output response(s).

Random

Random problem formulations take into account the variability in the design and study the corresponding variability in the performances. This aspect is studied under reliability and robustness.

Random constraints require you to modify the CFD Limit if the reliability requirement is different than the default value of 99.00%. The CFD Limit is the reliability requirement on the constraint; that is the probability of (Output Response >= 0) > 99.00%.

Standard Constraint Enforcement

Constraints violations can be treated in the following ways:

Standard Enforcement

Constraints are considered feasible when they are within a small percentage of difference between their bounds. This type of enforcement is conventional.

Strict Enforcement

Constraints must be perfectly satisfied with no margin. This type of enforcement may require additional iterations from an optimizer for convergence.

Percent of Constraint Bound

Constraints must be violated by more than this value in the converged design. Strict enforcement only uses this tolerance for equality constraints.

When the Constraint Bound = 0.0

In general, constraint values are normalized to their bound value. One exception is if the absolute bound value is less than this parameter.

> **Tip:** The recommended range is 1.0e-6 ~ 1.0.

Define Gradients

Gradients calculate the change of the output responses with respect to the input variables. The gradient information can be used in gradient enabled methods inside of a DOE, Fit, Optimization, or Stochastic approaches.

Gradients must be defined for all combinations of output responses and input variables; some or all gradients can remain undefined. An undefined gradient is not equivalent to setting a defined gradient to be equal to 0.

Gradients may also be defined using a combination of results read or extracted from a file and input variable values. These equations may use the functions available or external functions, including Compose/OML and Python functions. HyperStudy uses HyperWorks readers, and hence can read the binary output files of most popular CAE solvers. Additional data can be read using custom external readers and import templates. External readers, external functions and import templates may be manually added in the preference file or registered through the Edit menu.

1. In the Define Output Responses step, click the **Gradients** tab.
2. Click **Add Gradient**.
   A new gradient is created and added to the work area.
3. In the Label column, enter a name for the gradient.
4. In the Derivative of column, enter the varname of the output response to be associated with this gradient.
5. In the Respect to column, enter the varname of the input variable to be associated with this gradient.
6. Define the output response expression.
   - Manually enter the expression in the Expression column.
   - Build an expression using the Expression Builder, which can be accessed by clicking ... in the Expression column.
7. Click **Evaluate** to extract the gradient value.

Tip: Automatically generate all possible gradient combinations by clicking **Add All**.

4.2 Setup the Approaches

Once the study Setup is complete, an unlimited combination of approaches can be added to a study. A study approach is a specific set of steps taken to study the mathematical model of a design.

Each approach in HyperStudy serves a different purpose in the design study. The required steps for each approach may be unique. For example, you can use the DOE approach if you need to learn the main factors affecting your design, but you need to use the optimization approach if you want to find the design that achieves the design objectives while satisfying design requirements.

4.2.1 Setup DOE Studies

A DOE is a series of tests in which purposeful changes are made to the input variables to investigate their effect upon the output responses and to get an understanding of the global behavior of a design problem. By running a DOE, you can determine which factors are most influential on an output response.

Add a DOE Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
    a) In the Label field, enter a name for the DOE.
    b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.

       By default, the Definition defined in the study Setup is selected.
    c) Under Select Type, select **DOE**.
    d) Click **OK**.

A new DOE is added to the Explorer.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.
3. Test Models.
4. Define Output Responses.
5. Review definitions in the following ways:

To:	Do this:
Review status	Review the status of a Definition to verify that each step is complete.

Review status

Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

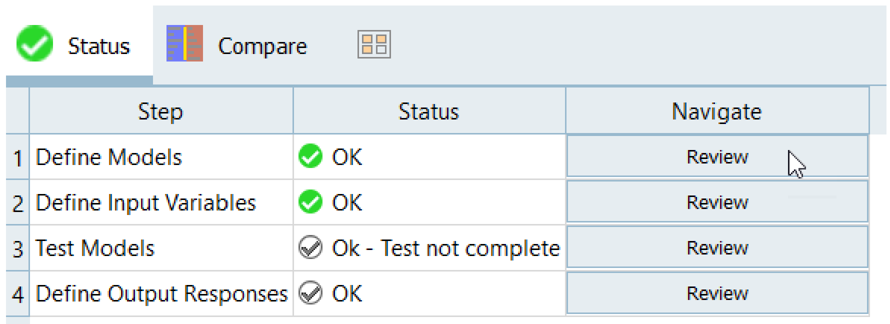


Figure 54:

Compare definitions

Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

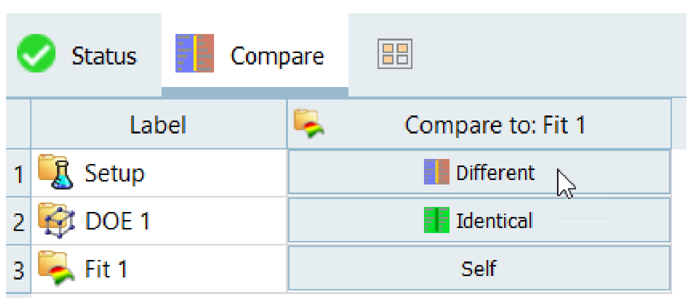
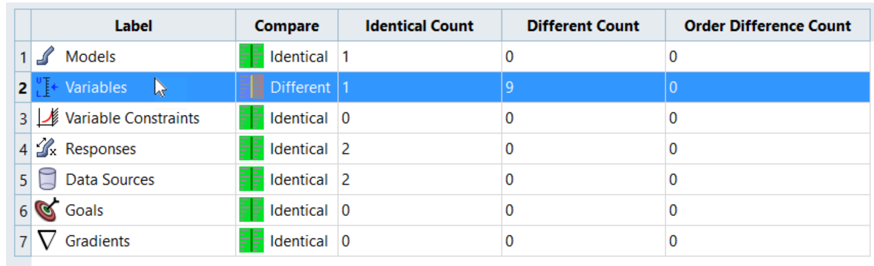


Figure 55:

   The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.
4. Click a channel to display a detailed comparison.

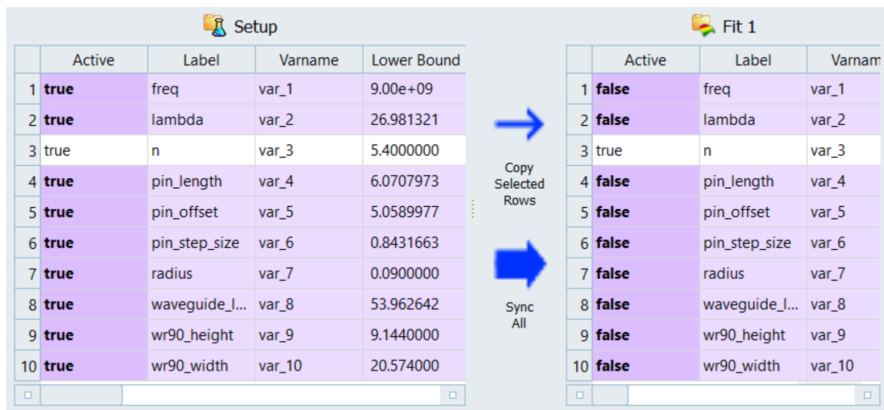
To:	**Do this:**



	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 56:

5. Sync data.
- Click **Copy Selected Rows** to sync the single row.
- Click **Sync All** to sync all rows.



Setup

	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

Fit 1

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 57:

Select a Numerical Method

Select a numerical method to use when evaluating the DOE.

1. In the Specifications step, Mode column, select a numerical method.
2. Optional: In the Settings tab, change settings as needed.
3. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog. For more information, see Edit the Run Matrix.

DOE Methods

Numerical methods available for a DOE approach.

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
Box Behnken	Space Filling	3	Click **Apply** for AutoSelect or select a table using the Design pull-down menu.	Use to build quadratic response surfaces if the responses are known to be quadratic and predictions are not required at the edge of the design space. Number of points can be 13, 25, 41, 49. 57. Selecting Autoselect will pick bbdgn13 if N < 4, where N is the number of design variables; bbdgn25 if N = 4, bbdgn41 if N = 5, etc. Limited to 7 design variables. Discrete variable must have at least 3 levels. Categorical variables must have exactly 3 levels.
Central Composite Design (CCD)	Space Filling	5		Use when the responses are known to be quadratic. Limited to 20 design variables.

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
D-Optimal	Space Filling	Any	You can either accept the default number of runs or enter a different value. You can also select the appropriate regression model.	Use when the known goal is to build a regression. This method is also useful when corner coverage is important, and you have problems with input variable constraints.
Fractional Factorial	Screening	Any	Select the appropriate resolution.	Resolution indicates the level of accuracy of the interactions. Interactions should not be used with Resolution III.
Full Factorial	Screening	Any		Requires a high number of simulations and is therefore unsuitable for most studies. Total number of runs should be less than 1,000,000.
Hammersley	Space Filling	Any	You can either accept the default number of runs or enter a different value.	Use when the response surface is highly nonlinear. This method is a better space filler than Latin HyperCube. The default number of runs is 1.1*((N+1)*(N+2))/2, where N

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
				is the number of design variables.
Latin HyperCube	Space Filling	Any	You can either accept the default number of runs or enter a different value.	Use when the response surface is highly nonlinear. The default number of runs is 1.1*((N+1)*(N+2))/2, where N is the number of design variables. You must maintain the value of the random seed in order to get repeatable designs.
Modified Extensible Lattice Sequence (Mels)	Space Filling	Any	You can either accept the default number of runs or enter a different value.	Use when the response surface is highly nonlinear. This method is a better space filler than Latin HyperCube. The default number of runs is 1.1*((N+1)*(N+2))/2, where N is the number of design variables.
Plackett Burman (PB)	Screening	Any		Computationally least expensive. Number of points can be 12, 20, 24, 28 or 36. Selecting Autoselect will pick pbdgn12 if N < 12, where N is the number of design variables; pbdgn20

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
				if 12 <= N < 20, etc. Limited to 35 design variables. Categorical variables must have exactly two levels.
Run Matrix	Custom	Any	Select the perturb file.	Use to create a design matrix using literal variable values.
Taguchi	Screening	Varies	You can either choose AutoSelect or a specific design matrix.	The levels of each variable must be set accordingly to ensure compatibility with a specific design matrix.
User Defined Design	Custom	Any	Select the perturb file.	Use to create a design matrix using abstract variable levels.

Box Behnken

Generates higher order response surfaces using fewer required runs than a normal factorial.

Box Behnken designs place points on the midpoints of the edges of the cubical design region, as well as points at the center.

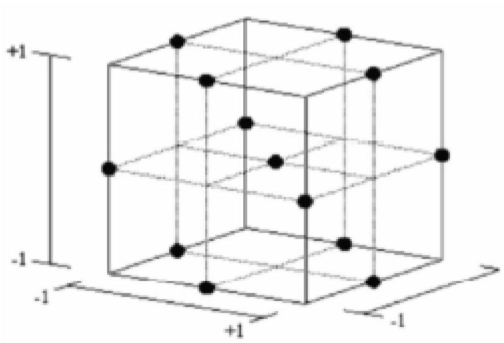


Figure 58:

Box Behnken designs and CCF Central Composite Design can be visualized as near compliments of each other. They both essentially suppress selected runs from a Full Factorial matrix in an attempt to maintain the higher order surface definition. For example, for three three-level variables, the Full Factorial run size is 27. The Central Composite Design drops all of the middle edge nodes, resulting in only 15 runs. The Box Behnken design is nearly the opposite in that it uses the twelve middle edge nodes and the center node to fit a 2nd order equation. A Central Composite Design plus a Box Behnken design becomes a Full Factorial with extra samples taken at the center.

Usability Characteristics

- Generally used for fitting a second-order response surface.
- Only defined when all of the variables have three levels.
- Should not be used when accurate predictions at the extremes are important.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Design	Auto Select	Auto Select bbdgn13 bbdgn25 bbdgn41 bbdgn49 bbdgn57	Select the lowest number of runs sufficient to study effects.

Parameter	Default	Range	Description
		Note: bbdgn stands for Box Behnken design.	
Number of Runs	Dependant upon the design selected.	13-57	Number of new designs to be evaluated.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Central Composite Design (CCD)

Central Composite Design contains an imbedded factorial or fractional factorial design with center points that are augmented with a group of `star points' that allow the estimation of curvature.

If the distance between the center of the design space and a factorial point is ±1 unit for each variable, then the distance between the center of the design space and a star point is $\pm\alpha$ with $|\alpha| > 1$. The precise value of α depends on certain properties desired for the design and on the number of factors involved. The star points represent new extreme values (low and high) for each variable in the design. Similarly, the number of center point runs that the design is to contain also depends on certain properties required for the design.

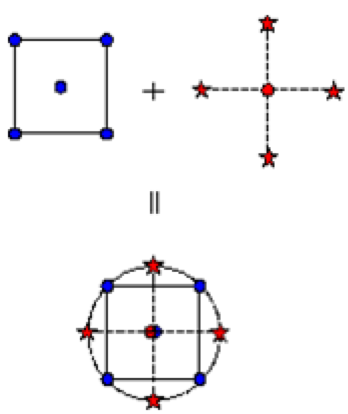


Figure 59: Generation of a Central Composite Design for Two Factors

Run #	X1	X2	X3
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1
9	-α	0	0
10	+α	0	0
11	0	-α	0
12	0	+α	0
13	0	0	-α
14	0	0	+α
15	0	0	0

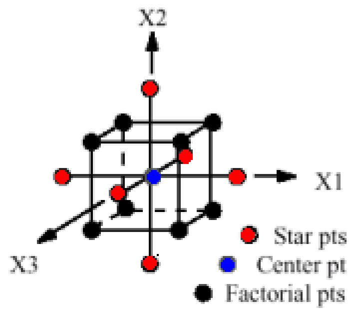


Figure 60: Generation of a Central Composite Design for Three Factors

Table 1: Values of α Defining the Type of Central Composite Design

CCD Type	Terminology	Definition
Circumscribed	CCC	Designs are the original form of the central composite design. The star points are at some distance α from the center, based on the properties desired for the design and the number of factors in the design. The star points establish new extremes for the low and high settings for all factors. These designs have circular, spherical, or hyperspherical symmetry and require five levels for each variable. Augmenting an existing factorial or resolution V fractional factorial design with star points can produce this design.
Inscribed	CCI	Uses the variable settings as the star points and creates a factorial or fractional factorial design within those limits (in other words, a CCI design is a scaled down CCC design with each variable level of the CCC design altered to generate the CCI design). This design also requires five levels of each variable. **Note:** Used for situations in which the limits specified for variable settings are truly limited.
Face Centered	CCF	The star points are at the center of each face of the factorial space, so α = ± 1. This variety

CCD Type	Terminology	Definition
		requires three levels of each variable. Augmenting an existing factorial or resolution V design with appropriate star points can also produce this design.

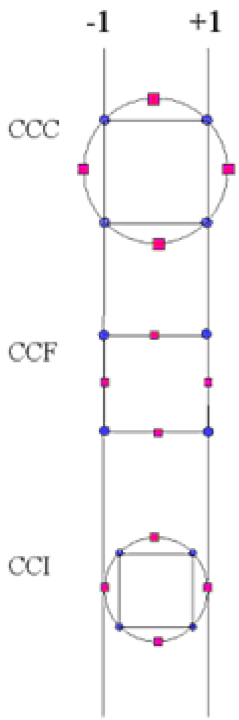


Figure 61: Comparison of the Different Types of Central Composite Designs

> **Note:** The CCC explores the largest process space and the CCI explores the smallest process space. Both the CCC and CCI are rotatable designs, but the CCF is not. In the CCC design, the design points describe a circle circumscribed about the factorial square. For three factors, the CCC design points describe a sphere around the factorial cube.

The Box Behnken design and the CCF Central Composite Design can be visualized as near compliments of each other. They both essentially suppress selected runs from a Full Factorial matrix in an attempt to maintain the higher order surface definition. For example, for three three-level variables, the Full Factorial run size is 27. The central composite plan drops all of the middle edge nodes, resulting in only fifteen runs. The Box Behnken design is nearly the opposite in that it uses the twelve middle edge nodes and the center node to fit a 2nd order equation. Central Composite Design plus Box Behnken becomes a Full Factorial with extra samples taken at the center.

Usability Characteristics

- Generally used for fitting a second-order response surface.

- In HyperStudy, the number of centre runs and axial distance, a, are parameters that you need to enter. HyperStudy also offers some preset values for a.

Preset Name	Axial Distance	No. of Centre Runs
Rotatable	2	User Defined
Orthogonal	1.41421	User Defined
Rotatable & Orthogonal	2	12
On Face	1	User Defined
User Defined	User Defined	User Defined

- The total number of runs is a function of the number of input variables and the number of center points as the Total runs = 2^k + 2k + N center points (k = input variables).
- When using Central Composite Design, HyperStudy has a limit of 20 input variables.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Axial Distance	0	0	Automatically calculated when ityp is 1 ~ 4; it can be modified when ityp = 5.
Inscribe	On	Off or On	Choose whether to force the points within the input variable bounds or not. **Off** Points do not need to fill in the bounds. **On** Points need to fill in the bounds.
Center Runs	1		Automatically determined when ityp = 3; you can modify it if ityp is not equal to 3.
Type	Rotatable	Rotatable Orthogonal Rotatable & Orthogonal On Face	Type of axial scaling.

Parameter	Default	Range	Description
		User Specified	
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

D-Optimal

Primarily intended to be used as the input matrix for a Least Squares Regression Fit. By identifying the type of regression that will be used, samples are selected to maximize the determinant of the information matrix that is inverted during the Fit's regression analysis, which in turn improves the numerical efficiency of the DOE.

This optimal determinant is the source of the name D-Optimal. Unlike factorial designs, which also have high information matrix determinants, D-Optimal DOEs can have an arbitrary number of runs. The evaluation points are selected from a candidate pool that uses an advanced combination of factorial and space filling designs.

Usability Characteristics

- The minimum number of allowable runs is equal to the number of unknown coefficients in the selected regression, which in turn is a function of the number of variables.
- The iterative search to improve the information matrix determinant can become computationally expensive. The search contains some internal logic to reduce excessive run times, but overall the time will scale with the number of variables, the number of terms in the regression, and the number of runs in the DOE.
- For linear regression, D-Optimal runs will tend to cluster near the corners and resemble a factorial design pattern. This is expected as edge points will have the largest impact on a linear function. This effect is less pronounced as the order of the regression is improved.
- Any data in the inclusion matrix is combined with the run data for post-processing. The optimal determinant is calculated using candidate and included points.
- Supports input variable constraints.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	Twice the number of runs required for a linear regression.	> 0 integer	Number of new designs to be evaluated.

Parameter	Default	Range	Description
Regression Model	Linear	Linear Squared Cubic Interaction Full Quadratic Full Cubic	The target regression structure used to build the information matrix. Maybe limited by problem size.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix. The D-Optimal solution considers the locations of the existing inclusion runs when generating its data.

Fractional Factorial

A factorial experiment in which only a chosen fraction of the combinations required for the Full Factorial DOE is run.

Fractional Factorial designs are used to reduce the number of runs required to extract pertinent information about the main effects and two-factor interactions.

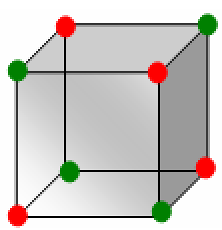


Figure 62:
The green dots illustrate the combinations a Fractional Factorial design considers.

This reduction in computational effort comes at the cost of an inability to completely resolve all of the main effects and interactions. Higher order interactions are often confounded with each other and, in some cases, can be confounded with the two-factor interactions. For example, for factors A, B, and C, each at two levels, only four runs would be required to resolve the main effects. However, all of the two-factor interactions would be confounded with the main effects (A:BC, B:AC, C:AB).

When applicable, the orthogonal arrays from other schemes, such as Plackett Burman or Taguchi, are used internally to reduce run count.

The amount of confounding in a Fractional Factorial design is classified by its resolution.

Resolution III	The effects are resolved with respect to each other, but all of the effects are confounded with the two-factor interactions.

Resolution IV	The effects are resolved with respect to each other and the two-factor interactions, but the two-factor interactions are confounded with respect to each other.
Resolution V	The effects are resolved with respect to each other and the two-factor interactions, and the two-factor interactions are resolved with respect to each other.

Usability Characteristics

- The desired resolution will determine the number of runs in the DOE.
- Resolution type III should only be used when on applications in which the interactions are known to be small with respect to the effects. This makes the confounding unimportant.
- When all variables have only two levels and the resolution is type III, avoiding the confounding between the effects and specified two-factor interactions can be achieved by using the Interactions tab. An enabled checkbox indicates the main effects are free from confounding with specified interaction.
- The techniques used to generate the run matrix work most effectively when variables have the same number of levels.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.
- Having consistent levels across the input variables makes Fractional Factorial more efficient.
- When the number of levels is less than the defined number of states of a discrete or categorical variable, the assigned levels are based on an equally spaced sample of the ordinal indices.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Resolution	III	III, IV, V	Select the resolution.
Number of Runs	Dependent upon the selected resolution	Any positive integer	Number of new designs to be evaluated.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Full Factorial

Evaluates all possible combinations of input variable levels. This will resolve all the effects and interactions.

Table 2:

Full Factorial run matrix for a three variable problem (variables A and B have two levels and variable C has three levels).

Run Number	A	B	C
1	1	1	1
2	1	1	2
3	1	1	3
4	1	2	1
5	1	2	2
6	1	2	3
7	2	1	1
8	2	1	2
9	2	1	3
10	2	2	1
11	2	2	2
12	2	2	3

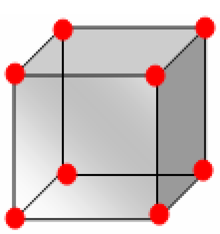


Figure 63:

Usability Characteristics

- For a case with k input variables, each at L levels, a Full Factorial design has L^k runs. For studies with a large number of input variables and levels, the total number of runs is larger. For example, for a study with eight factors and each with three levels, 6561 runs are needed (3^8 = 6561).

- This method may be practical for studies where there is a small number of variables and each variable has two levels, such as yes or no; -1 or 1. This method is not practical for most CAE applications where there are many factors possibly at several levels, and the simulations are costly.
- If the number of levels is not equal across variables, then the total number of runs is calculated by the product of the L^k terms. For example, consider eight variables: five variables with two levels, two variables with three levels and one variable with four levels. The number of full factorial runs is 1152 = 2^5 * 3^2 * 4^1.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.
- When the number of levels is less than the defined number of states of a discrete or categorical variable, the assigned levels are based on an equally spaced sample of the ordinal indices.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	$2^{ndv_2}3^{ndv_3}\ldots$	2-1,000,000	Number of new designs to be evaluated. ndv_i is the number of input variables with i levels. This number is determined automatically based on the number of input variables and levels.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Hammersley

Hammersley sampling belongs to the category of quasi-Monte Carlo methods. This technique uses a quasi-random number generator, based on the Hammersley points, to uniformly sample a unit hypercube.

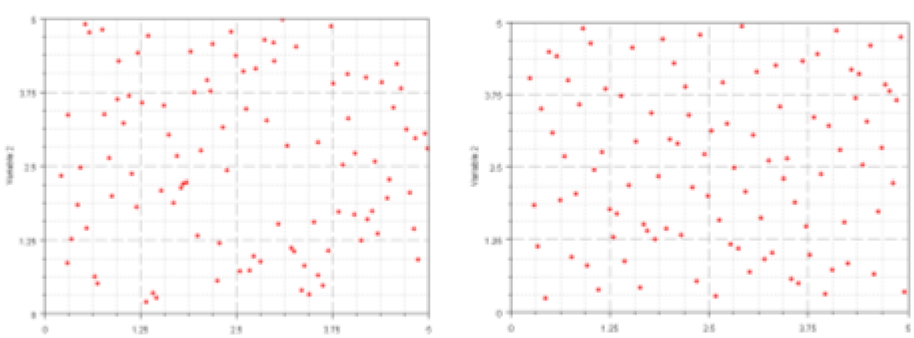


Figure 64:

Latin HyperCube (left) and Hammersley (right) for 100 runs.

Usability Characteristics

- An efficient sampling technique that provides reliable estimates of output descriptive statistics using fewer samples than random sampling. For example, for the same number of runs, a Hammersley sample will be closer to the theoretical mean than a truly random sample.
- Provides good, uniform properties on a k-dimensional hypercube. This is an advantage over Latin HyperCube sampling, which provides good uniform properties of each dimension individually.
- To get a good quality fitting function, a minimum number of runs should be evaluated. (N+1)(N+2)/2 runs are needed to fit a second order polynomial, assuming that most output responses are close to a second order polynomial within the commonly used input variable ranges of -+10%. An additional number of runs equal to 10% is recommended to provide redundancy, which results in more reliable post-processing. As a result, this equation is recommend to calculate the number of runs needed or a minimum of 1.1*(N+1)(N+2)/2 runs.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	$\frac{1.1(N+1)(N+2)}{2}$	> 0 integer	Number of new designs to be evaluated.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Latin HyperCube

A square grid containing sample positions is a Latin square if, and only if, there is only one sample in each row and each column. A Latin HyperCube DOE, categorized as a space filling DOE, is the generalization of this concept to an arbitrary number of dimensions.

When sampling a design space of N variables, the range of each variable is divided into M equally probable intervals. M sample points are then placed to satisfy the Latin HyperCube requirements. As a result, all experiments have unique levels for each input variable and the number of sample points, M, is not a function of the number of input variables.

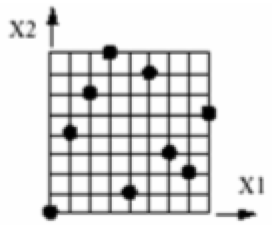


Figure 65:

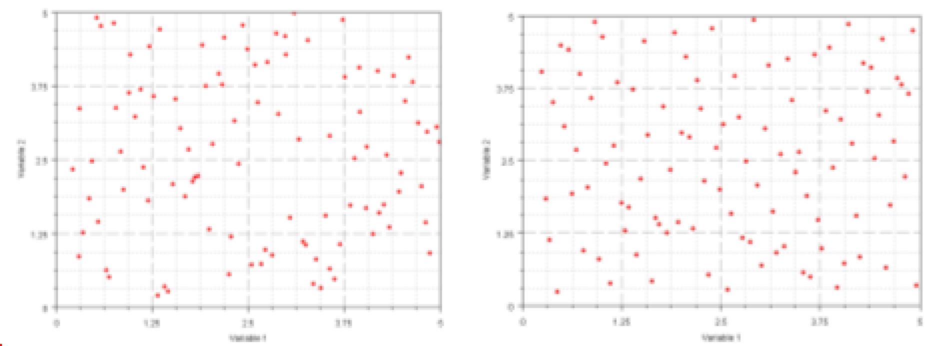


Figure 66:

Latin HyperCube (left) and Hammersley (right) for 100 runs.

Usability Characteristics

- To get a good quality fitting function, a minimum number of runs should be evaluated. (N+1)(N+2)/2 runs are needed to fit a second order polynomial, assuming that most output responses are close to a second order polynomial within the commonly used input variable ranges of -+10%. An additional number of runs equal to 10% is recommended to provide redundancy, which results in more reliable post-processing. As a result, this equation is recommend to calculate the number of runs needed or a minimum of 1.1*(N+1)(N+2)/2 runs.
- The structure of a Latin HyperCube run matrix ensures that the runs are orthogonal. Orthogonality is desirable because it is less likely to result in singularities when creating Least Squares Regression fits.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	$\frac{1.1(N+1)(N+2)}{2}$	> 0 integer	Number of new designs to be evaluated.
Random Seed	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated.

Parameter	Default	Range	Description
			0 Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Modified Extensible Lattice Sequence (Mels)

A lattice sequence is a quasi-random sequence, or low discrepancy sequence, designed to equally spread out points in a space by minimizing clumps and empty spaces.

This property makes lattice sequences an excellent space filling DOE scheme. This DOE type also has the property of extensibility, which means the method can take an existing set of data in a space, and add more data points to provide equal coverage; although with Modified Extensible Lattice Sequence it is optimal to extend on to Modified Extensible Lattice Sequence data. The number of runs is specified by the user.

Usability Characteristics

- Use for exploring the entire design space and creating fitting functions to the exact output responses. It is the recommended default space filling scheme.
- To get a good quality fitting function, a minimum number of runs should be evaluated. (N+1)(N+2)/2 runs are needed to fit a second order polynomial, assuming that most output responses are close to a second order polynomial within the commonly used input variable ranges of -+10%. An additional number of runs equal to 10% is recommended to provide redundancy, which results in more reliable post-processing. As a result, this equation is recommend to calculate the number of runs needed or a minimum of 1.1*(N+1)(N+2)/2 runs.
- Add existing data to the inclusion matrix to use the extensibility feature. While any data can be used as an inclusion, the best performance can be expected when the inclusion is an existing data set from a Modified Extensible Lattice Sequence DOE.
- Supports input variable constraints.
- When building a Modified Extensible Lattice Sequence DOE with the intention of using it as a Testing matrix, the resulting Testing matrix could be a subset of the Modified Extensible Lattice Sequence based input matrix due to the extensible property of Modified Extensible Lattice Sequence. To prevent this from happening, change the Random Seed setting of the Testing matrix to be a number larger than the number of runs in the Input data before building it.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	$\frac{1.1(N+1)(N+2)}{2}$	> 0 integer	Number of new designs to be evaluated.
Sequence Offset	1	Integer 0 to 10000	Controls the starting offset for the Modified Extensible Lattice Sequence sequence. For example, a value of 101 starts the generated evaluation points from the 101st point of the Modified Extensible Lattice Sequence sequence. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Use Inclusion Matrix	Off	Off or On	The use of an inclusion matrix will trigger the DOE to be extensible as it tries to fill in the space already covered by the existing data set.

Plackett Burman (PB)

Screens the maximum number of main effects with the least number of experimental runs in case of two-level factors.

Plackett Burman designs are economical designs, and are efficient in screening when only main effects are of interest. This is because the main effects in a Plackett Burman design are, in general, heavily confounded with two-factor interactions. The Plackett Burman design in 12 runs, for example, may be used for an experiment containing up to 11 factors.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	-1	+1	-1	+1	+1	+1	-1	-1	-1	+1	-1
3	-1	-1	+1	-1	+1	+1	+1	-1	-1	-1	+1
4	+1	-1	-1	+1	-1	+1	+1	+1	-1	-1	-1

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
5	-1	+1	-1	-1	+1	-1	+1	+1	+1	-1	-1
6	-1	-1	+1	-1	-1	+1	-1	+1	+1	+1	-1
7	-1	-1	-1	+1	-1	-1	+1	-1	+1	+1	+1
8	+1	-1	-1	-1	+1	-1	-1	+1	-1	+1	+1
9	+1	+1	-1	-1	-1	+1	-1	-1	+1	-1	+1
10	+1	+1	+1	-1	-1	-1	+1	-1	-1	+1	-1
11	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1	+1
12	+1	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1

Usability Characteristics

- Only 2-level variables can be used.
- The Number of runs is a factor of four.
- The maximum number of variables you can use is 35.
- Best suited for problems where the interactions are expected to be minimal and the main effects dominate.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Design	Auto Select	Auto Select pbdgn8 pbdgn12 pbdgn16 pbdgn20 pbdgn24 pbdgn28 pbdgn32 pbdgn36	Select the lowest number of runs sufficient to study effects.

Parameter	Default	Range	Description
		Note: pbdgn stands for Plackett Burman design.	
Resolution	III	III	Select the resolution.
Number of Runs	Dependant upon design selected.	8-36	Number of new designs to be evaluated.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Run Matrix

Load your own design matrix.

The run matrix is read by HyperStudy and used like any other design. Spaces, tabs, or commas can delimit the individual elements of the matrix. The rows define the different runs and the columns define the input variable values.

Table 3: Example: Run Matrix

1.0	2.0	3.0	4.0
4.1	4.3	4.5	4.6
6.7	8.1	10.0	11.0
17.2	1.0	1.0	3.0
.02	0.4	0.5	1.7
3.4	2.1	7.3	9.1

Usability Characteristics

- Create your own design based on individual requirements.
- The matrix does not have to fit any DOE type requirements. You can use it to automate a parameter study.

- The run matrix uses exact values of the variable. This is in contrast to the user defined DOE, which contains integers to represent the corresponding level of the variable.
- It is not necessary to utilize all designs in a study. Designs that are not desired can be turned off from the Write/Execute runs panel.
- Imported values are mapped to the independent variables.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Matrix File			File that has the dataset for your own design of matrix.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Taguchi

Explores how controllable variables can be used to mitigate the effects from the uncontrolled variables.

This DOE array is particularly popular in the field of robust design. A fundamental concept of this methodology is the classification of variables into two distinct groups: independent controlled variables and uncontrolled variables (sometimes called noise).

Consider the simple example of how the selection of a particular alloy for manufacturing could reduce the frequency of manufacturing outliers due to thermal effects (for example, a particular alloy is less sensitive instead of noise).

Usability Characteristics

- Taguchi arrays have a resolution type III. However, unless there is intent to use Taguchi arrays specifically, in most cases it is recommended that you use the Fractional Factorial DOE with resolution III, which can result in Taguchi arrays in some conditions.
- The effects between the controlled variables are confounded with respect to their two-factor interactions so the controlled variables should be selected to have no interactions between themselves. In this condition the calculation of the controlled variables are valid.
- The interaction between the controlled and uncontrolled variables are valid and often are the main result of interest.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Design	Auto Select	Auto Select L4 L8 L9 L12 L16a L16b L18 L25 L27 L32a L32b L36a L36b L50 L54 L64a L64b L81	Choose the array for the variable set.
Resolution	III	III	Select the resolution.
Number of Runs	$\frac{1.1(N+1)(N+2)}{2}$	> 0 integer	Number of new designs to be evaluated.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

User Defined Design

Load your own design matrix.

The User Defined Design is read by HyperStudy and used like any other design. You must specify the number of runs (rows) and the number of columns in the specified matrix in the first row of the supplied file. Spaces, tabs, or commas can delimit the individual elements of the matrix. Each entry in a row is an integer to represent the assigned level for the corresponding variable: 1 is the first level, 2 is the second, 3 is the third, and so on.

Table 4: Example: User Defined Design

9	3	
1	1	1
1	2	2
1	3	3
2	1	3
2	2	2
2	3	1
3	1	2
3	2	3
3	3	1

Usability Characteristics

- Create your own design based on individual requirements.
- The User Defined matrix uses integers to represent the corresponding level of the variable. This is in contrast to the run matrix DOE, which contains exact values of the variables.
- The number of levels specified in the file must be consistent with the number of variable levels specified in the HyperStudy user interface.
- Imported values are mapped to the independent variables.
- Any data in the inclusion matrix is combined with the run data for post-processing. Any run matrix point which is already part of the inclusion data will not be rerun.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Perturb File			File that has the dataset for your own design of experiments.
Use Inclusion Matrix	Off	Off or On	Concatenation without duplication between the inclusion and the generated run matrix.

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

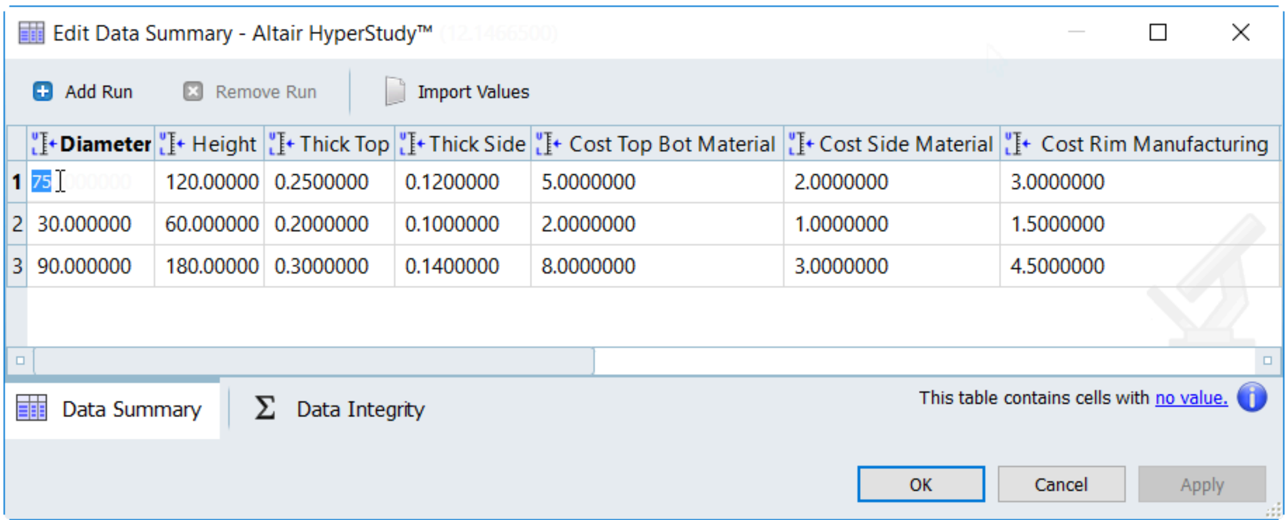


Figure 67:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.
   - Manually enter run data.
   - Copy and paste run data into the run matrix.

     Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

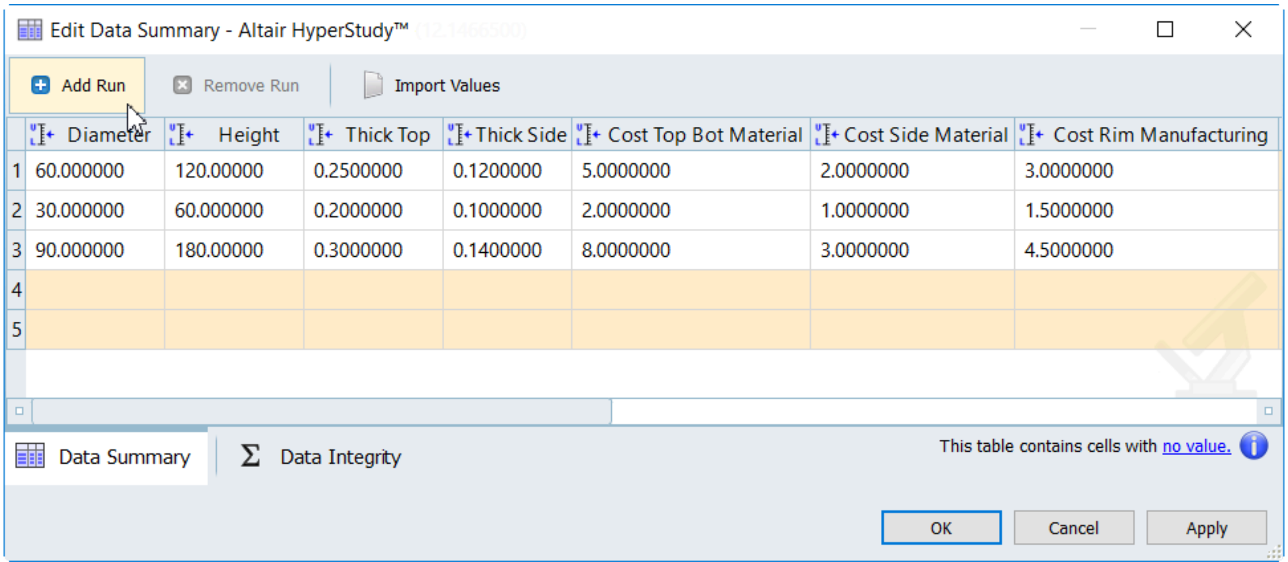
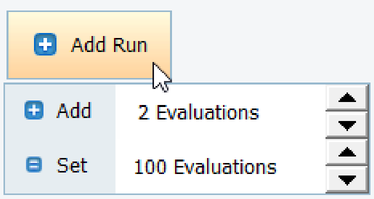


Figure 68:

> **Tip:** Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.
>
> 
>
>
> *Figure 69:*

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.
5. Select the source that contains run data.
   - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
   - For Approach evaluation data, select the approach that contains run data.

- For HyperStudy post processing file, select the source file.

6. Click **Next**.
7. Define the variable to column assignment(s).
   a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
   b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
   c) Click **Assign**.
8. Click **Finish**.

Reuse Run Data

An Inclusion matrix contains existing data that will be appended into the newly created approach as known data points. This data typically comes from other approaches, such as DOEs or previously run Optimizations.

In a DOE, an inclusion can be used to extend an existing DOE or concatenate data sets.

1. Go to the **Specifications** step for the DOE.
2. In the top, right of the work area, click **Edit Matrix** > **Inclusion Matrix**.
3. In the **Edit Inclusion Matrix** dialog, click **Import Values**.
4. In the **Import Values** dialog, select **Approach evaluation data** and click **Next**.
5. For Approach evaluation data, select the approach that contains run data.
6. Click **Next**.
7. Define the variable to column assignment(s).
   a) From the Variable to ColumnAssignment table, select a variable to which run data will be assigned.
   b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
   c) Click **Assign**.
8. Click **Finish**.
9. Review the imported run data.
10. Click **OK**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.

2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.
3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

   By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

	Active	Task	Batch
1	☐	Create Design	☐
2	☑	Write Input Files	☐
3	☑	Execute Analysis	☐
4	☑	Extract Output Responses	☐
5	☐	Purge ...	☐
6	☐	Create Reports	☐

Figure 70:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to stop the submission of new jobs. Click ≡ to select an execution mode.

   Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

   Each evaluation is independent so multi-execute can be used to run in parallel.

   Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

   - Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
   - Horizontal execution mode sequences the write, execute, and extract task for each run independently.
   - Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.
5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

DOE Output Files

Output files generated from a DOE.

<doe_variable_name>.hstds

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<doe_variable_name>/<doe_variable_name>.hstds`

File Contents

Result	Format	Description
Run Matrix Data	hstds, binary	Hstds files stores the retained data sources; direct access data using the `.hstds` file is not suggested.

<doe_variable_name>.hstdf

File Creation

This file is created when **Apply** is selected during the Specifications step.

File Location

`<study_directory>/approaches/<doe_variable_name>/<doe_variable_name>.hstdf`

File Contents

Result	Format	Description
Run Matrix Data	hstdf, binary	Hstdf files store the run data; however, direct access to the data using the hstdf files are not suggested.

Evaluation Parameters

Modify the run environment settings for the Evaluation tasks.

1. From the Evaluation step, click the **Evaluation Parameters** tab.
2. In the Value column, modify settings accordingly.

> **Note:** Review the Effectuation column to determine the scope at which each setting takes effect.

Review Evaluation Results

Review the input variable and output response values for each run, as well as review the run files.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Evaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.
4. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.

	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Channel

Figure 71:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

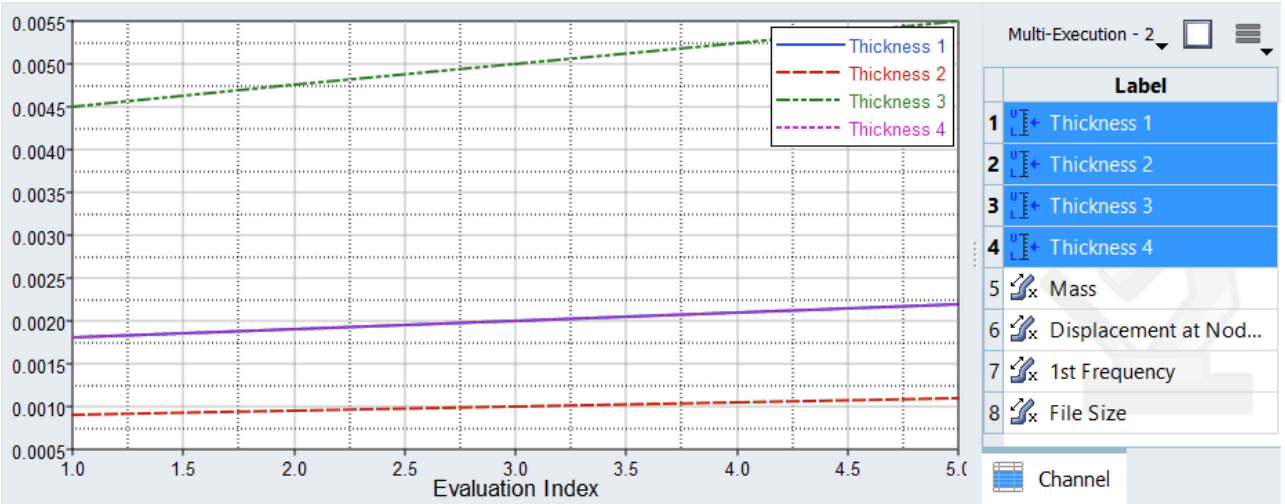


Figure 72:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Evaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. Select data to display in the scatter plot.
   - Use the Channel selector to select two dimensions of data to plot.

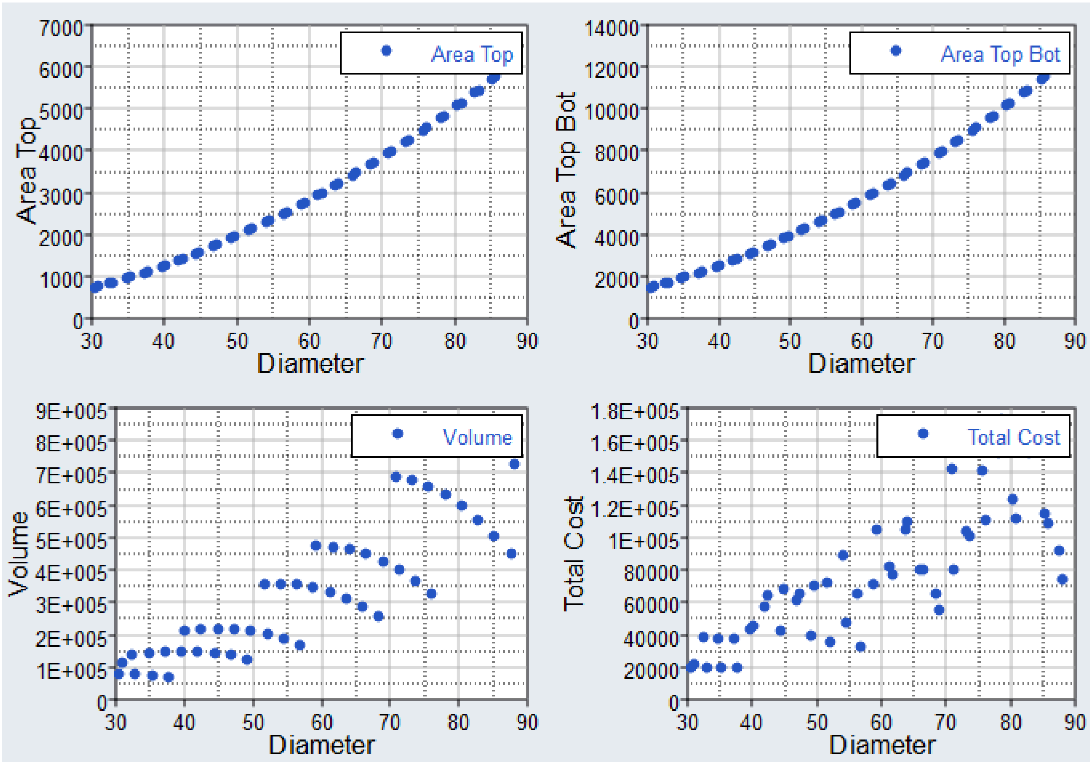


Figure 73:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

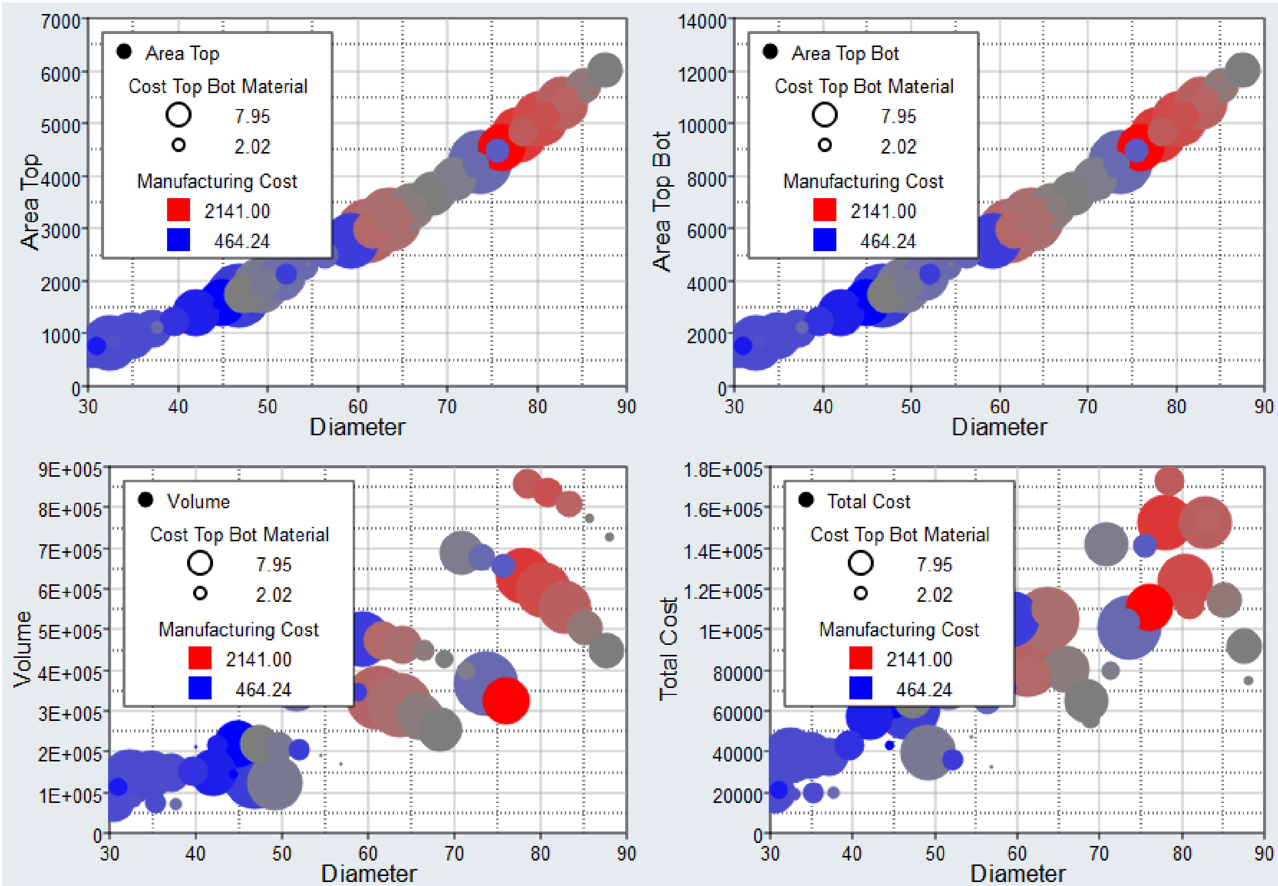


Figure 74:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between ☐ (single plot) and ▦ (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation Scatter Tab Settings

Settings to configure the plots displayed in the Evaluation Scatter tab.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.

Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Review Evaluation Time

Inspect task wall-clock times.

Review the time spent in each task within the Evaluation Time tab. Identify bottlenecks in tabular or plot form.

1. From the Evaluate step, click the **Evaluation Time** tab.
2. Use the top channel selector to select the model(s) to review.
3. Use the bottom channel selector to identify the time categorises for review.

Option	Action
Write	Time spent in the write task.
Execute	Time spent in the execute task.
Extract	Time spent in the extract task.
Model Total	Total time of the write, execute, and extract tasks.
All Models Total	Summation of all Model Totals.

Option	Action
	Note: This category is independent of the selected models.

4. Switch the view between table and plot by clicking Table or Plot, located above the Channel selector.

Evaluation Time Settings

Settings to configure the plots and tables displayed in the Evaluation Time tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Cumulative Rows
: Each row entry is a summation of its own wall time and the wall times preceding it with a lower evaluation index.

Plot Time-Unit
: Controls the units of time when plotting the wall times.

Post Processing

View the computational results from the DOE.

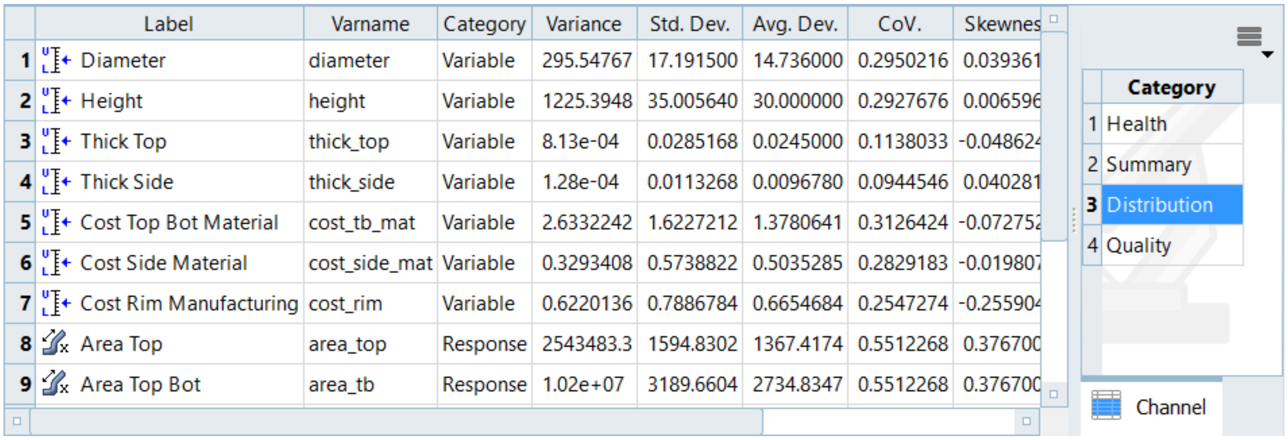
Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
   - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
   - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.
   - **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
   - **Quality** Values typically used in Quality Engineering.



	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.072752
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.019807
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.255904
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

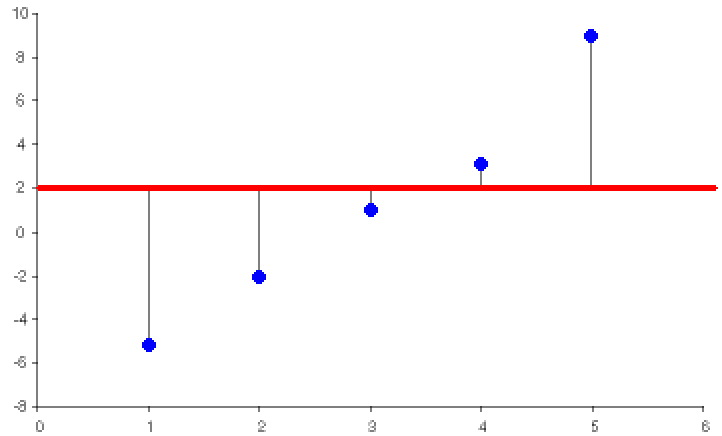
	Category
1	Health
2	Summary
3	Distribution
4	Quality

Figure 75:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	**Description**
Avg Dev (Average Deviation)	Average deviation is evaluated using: $$\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$$ In Figure 76, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 76:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $CoV = \frac{Standard\ Deviation}{Mean}$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $\sqrt{\frac{\sum x_i^2}{N}}$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}$

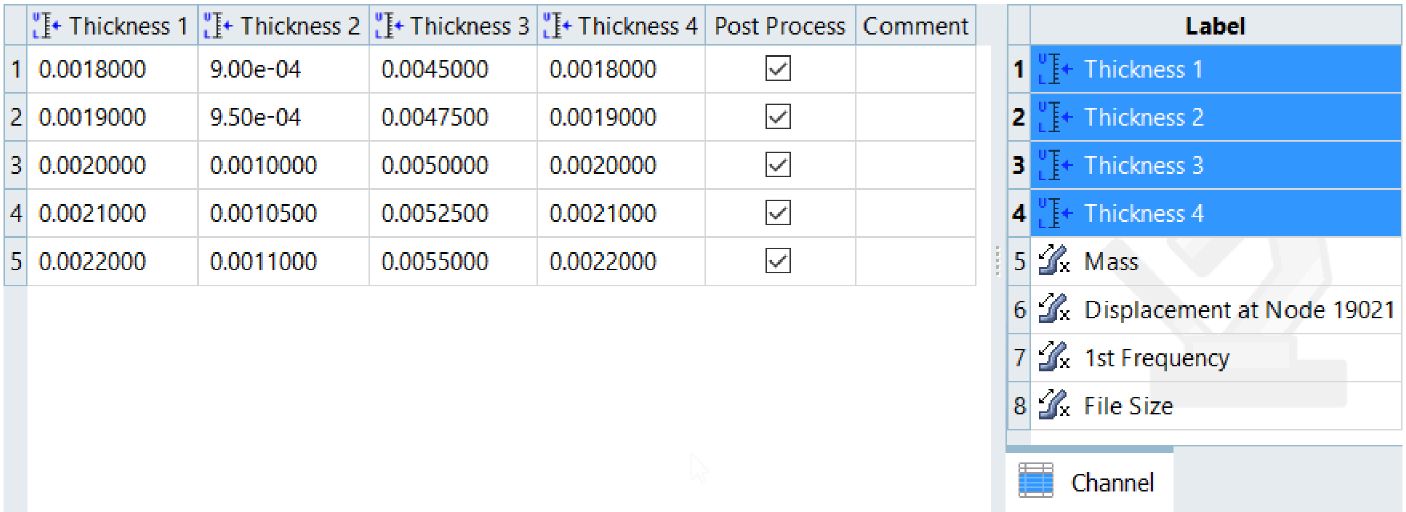
Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processing tab.

1. From the Post Processing step, click the **Summary** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Figure 77:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot.

Option	Description
	Note: Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 78 run data was selected for a set of lines. In Figure 79, you can see that when Styling is low, Height is high.

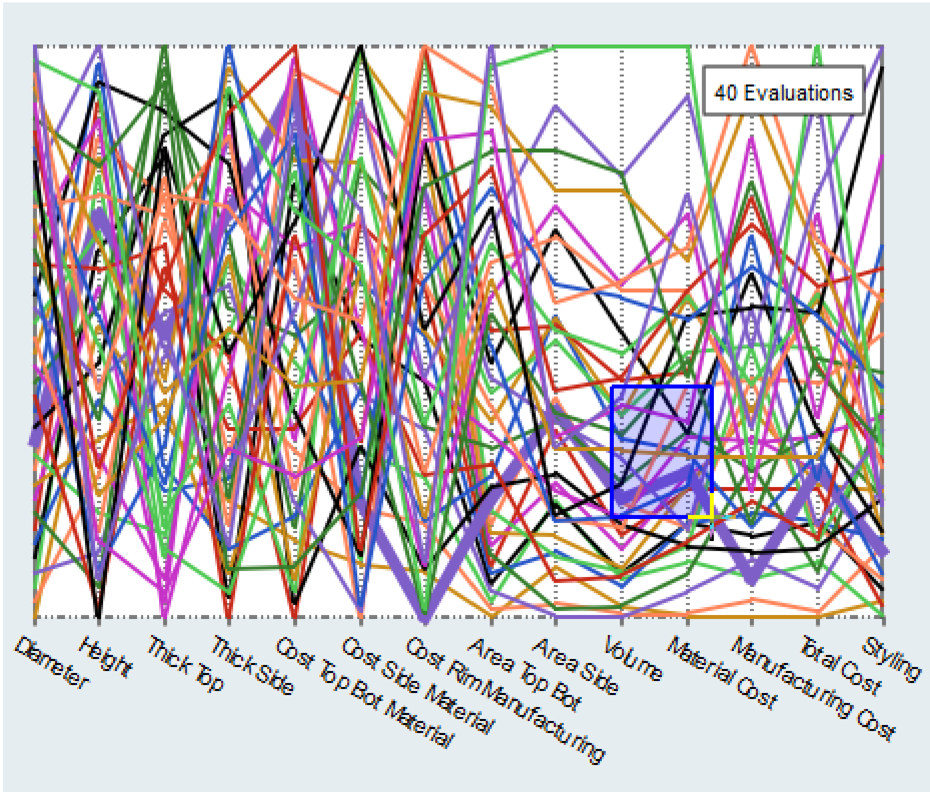
Option	Description
	

Figure 78:

Option **Description**

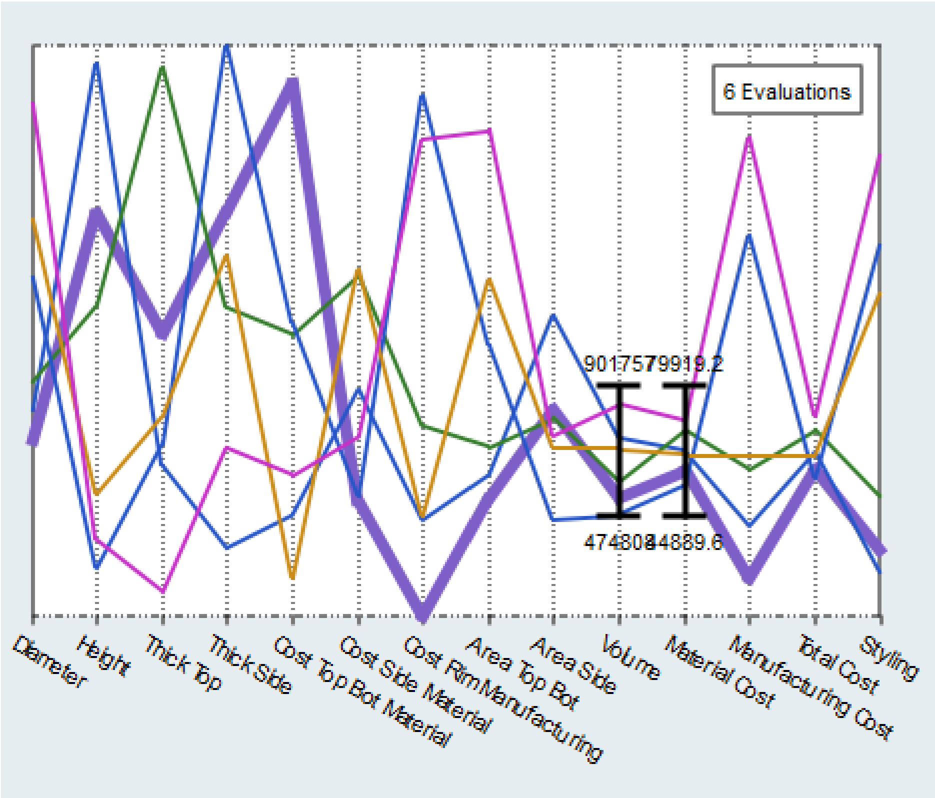


Figure 79:

Filter run data

Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog.

- From the Filter column, select the input variables and output responses to plot.
- From the Filter Min and Filter Max columns, enter values to filter.

The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

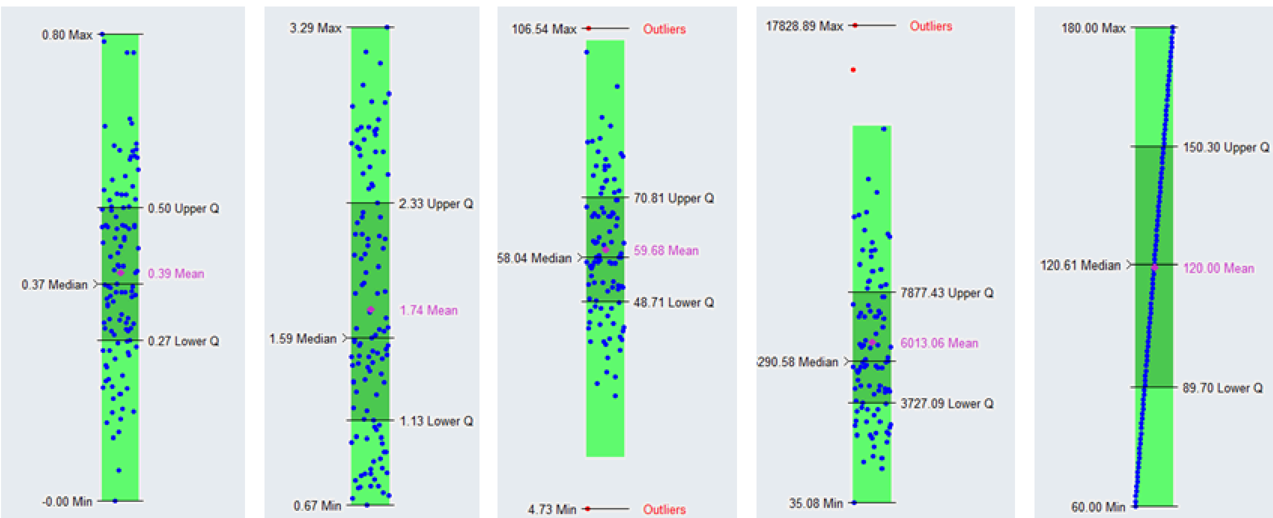


Figure 80:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 81 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

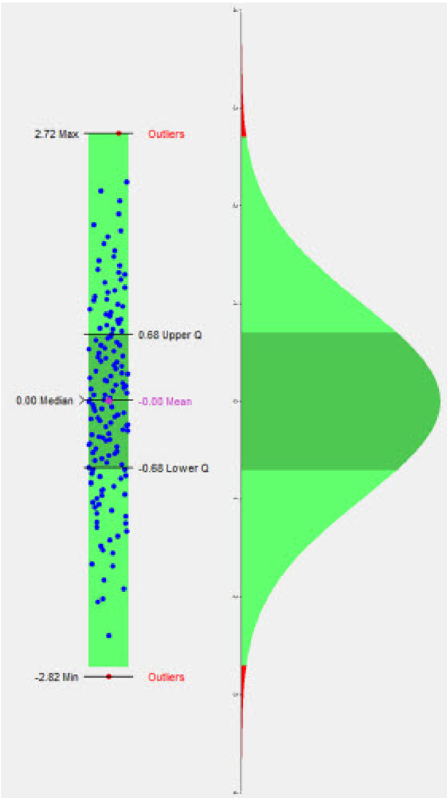


Figure 81:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

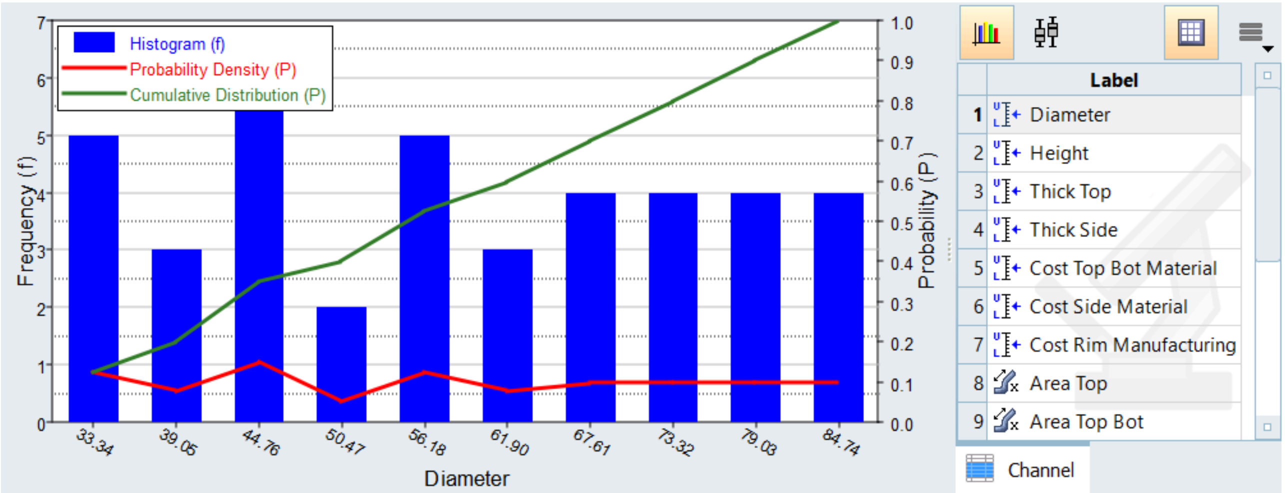


Figure 82:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processing tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Post Processing step, click the **Scatter** tab.
2. Select data to display in the scatter plot.
   - Use the Channel selector to select two dimensions of data to plot.

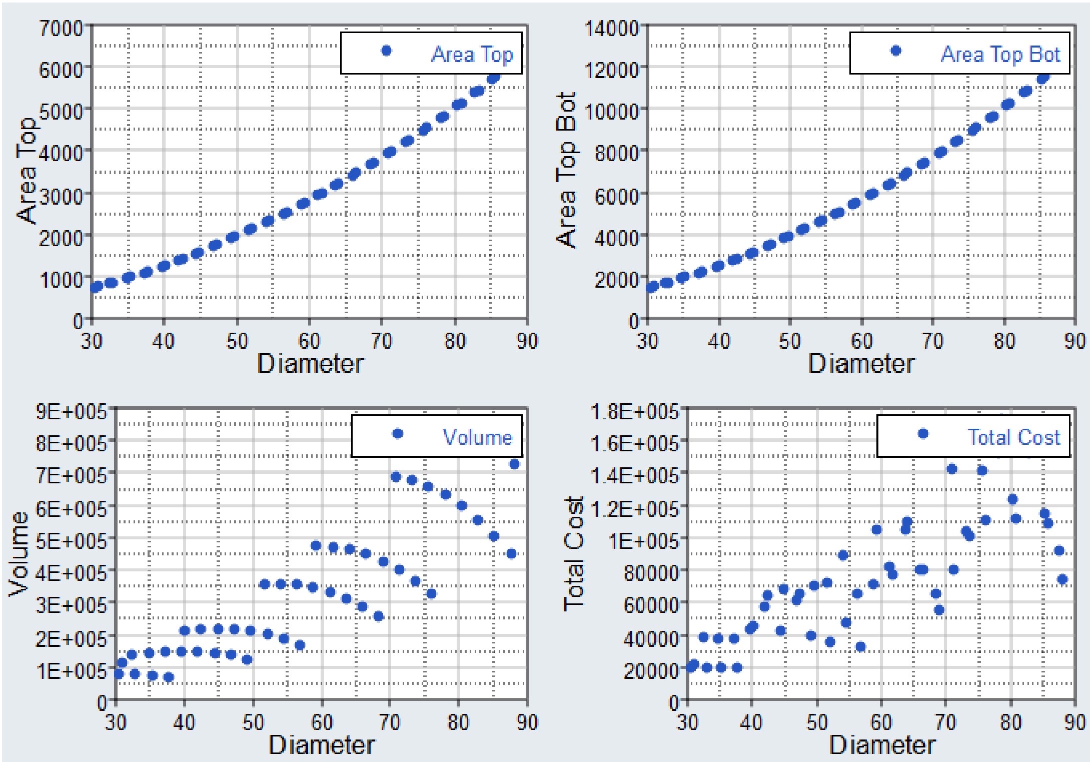


Figure 83:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

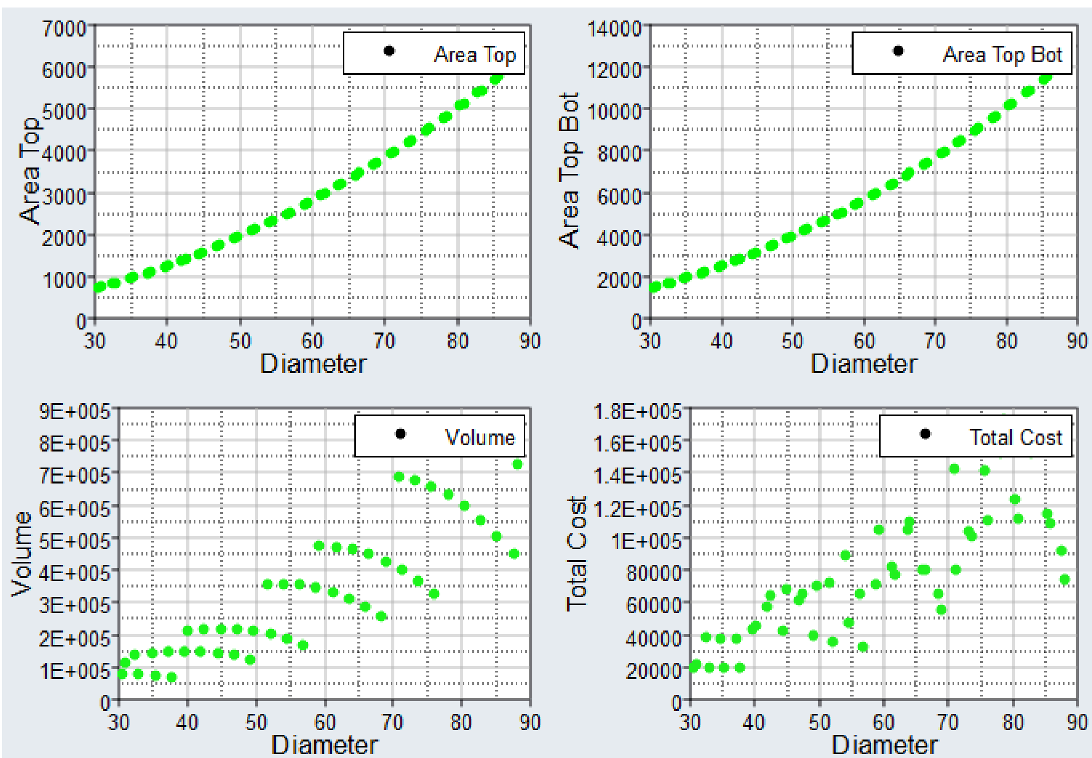


Figure 84:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

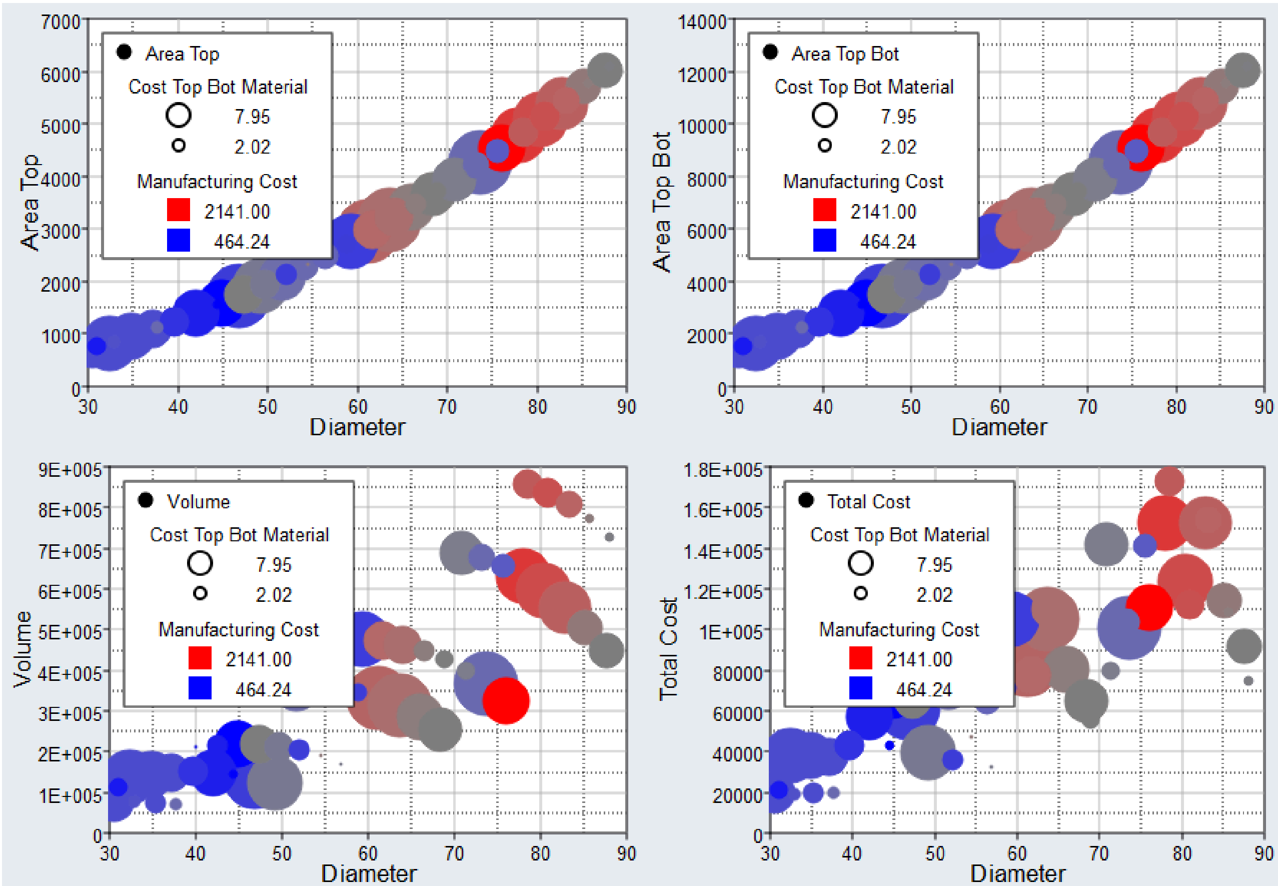


Figure 85:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Scatter Tab Settings

Settings to configure the plots displayed in the Scatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds

Display the X bounds in the plot.

Y-Bounds

Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)

Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank

Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 86: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 87: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥

Show only the column/rows with cells over the specified threshold.

Show Variables and Responses

Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

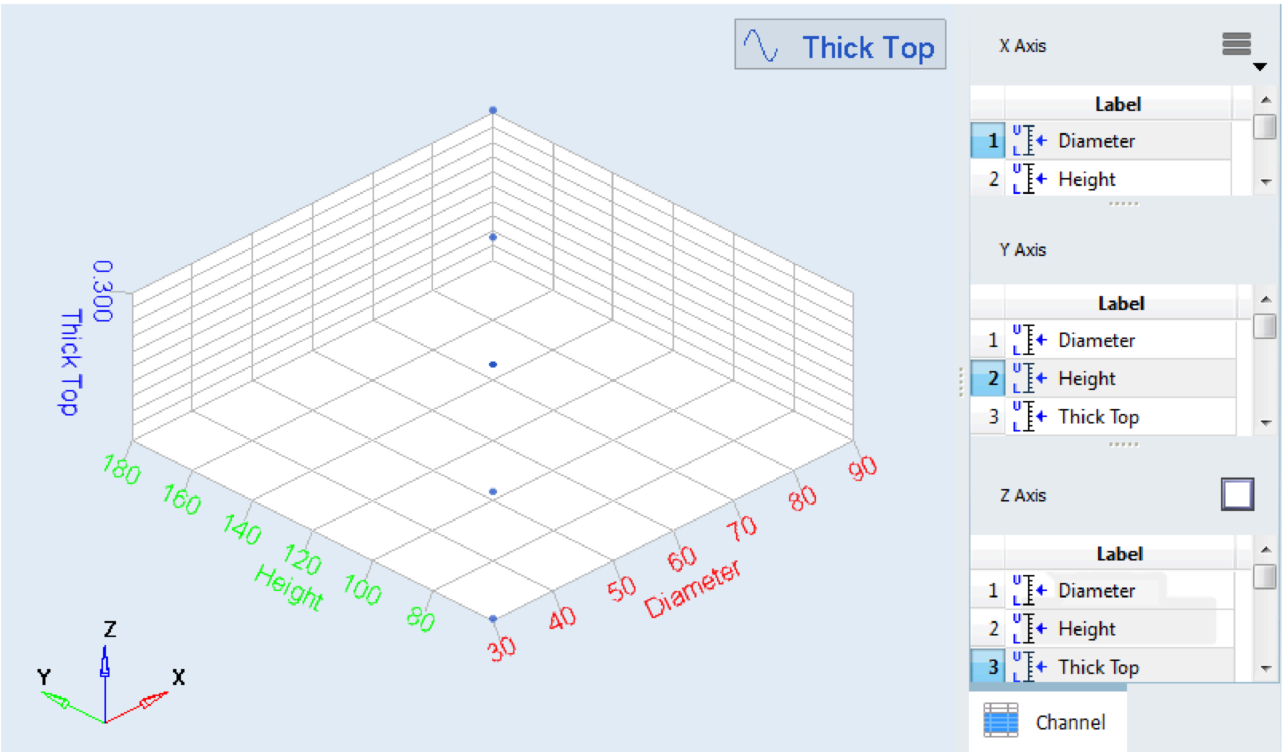


Figure 88:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

Tip: For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the biplot.

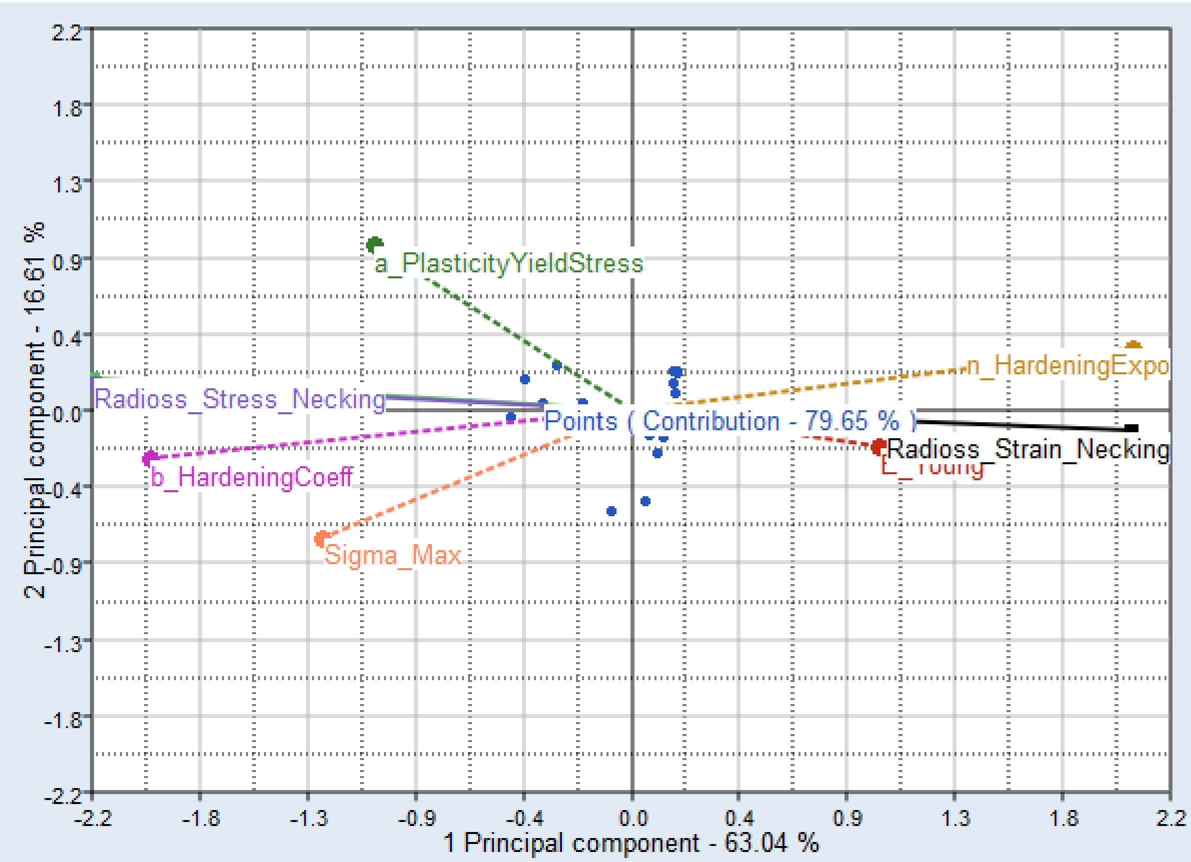


Figure 89:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
   a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

> **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

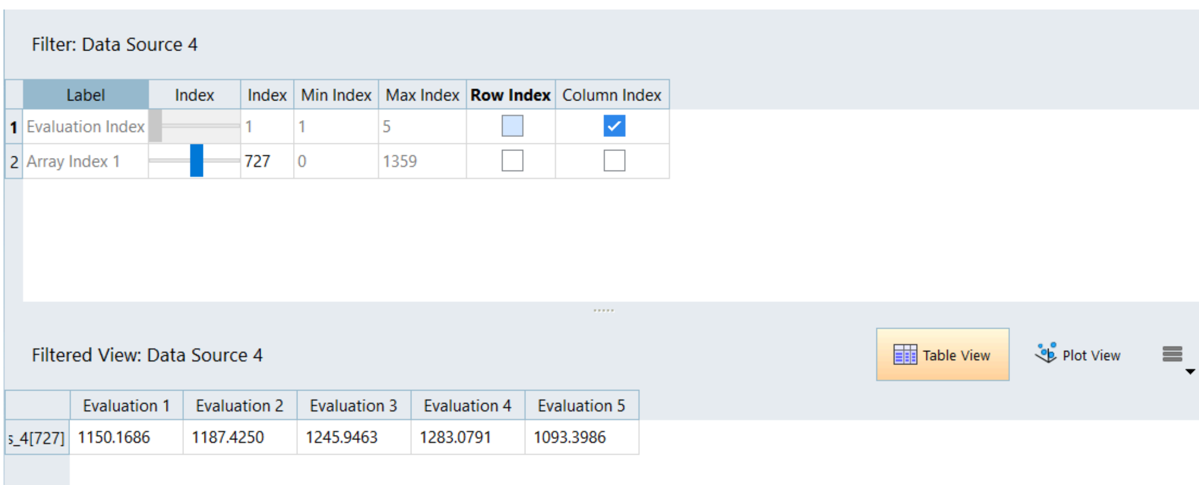


Figure 90:

5. Analyze the table.

Gradient Post Processing

Visualize gradients using vectors.

Analyze Vector

Analyze the vector in a gradient plot from the Gradient tab. Representing gradients as a vector field is an effective way to see gradients in space.

1. From the Post-Processing step, click the **Gradient** tab.

2. Use the Inputs and Output tabs of the Channel selector to select three dimensions of data to plot.
3. Analyze the direction and intensity of the vector created using the selected data sets.

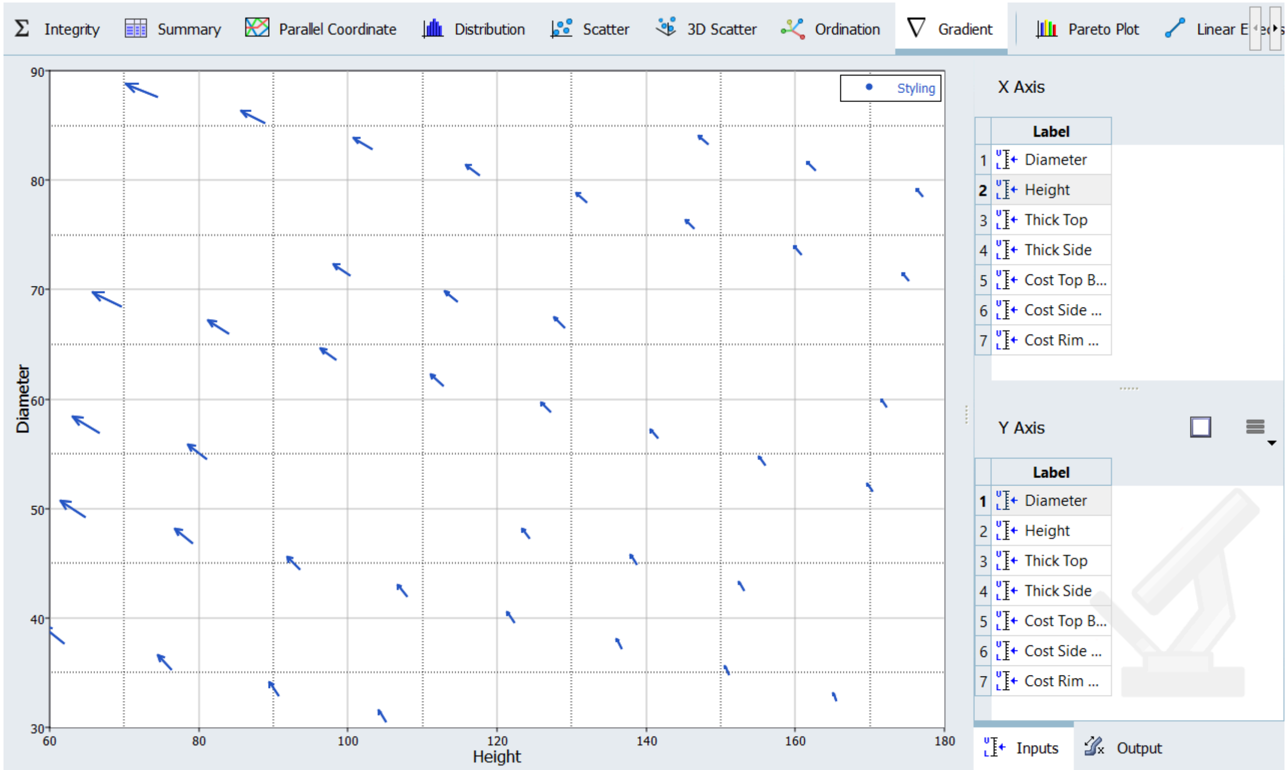


Figure 91:

Gradient Tab Settings

Settings to configure the plots displayed in the Gradient tab.

In the Gradient tab, use the Channel selector to select data to display in the gradient plot.

Channel Settings

Inputs	X-Bounds: display the X bounds in the plot. Y-Bounds: display the Y bounds in the plot.
Output	Gradient: display the vector in the plot.

Pareto Plot Post Processing

Plot the effects of input variables on output responses in hierarchical order (highest to lowest).

Plot the Effects of Variables on Responses in Hierarchical Order

Rank the effects of input variables on output responses in hierarchical order (highest to lowest) in the Pareto Plot post processing tab.

1. From the Post Processing step, click the **Pareto Plot** tab.
2. Using the Channel selector, select the response to plot.

> **Tip:** Analyze multiple responses simultaneously by switching the Multiplot option to (multiple plots) and selecting the responses to plot using the Channel selector.

3. Analyze the pareto plot.

   The effect of input variables on output responses is indicated by bars. Hashed lines with a positive slope indicates a positive effect. If an input variable increases, the output response will also increase. Hashed lines with a negative slope indicates a negative effect. Increasing the input variable lowers the output response.

   A line represents the cumulative effect.

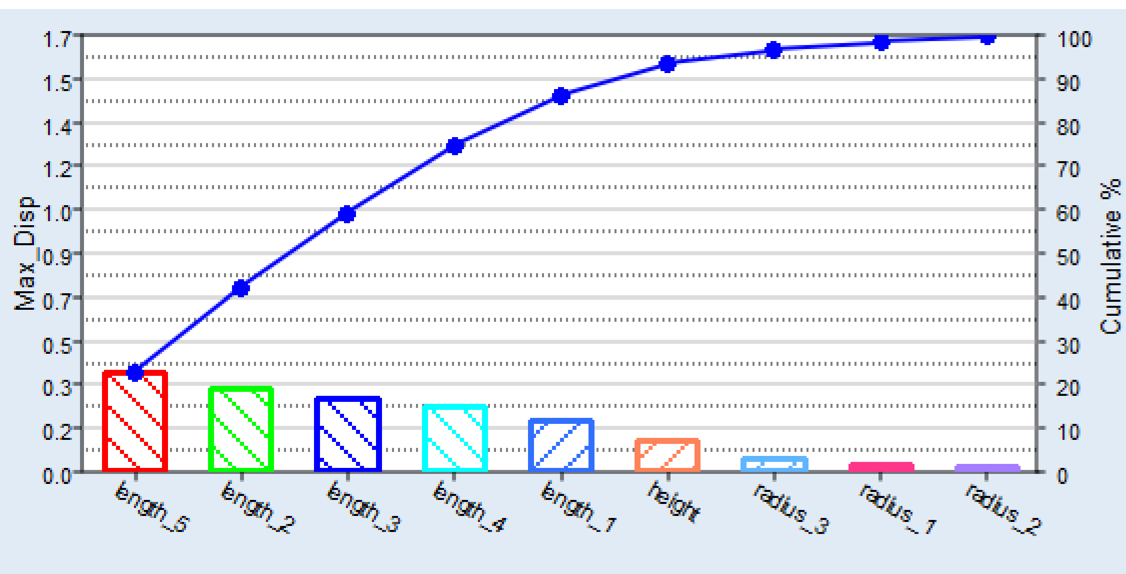


Figure 92:

Configure the pareto plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Pareto Plot Tab Settings.

Pareto Plot Tab Settings

Settings to configure the plots displayed in the Pareto Plot post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Effect curve	Show line to represent the cumulative effect.
# Top factors displayed	Specify the number of input variables (bars) displayed in the plot. **Note:** This settings does not change the calculated effects.
Multivariate Effects	Calculate the effect using all input variables simultaneously.
Linear Effects	Calculate the effect using each input variable independently. For more information about linear effects, refer to Linear Effects Post Processing.
Include Interactions	Include first order, two way interactions along with first order effects, and calculate interactions consistently with the choice of linear or multi-variate effects. For more information about interactions, refer to Interactions Post Processing.
Exclude dependent/linked inputs	Only show the independent input variables. **Tip:** Excluding dependent/link inputs reduces redundant information.

Linear Effects Post Processing

Measure the result of a single variable moving.

Analyze Linear Effects

Analyze the effects of input variables on output responses while ignoring the effects of other input variables in the Linear Effects post processing tab.

Linear effects are calculated using a linear regression model for the normalized input variable ranges of [-1, 1]. The linear effect value of input variable x on output response f(x, y) doubles the coefficient a1 of the regression model for f(x)=a0+a1*x. This term, A1, can be written compactly as cov(x,y)/var(x)*(x_max-x_min).

1. From the Post Processing step, click the **Linear Effects** tab.
2. Using the Channel selector, select the input variable(s) and output response(s) to analyze.
3. Change the format to display linear effects by switching the view between (Linear Effects plot) and (Linear Effects Table).
   - In the Linear Effects Plot view, effects are represented by the slope of a line.
   - In the Linear Effects Table view, effects are presented in a tabular format.

4. Analyze the linear effects.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the linear effect table or plot display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Linear Effects Tab Settings.

Linear Effects Tab Settings

Settings to configure tables/plots displayed in the Linear Effects post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Normalize X-Axis	
Scatter	Show scatter points in Linear Effects table.

Example: Linear Effects

For two-level design of experiments, linear effect values can also be calculated as the difference between the average output responses when the input variable is at its lower value and when the input variable is at its upper value.

Given a two variable, two-level Full Factorial DOE matrix over the design space of [0:2] on both parameters as:

Table 5: Design Matrix

Run	X	Y	F(X,Y)
1	0	0	1
2	0	2	401
3	2	0	1601
4	2	2	401

- When X is at lower level, the mean output response is (1 + 401) / 2 = 201
- When X is at upper level, the mean output response is (1601 +401) / 2 = 1001
- The effect of X on F is then (1001 - 201) = 800
- When Y is at lower level, the mean output response is (1 + 1601) / 2 = 801
- When Y is at upper level, the mean output response is (401 + 401) / 2 = 401
- The effect of Y on F is then (401 - 801) = -400

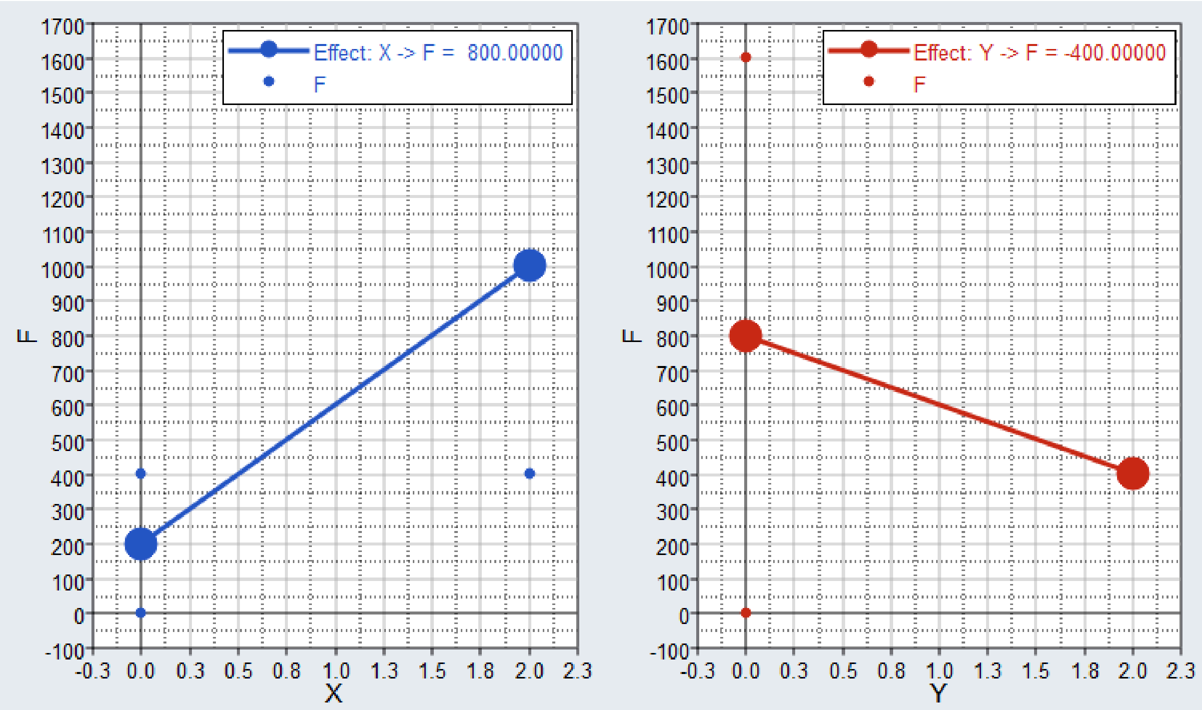


Figure 93: Linear Effects Plot

	Label	Varname	F
1	X	x	800.00000
2	Y	y	-400.00000

Figure 94: Linear Effects Table

Note: HyperStudy calculates linear effects using a regression based on the data set. In this example, the regression is f(x) is ao+a1*x; where a1 is equal to 400.0 and x is between -1.0 and 1.0. The linear effect of the input variable x on the output response f(x,y) is 800.0.

Interactions Post Processing

Measure the result of two variables moving simultaneously.

Analyze Interactions

Analyze the effect of an input variable on an output response at varying levels of other input variables in the Interactions post processing tab.

An interaction is the failure of one input variable to produce the same effect on the output response at different levels of another input variable. In other words, the strength or the sign (direction) of an effect is different depending on the value (level) of some other variable(s). An interaction can be either positive or negative.

1. From the Post Processing step, click the **Interactions** tab.
2. Using the Channel selector, select the input variable(s) and output response(s) to analyze.
3. Change the format to display interactions by switching the view between (Interactions Plot) and (Interactions Table).
4. Analyze interactions.

Tip: Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Example: Interactions

In Table 6, interactions are calculated as:

- Effect of X when Y = +1 is (401 - 401) / 2 = 0
- Effect of X when Y = -1 is (1601 - 1) / 2 = 800

Interaction of X on Y is then (0 - 800) = -800

- Effect of Y when X = +1 is (401 - 1601) / 2 = -600
- Effect of Y when X = -1 is (401 - 1) / 2 = 200

Interaction of X on Y is then (-600 - 200) = -800

Table 6: Design Matrix

Run	X	Y	F(X,Y)
1	0	0	1
2	0	2	401
3	2	0	1601
4	2	2	401

Interactions are symmetric; that is:

Interaction XY = effect of (X) on effect of (Y) = effect of (Y) on effect of (X)

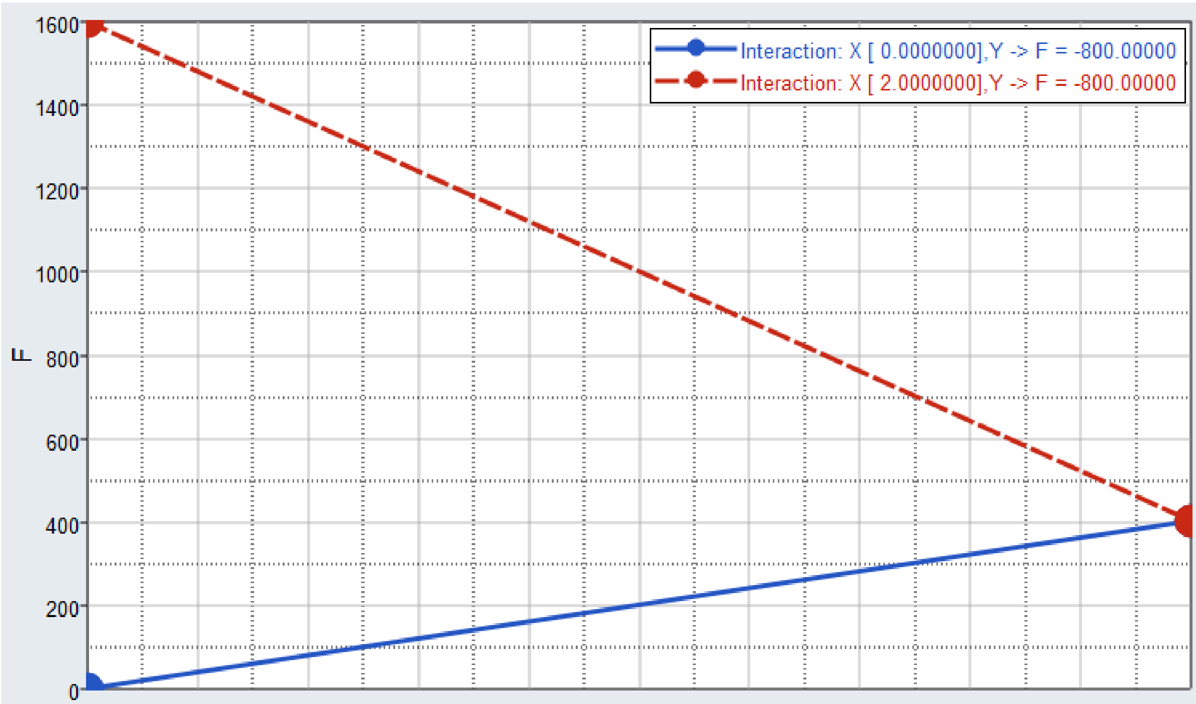


Figure 95: Interactions Plot

	Label	Varname	F
1	X	x	-800.00000
2	Y	y	N/A

Figure 96: Interactions Table

> **Note:** HyperStudy calculates interactions using a regression based on the data set. In this example, the regression is f(x,y) is ao+a1*x+a2*y+a3*x*y; where a3 is equal to -400.0 and x and y are between -1.0 and 1.0. The interaction of the input variables x and y on the output response f(x,y) is -800.0.

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	Description
HyperStudy Data	Generates a data report (`*.data`).
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.

Report Type	Description
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel.

3. Click **Create Report**.

4.2.2 Setup Fit Studies

A Fit is a mathematical model that is trained by data and is capable of predicting output response variables for a given set of input variables.

Add a Fit Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
   a) In the Label field, enter a name for the Fit.
   b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.

      By default, the Definition defined in the study Setup is selected.
   c) Under Select Type, select **Fit**.
   d) Click **OK**.

A new Fit is added to the Explorer.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.
3. Test Models.

4. Define Output Responses.
5. Review definitions in the following ways:

To:	Do this:
Review status	Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

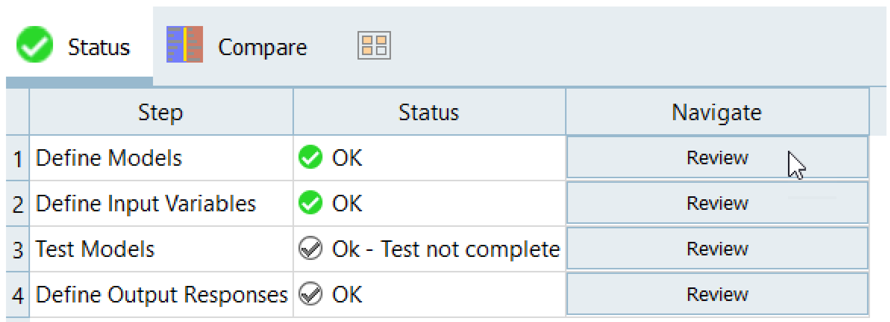


Figure 97:

Compare definitions	Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

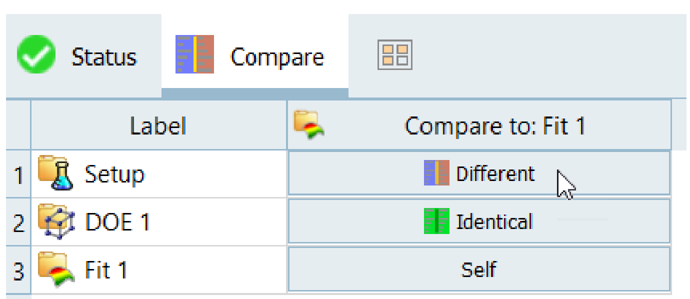
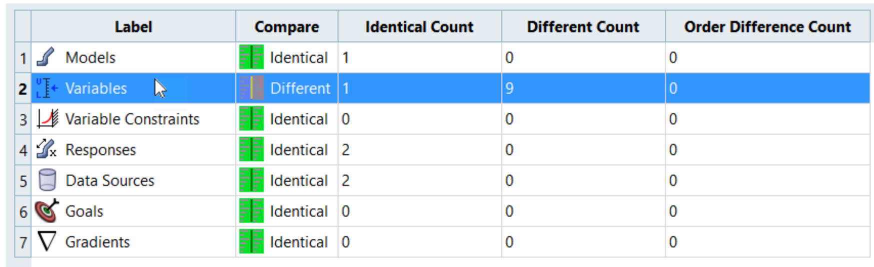


Figure 98:

The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.

To:	**Do this:**

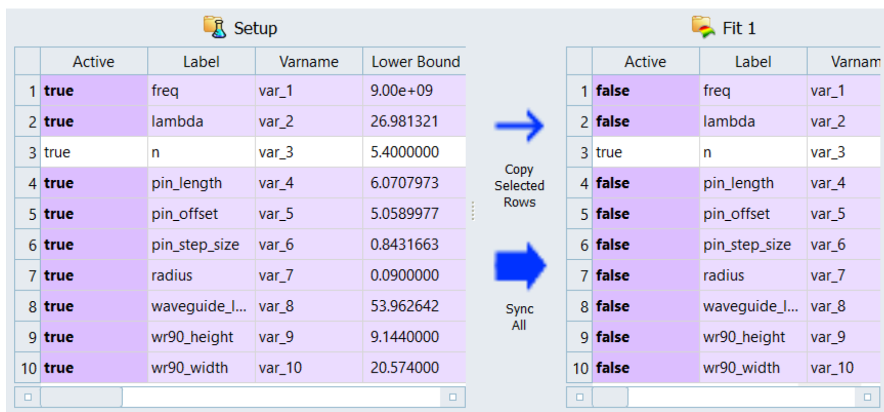
4. Click a channel to display a detailed comparison.



	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 99:

5. Sync data.
   - Click **Copy Selected Rows** to sync the single row.
   - Click **Sync All** to sync all rows.



	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 100:

Select a Numerical Method

Select a numerical method to use when evaluating the Fit.

1. In the Specifications step, go to the Specifications tab.
2. In the work area, Fit Type column, select a numerical method for each output response.
   By default, FAST is selected.
3. Optional: In the Settings tab, change settings as needed.
4. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog.

Fit Methods

Numerical methods available for a Fit approach.

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
Fit Automatically Selected by Training	General	N/A	N/A	Choose methods for Fit Automatically Selected by Training to consider.	Selects the most appropriate method and settings. It it recommended that you use this method unless you desire a specific method and settings.
HyperKriging	Interpolated data	###	##		The time to build the Fit and use the Fit (Evaluate From) increases with both the number of runs and the number of design variables in the input matrix. The number of design variables has more influence than the number of runs if order is larger than 1.
Least Squares Regression	Data trend lines	#	###		Noises can be screened out with this method.

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
					Closed form equations are available.
Moving Least Squares Method (MLSM)	General	##	##		The time to build the Fit and use the Fit (Evaluate From) increases with both the number of runs and the number of design variables in the input matrix. The number of design variables has more influence than the number of runs if order is larger than 1.
Radial Basis Function	Interpolate data	###	##		The time to build the Fit increases with both the number of runs and the number of design variables in the input matrix. The number of runs has more influence than the number of design variables. The run time for using the Fit in another

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
					approach (Evaluate From) is very small regardless of the size of the input matrix.

Fit Automatically Selected by Training

Selects the best available Fit from a list of available methods you have chosen. In addition to selecting the best method, Fit Automatically Selected by Training also automatically adjusts the individual settings (often called hyperparameters) to find the optimizing, predictive performance while avoiding overfitting.

Usability Characteristics

- Fits both noisy and non-noisy data.
- Reduces the methods on which Fit Automatically Selected by Training iterates in order to reduce the run time used to build the Fit.
- Can run in multi-execute, while simultaneously iterating over multiple responses.
- The Stepwise Regression Terms option for Least Squares Regression reduces the number of coefficients in the regression model to contain only the set that is statistically significant.
- The behavior and characteristics of the underlying methods are the same as when the methods are directly applied. See their respective documentation pages for details.
- Gradient information can be used to boost performance for the methods that support gradients.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Least Square Regression	On	On or Off	**On** Use Stepwise Regression Terms to reduce the number of terms in the regression to the statistically significant set. **Off** Do not consider Least Squares Regression in

Parameter	Default	Range	Description
			the ensemble list of methods.
Stepwise Regression Terms	Full Quadratic	Linear Squared Cubic Interaction Full Quadratic Full Cubic	Controls the maximal set of terms considered in stepwise Least Squares Regression. **Linear** First order terms only. y=A+Bx+Cy **Squared** Second order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2 **Cubic** Third order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2+Fx^3+Gy^3 **Interaction** Linear and the cross terms. y=A+Bx+Cy+Dxy **Full Quadratic** Complete second order polynomial. **Full Cubic** Complete third order polynomial.
Moving Least Squares	On	On or Off	**On** Consider Moving Least Squares Method in the ensemble list of methods. **Off** Do not consider Moving Least Squares Method in the ensemble list of methods.
Radial Basis Function	On	On or Off	**On** Consider Radial Basis Function in the ensemble list of methods.

Parameter	Default	Range	Description
			Off Do not consider Radial Basis Function in the ensemble list of methods.
Use Gradient Data	On	On or Off	**On** Allow methods to be enhanced by gradient information when it is available. **Off** Do not allow methods to be enhanced by gradient information.

HyperKriging

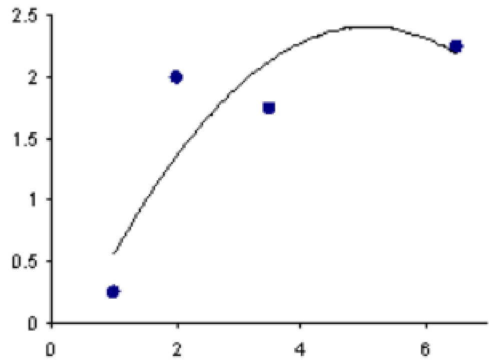
Creates predictive models with data sets coming from deterministic computer simulations, an area of application commonly known as the Design and Analysis of Computer Experiments (DACE).

These experiments are unique because they do not require some concepts such as replication. This approximation method is designed to tightly pass through and smoothly interpolate between the known points.

Usability Characteristics

- Attempts to go through the exact sampling points, and in general, the residuals are small, if not zero. As a result, diagnostic measures using only the complete input matrix do not produce meaningful values. Cross-validation results provide some diagnostics using a special scheme using only the input points. To get detailed diagnostics on the quality of a HyperKriging Fit, it is suggested that you use a testing matrix.
- Suitable for modeling highly nonlinear output response data that does not contain numerical noise.
- Applicability of HyperKriging and Radial Basis Function methods are similar in terms of physics (they both are suggested for highly nonlinear output responses with no noise). It is suggested that you use HyperKriging for large studies that contain a large number of sampling points, whereas, Radial Basis Function is suggested for studies with a large number of variables.

> **Note:** As a result, Radial Basis Function Fits are evaluated faster than HyperKriging Fits when used in approaches.



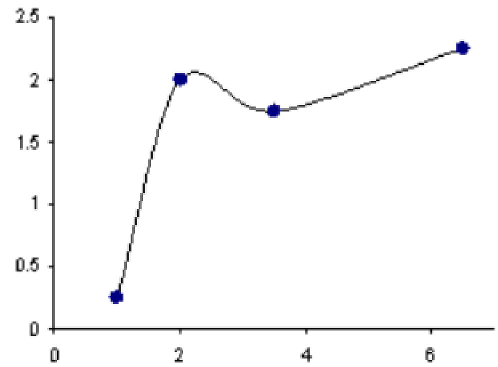


Figure 101: Comparison of a Quadratic Polynomial Model and aHyperKriging Model for a Function of a Single Variable

The plot on the left is a Least Squares Quadratic Regression, and the plot on the right is a HyperKriging model.

Settings

No settings available at this time.

Least Squares Regression

Creates a regression polynomial of the chosen order such that the sum of the squares of the differences (residuals) between output response values predicted by the regression model and the corresponding simulation model is minimized.

For example,

$$\min E = \sum_{i=1}^{n} \left(f_{i}^{predicted} - f_{i} \right)^{2}$$

where n is the number of designs, $f^{predicted}$ is the output response value predicted by the regression model for the ith design, and f is the output response value from the simulation of the ith design. This is achieved by finding the regression model coefficient values that sets the derivative of E, with respect to each unknown coefficient, to zero.

Least Squares Regression Model

The least squares regression model in HyperStudy is the polynomial expression that relates the output response of interest to the factors that were varied.

Selection of the proper model is required to create an accurate approximation. However this requires a prior knowledge of the behavior of the output responses (linear, non linear, noisy, and so on) and enough runs to feed the selected model.

Types of regression models include:

Linear Regression Model

$$F(x) = a_0 + a_1 x_1 + a_2 x_2 + (error)$$

Interaction Regression Model

$$F(x)=a_0+a_1x_1+a_2x_2+a_3x_1x_2+(error)$$

Quadratic Regression Model (2nd Order)

$$F(x)=a_0+a_1x_1+a_2x_2+a_3x_1x_2+a_4x_12+a_5x_22+(error)$$

An approximation is only as good as the uniformity of the design sampling and, for example, a two-level parameter only has a linear relationship in the regression. Higher order polynomials can be introduced by using more levels for the factors, but then, using more levels results in more runs.

If n is the number of input variables:

- A linear regression model requires $n+1$ runs.
- An interaction regression model requires $\frac{(n+1)(n+2)}{2}-n\ runs$.
- A quadratic regression model requires $\frac{(n+1)(n+2)}{2\ runs}$.

Usability Characteristics

- HyperStudy will create the least squares regression of any order, however, in most cases polynomials of the 4th order or higher do not increase accuracy.

  > **Note:** A custom order can be defined from the Regression Terms tab.

- Suppress regression terms that are known to be insignificant.
- Residuals and diagnostics should be used to gain an understanding of the quality of the Fit.
- Quality of a Least Squares Regression Fit is a function of the number of runs, order of the polynomial, and the behavior of the application.
- If the residuals and diagnostics are not good for a Least Squares Regression Fit, than you can increase the order of the Fit provided you have enough runs to fit that specific order.

  > **Note:** If n is the number of input variables:
  > - A linear model requires $n+1$ runs.
  > - An interaction model requires $\frac{(n+1)(n+2)}{2}-n$ runs.
  > - A quadratic model requires $\frac{(n+1)(n+2)}{2}$ runs.

- If increasing the order does not improve the Fit quality, then you may want to inspect the input matrix collinearity and optionally add more runs. You should try the other available Fit methods as your application may have more non-linearity than polynomials can handle.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Regression Model	Linear	Linear Squared Cubic Interaction Full Quadratic Full Cubic Custom	**Linear** First order terms only. y=A+Bx+Cy **Squared** Second order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2 **Cubic** Third order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2+Fx^3+Gy^3 **Interaction** Linear and the cross terms. y=A+Bx+Cy+Dxy **Full Quadratic** Complete second order polynomial. **Full Cubic** Complete third order polynomial. **Custom** User defined order and terms.

Moving Least Squares Method (MLSM)

Builds a weighted least squares model where the weights associated with the sampling points do not remain constant.

Weights are functions of the normalized distance from a sampling point to a point x, where the surrogate model is evaluated. The weight, associated to a sampling point, decays as the evaluation point moves away from it. The decay is defined through a decay function. For each point x it reconstructs a continuous function biased towards the region around that point.

Usability Characteristics

- Suggested to be used for nonlinear and noisy output responses.
- Residuals and diagnostics should be used to gain an understanding of the quality of the Fit.
- Use a Testing matrix in addition to an Input matrix for better diagnostics.

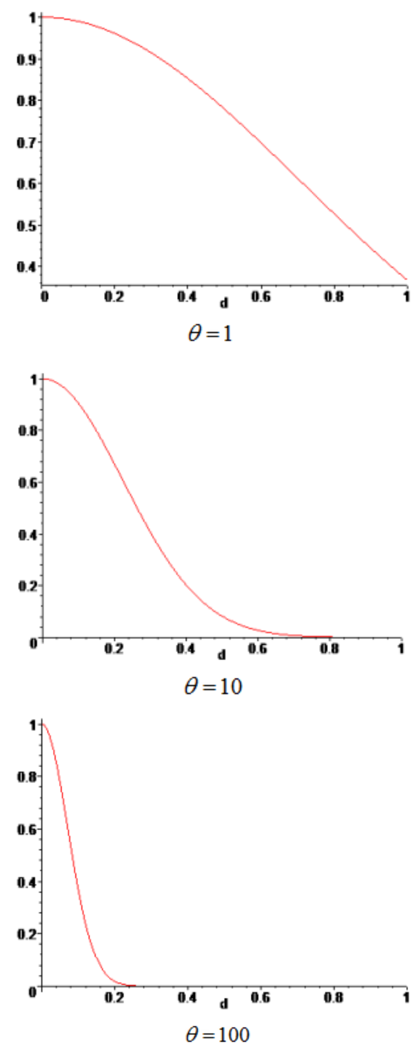
- Quality of a Moving Least Squares Method Fit is a function of the number of runs, order of the polynomial and the behavior of the application.
- If the residuals and diagnostics are not good for a Moving Least Squares Method Fit, than you can increase the order of the Fit provided you have enough runs to fit that specific order.
- Because the weights are not constant in Moving Least Squares Method, there is no analytical form and an equation can not be provided.

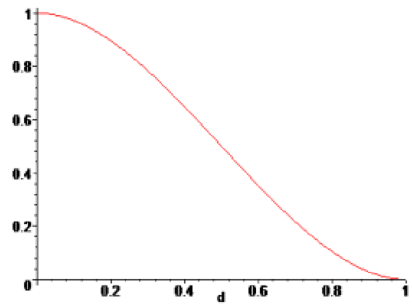
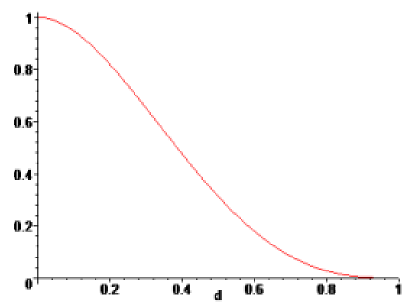
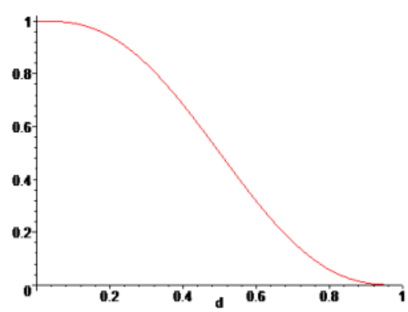
Settings

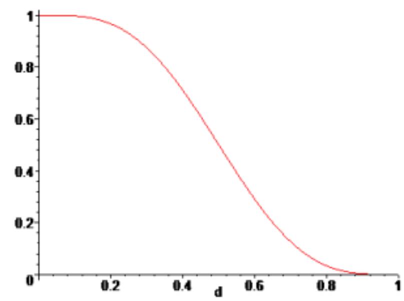
In the Specifications step, Settings tab, change method settings.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Order to improve the Fit quality.

Parameter	Default	Range	Description
Fit Parameter	5.0	>= 0.0 <= 10.0	Controls the effect of screening out noise; the larger value, the less effect.
Minimum Weight	0.001	> 0.0	Minimum weight.
Number of Excess Points	3	>=0	Number of excessive points to build Moving Least Squares Method.
Regression Model	Linear	Linear Squared Cubic Interaction Full Quadratic Full Cubic Custom	Order of polynomial function.
Weighting Function	Gaussian	Gaussian (Recommended Cubic Fourth Order Fifth Order Seventh Order	Type of weighting function. *Gaussian* $W_i = \exp(-\theta r_i 2)$ where r_i is the normalized distance from the i-th sampling point to a current point. The parameter θ defines the "closeness of fit", the case θ=0 is equivalent to the traditional Least Squares Regression. When the parameter θ is large, it is possible to obtain a very close fit

Parameter	Default	Range	Description
			through the sampling points, if desired. The images in Figure 102 illustrate the change of the weight over the interval [0,1] where the sampling point is at r = 0.  *Figure 102:* *Cubic* $w_i = 1 - 3p_i^2 + 2p_i^3$ where $p_i = r_i / R_{max}$, R_{max} is the normalized radius of the sphere of influence.

Parameter	Default	Range	Description
			 *Figure 103:* The normalized radius of the sphere of influence $R_{\max}$ inversely relates to the closeness of fit parameter, for example the smaller the value of $R_{\max}$, the closer fit is obtained. *Fourth Order* $w_i = 1 - 6p_i^2 + 8p_i^3 + 3p_i^4$  *Figure 104:* *Fifth Order* $w_i = 1 - 10p_i^3 + 15p_i^4 + 6p_i^5$  *Figure 105:*

Parameter	Default	Range	Description
			Seventh Order $w_i = 1 - 35p_i^4 + 84p_i^5 + 70p_i^6 + 20p_i^7$  *Figure 106:*

Radial Basis Function

Uses linear combinations of basis functions, such as linear, cubic, thin-plate spline, Gaussian, multiquadric, and inverse-multiquadric. These basis functions are observed to be accurate for highly nonlinear output responses but not for linear output responses.

To remedy this deficiency, in HyperStudy, a Radial Basis Function model is augmented with a polynomial function.

$$f(x) = \sum_{i=1}^{n} \lambda_i \phi(\| x - x_i \|) + \sum_{j=1}^{n} c_j p_j(x)$$

where n is the number of sampling points, x is a vector of input variables, x_i is the ith sampling point, $\| x - x_i \|$ is the Euclidean norm, ϕ is a basis function, and λ_i is the coefficient for the ith basis function. $p_j(x)$ is a low-order (constant or linear) polynomial function; k is the total number of terms in the polynomial, and $cj(j = 1, 2..k)$ are the unknown coefficients.

Usability Characteristics

- Attempts to go through the exact sampling points, and in general, the residuals are small, if not zero. As a result, diagnostic measures using only the complete input matrix do not produce meaningful values. Cross-validation results provide some diagnostics using a special scheme using only the input points. To get detailed diagnostics on the quality of a Radial Basis Function Fit, it is suggested that you use a testing matrix.
- Suitable for modeling highly nonlinear output response data that does not contain numerical noise.
- Applicability of HyperKriging and Radial Basis Function methods are similar in terms of physics (they both are suggested for highly nonlinear output responses with no noise). It is suggested that you use HyperKriging for large studies that contain a large number of sampling points, whereas, Radial Basis Function is suggested for studies with a large number of variables.

> **Note:** As a result, Radial Basis Function Fit are evaluated faster than HyperKriging Fits when used in approaches.

Settings

In the Specifications step, Settings tab, change method settings.

> **Note:** For most applications the default settings work optimally.

Parameter	Default	Range	Description
Augmented Function	Constant	Constant Linear Custom	Type of augmented function.
Maximum Points	2000	>=100	Maximum number of points for building Radial Basis Function; if number of building points is larger than maxnpt, then the point reduction algorithm is activated and a warning message is shown; the purpose of introducing maxnpt is to reduce computational effort for large scale problems.
RBF Type	CS21	Multiquadric CS21 (formally knows as Wu's Compactly Supported (2,1)) Gaussian	Type of Radial Basis Function.
Relaxation Parameters	1.0	>=0.0	Relaxation parameter d used in Radial Basis Function; if Radial Basis Function is CS21 or Gaussian, and d is set to 0.0 by users, then Radial Basis Function will automatically set d = 1.0e-6.

Select Matrices

Import and modify the design matrices and associated results for the creation of the approximation model.

The matrix and results should be imported from an existing DOE or Stochastic approach and can be further edited on the fly.

Active matrices are automatically imported.

1. In the Explorer, for the Fit, go to the Select Matrices step.
   A new matrix is created and added to the list of matrices.
2. Define the matrix by modifying its corresponding cells in the work area.
   a) Enter a label.
   b) Select a matrix type.

Input Matrix	Data will be used to create the fit and tune its parameters.
Testing Matrix	Data will be used to assess the quality of the fit.
Input + Testing Matrix	Data can be partitioned into input and testing by specifying the number of runs or percentage.

   c) Select a matrix origin.

   The origin settings names the approach from which the matrix is derived (in the current study).

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

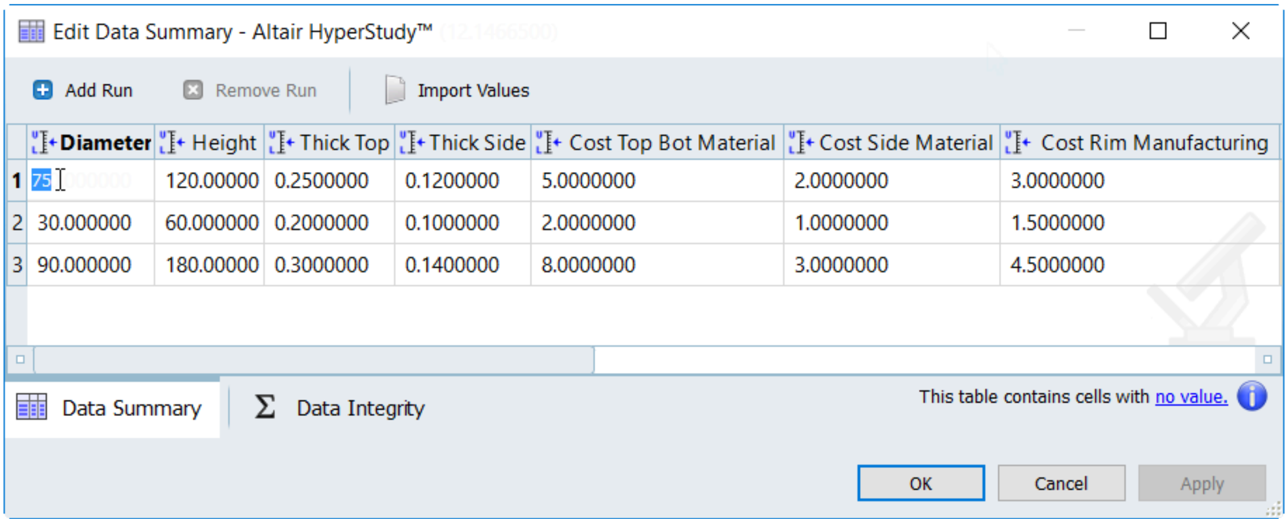


Figure 107:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.
   - Manually enter run data.
   - Copy and paste run data into the run matrix.

     Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

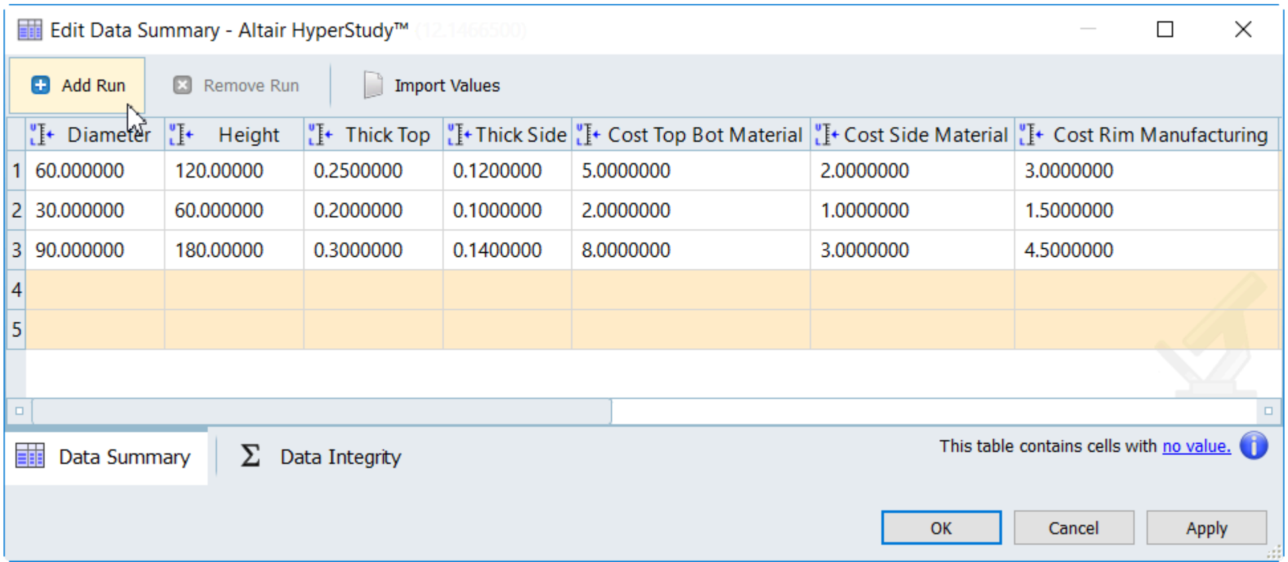


Figure 108:

Tip: Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.

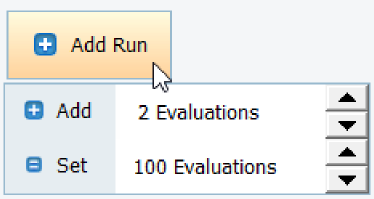


Figure 109:

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.
5. Select the source that contains run data.
   - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
   - For Approach evaluation data, select the approach that contains run data.

- For HyperStudy post processing file, select the source file.

6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.

Filter Run Data

Use the Filter tab to filter the run data included in a Fit approach.

1. Go to the **Fit** > **Specifications** step and select the **Filter** tab.
2. Click the corresponding checkbox to activate a filter.

Filter Name	Filter Description
Filter Outliers	Removes outliers. For more information about outliers, refer to About Box Plots.
Filter Duplicates	Removes duplicate information. **Note:** Filtering duplicates can cause unexpected results when using interpolating fits.
Filter Bad Numbers	Removes runs that contain undefined or unrepresentable data (NaN, inf, and so on).
Filter Excluded	Removes runs that were marked as excluded from post-processing in the source matricies. For more information about editing run matricies, refer to Edit the Run Matrix.

3. Click **Apply**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.
2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.
3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

   By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

	Active	Task	Batch
1	☐	Create Design	☐
2	☑	Write Input Files	☐
3	☑	Execute Analysis	☐
4	☑	Extract Output Responses	☐
5	☐	Purge ...	☐
6	☐	Create Reports	☐

Figure 110:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to stop the submission of new jobs. Click ≡ to select an execution mode.

Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

A Fit can run in multi-execute while simultaneously iterating over multiple responses.

Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

- Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
- Horizontal execution mode sequences the write, execute, and extract task for each run independently.

- Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

Fit Output Files

Output files generated from the a Fit.

<fit_variable_name>_anova.dat

File Creation

This file is created upon saving the study if Least Squares approximations have been created.

File Location

`<study_directory>/approaches/<fit_variable_name>/<fit_variable_name>_anova.dat`

File Contents

Result	Format	Description
ANOVA	ASCII	The analysis of variance (ANOVA) results are given in table form for each response approximation.

<fit_variable_name>_approximations.slk

File Creation

This file is created upon saving the study if Least Squares approximations have been created.

File Location

`<study_directory>/approaches/<fit_variable_name >/`
`<fit_variable_name>_approximations.slk`

File Contents

Result	Format	Description
Regression	Excel	Spreadsheet that holds the response surface created in the approximation. Use this file to make trade-off studies by modifying the Current Value of each input variable.

Result	Format	Description
		It also contains sensitivity information.

<fit_variable_name>.hstds

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<fit_variable_name >/<fit_variable_name>.hstds`

File Contents

Result	Format	Description
Run Matrix Data	hstds, binary	Hstds files stores the retained data sources; direct access data using the `.hstds` file is not suggested.

<fit_variable_name>.hstdf

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<fit_variable_name >/<fit_variable_name>.hstdf`

File Contents

Result	Format	Description
Run Matrix Data	hstdf, binary	Hstdf files store the run data; however, direct access to the data using the hstdf files are not suggested.

Review Evaluation Results

Review the input variable and output response values for each run, as well as review the run files.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Evaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.
4. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.

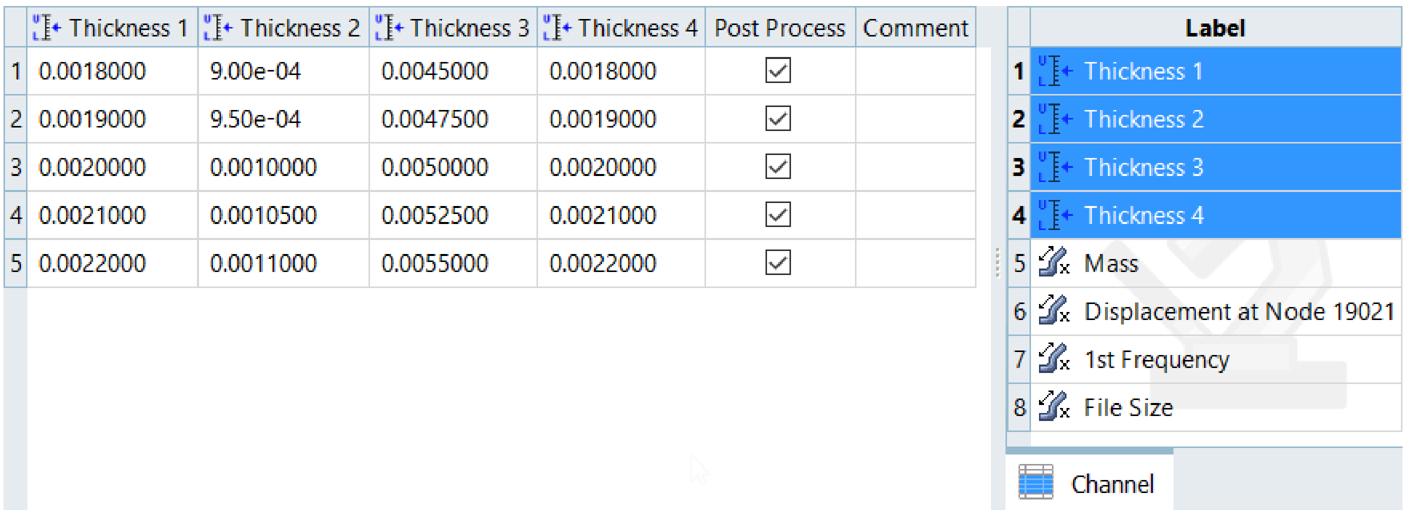


Figure 111:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

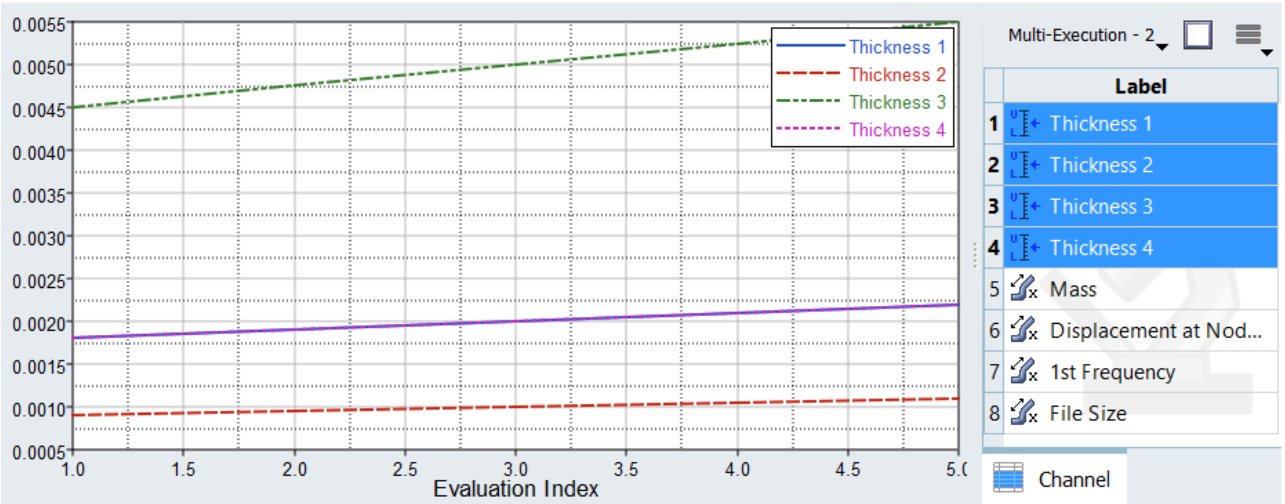


Figure 112:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Evaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

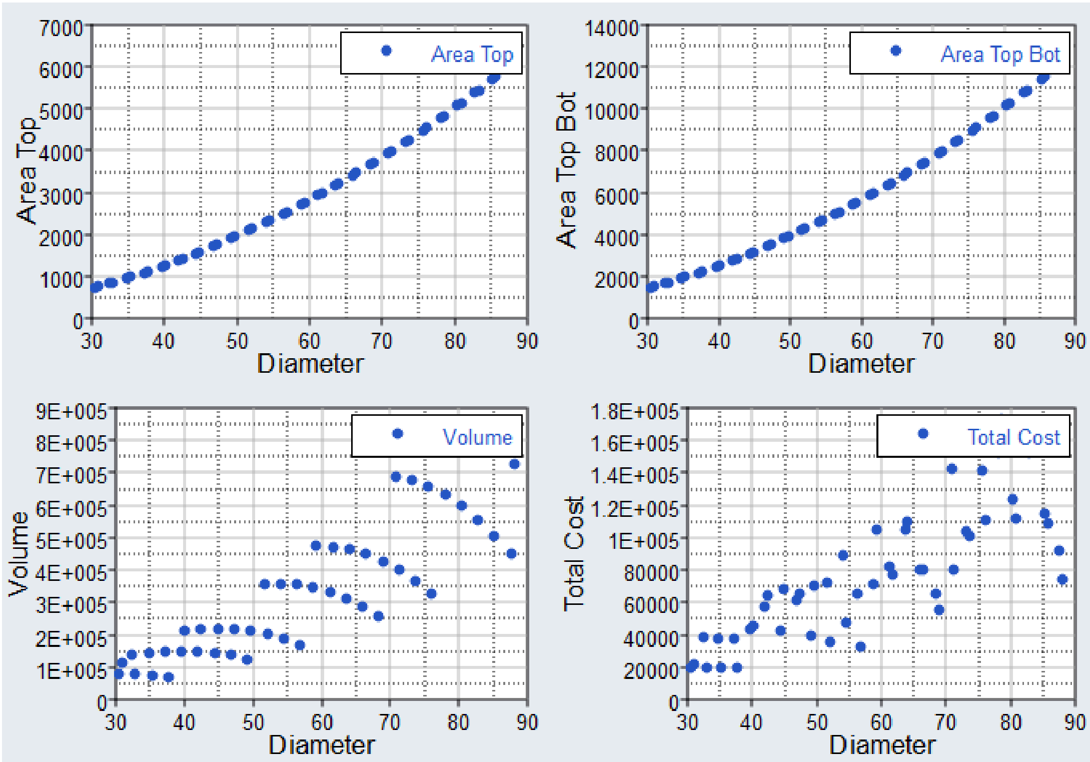


Figure 113:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

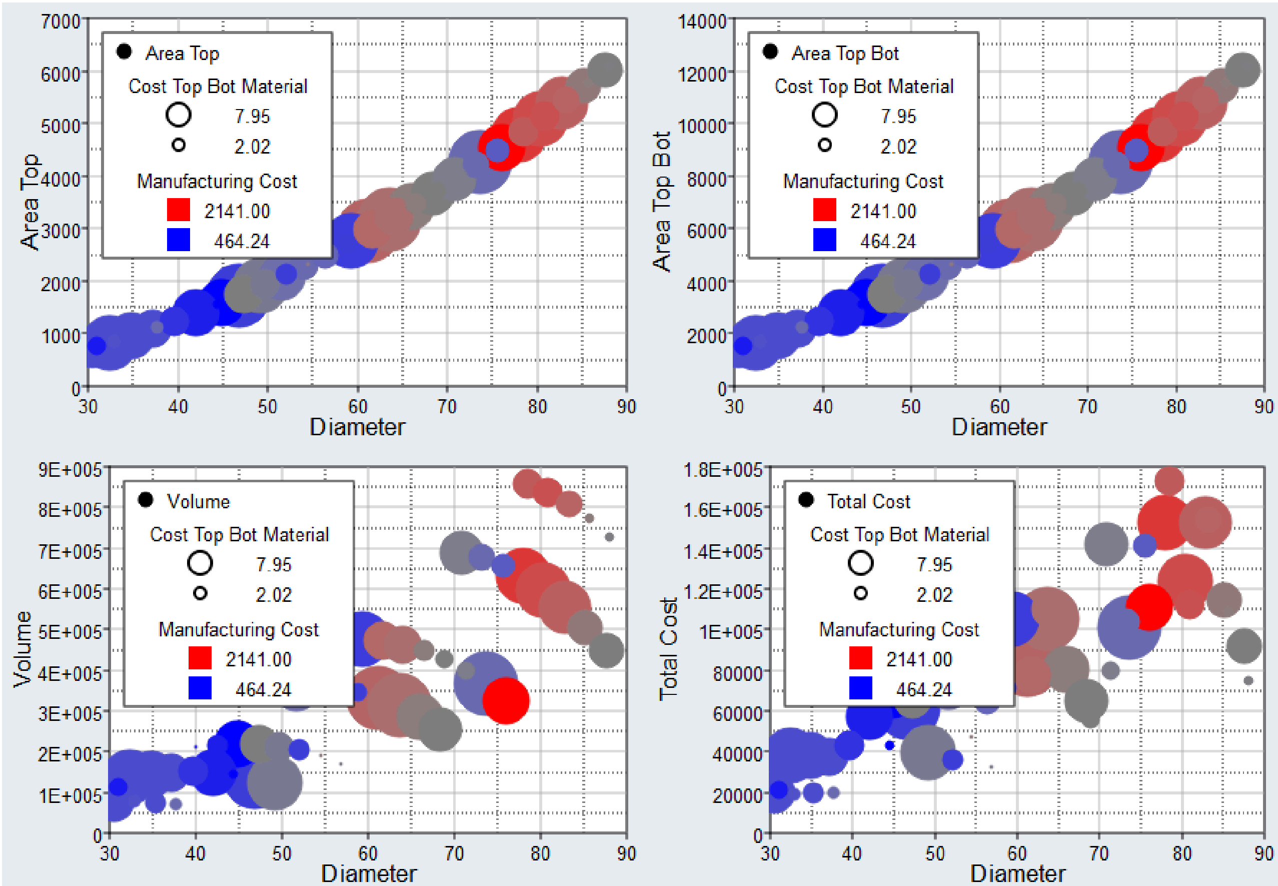


Figure 114:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation Scatter Tab Settings

Settings to configure the plots displayed in the Evaluation Scatter tab.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.

Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Post Processing

View the computational results from the Fit.

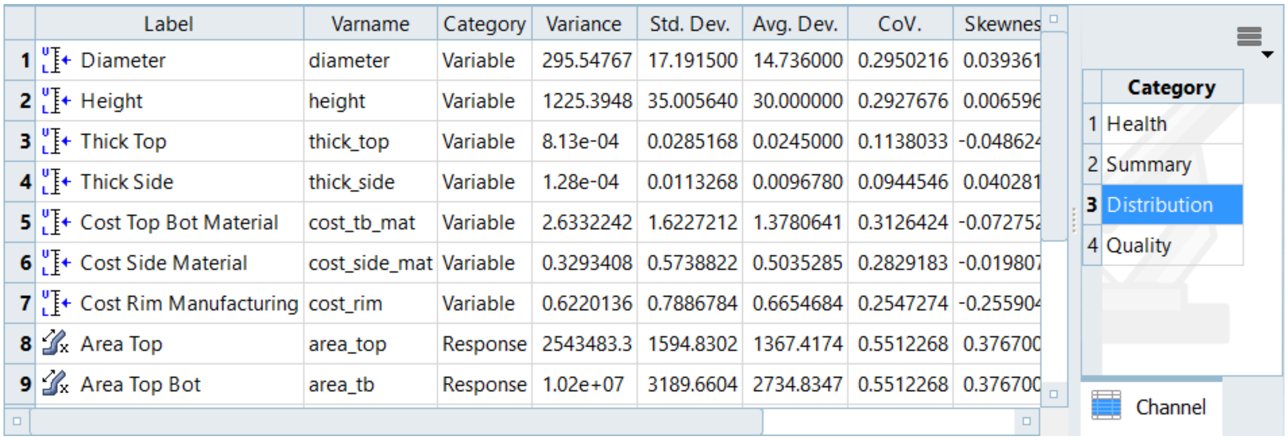
Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
    - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
    - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.
    - **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
    - **Quality** Values typically used in Quality Engineering.



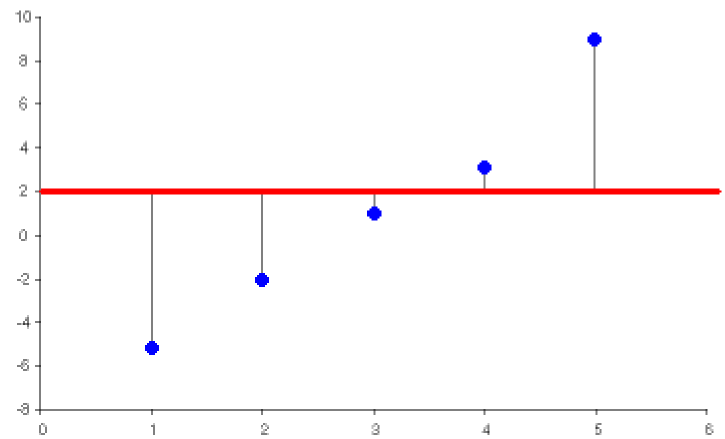
	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.072752
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.019807
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.255904
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

Figure 115:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	Description
Avg Dev (Average Deviation)	Average deviation is evaluated using: $$\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$$ In Figure 116, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 116:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $CoV = \frac{Standard\ Deviation}{Mean}$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $\sqrt{\frac{\sum x_i^2}{N}}$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}$

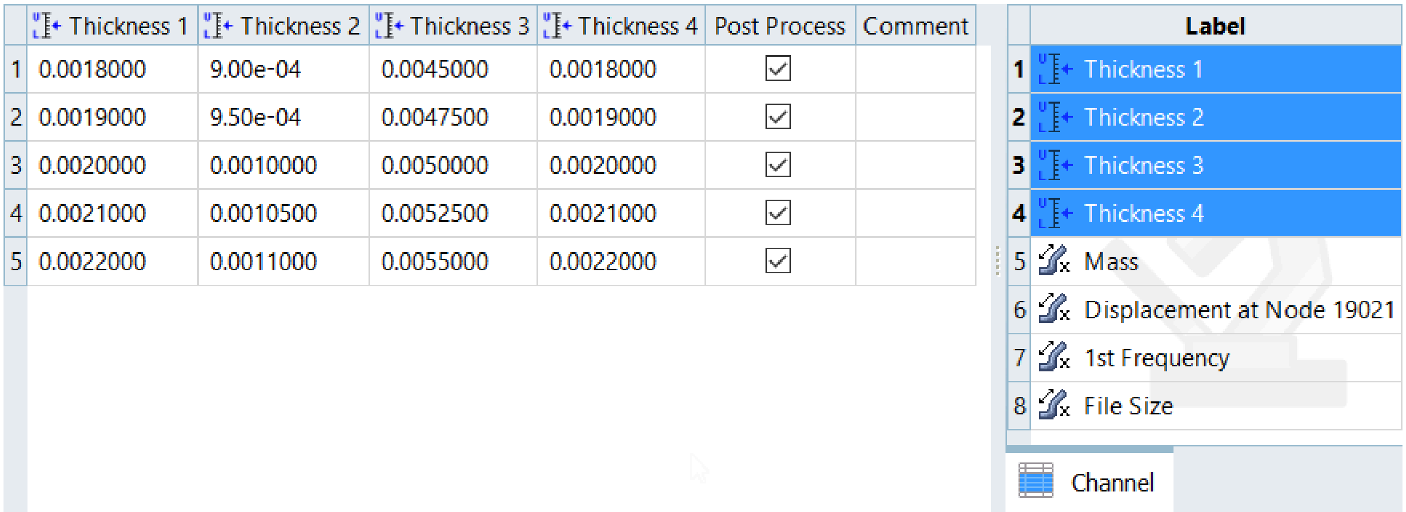
Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processing tab.

1. From the Post Processing step, click the **Summary** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Figure 117:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

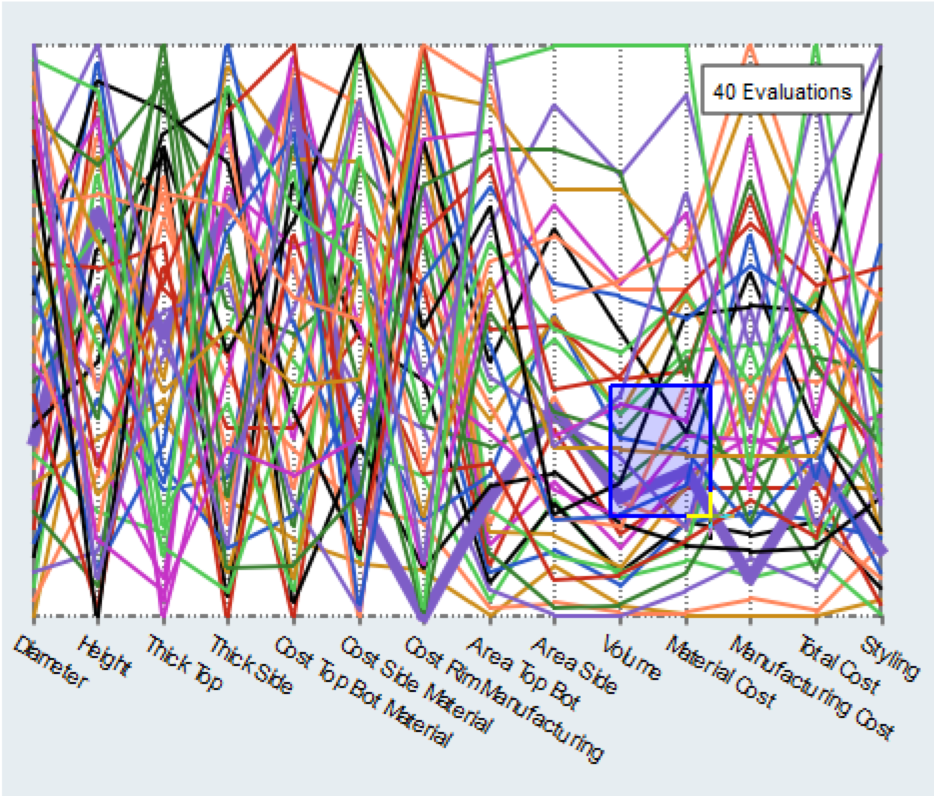
A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot.

Option	Description
	Note: Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 118 run data was selected for a set of lines. In Figure 119, you can see that when Styling is low, Height is high.

Option	Description
	 *Figure 118:*

Option	Description

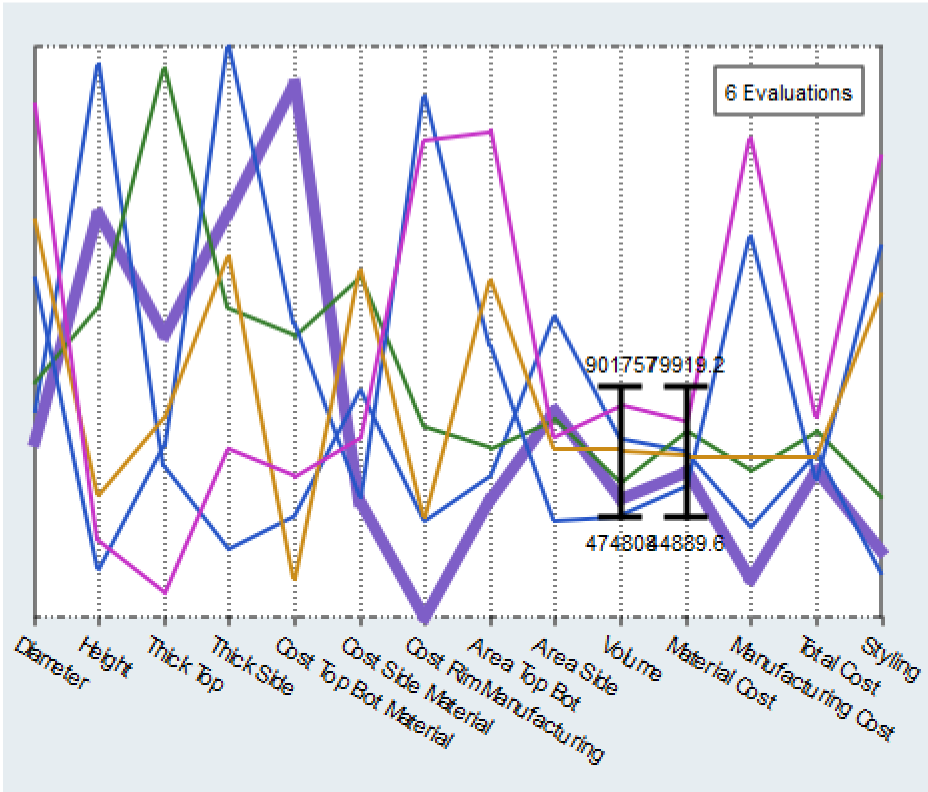


Figure 119:

Option	Description
Filter run data	Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog. • From the Filter column, select the input variables and output responses to plot. • From the Filter Min and Filter Max columns, enter values to filter. The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

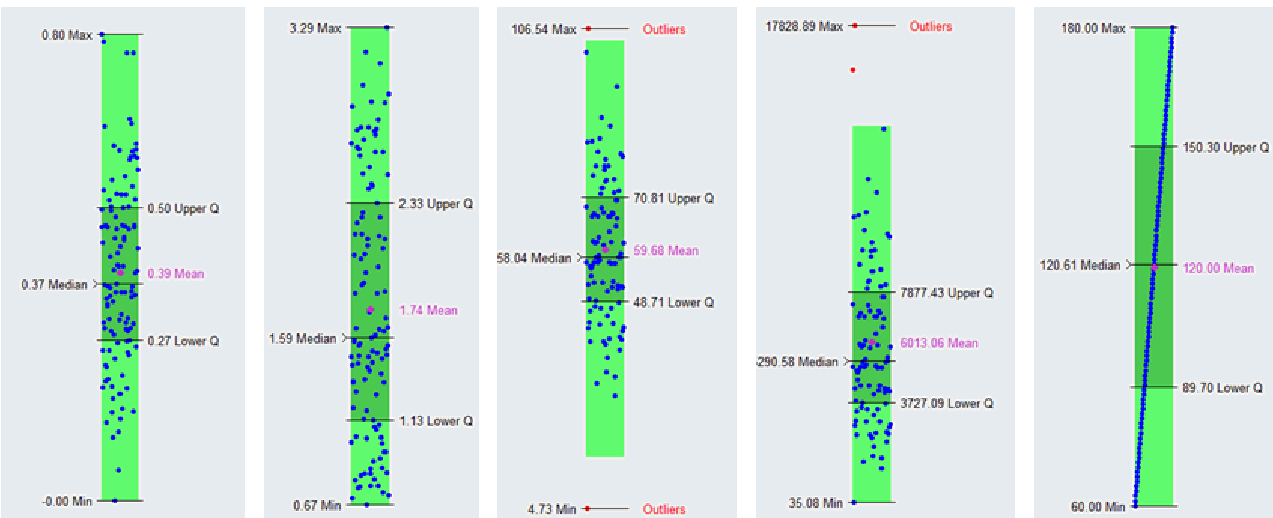


Figure 120:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 121 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

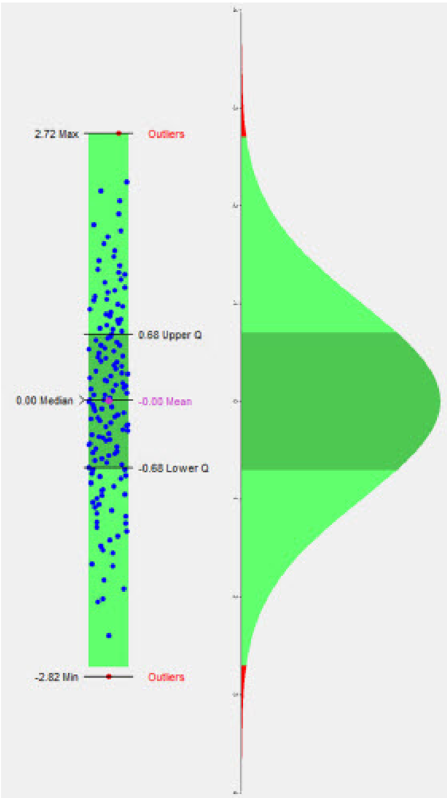


Figure 121:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

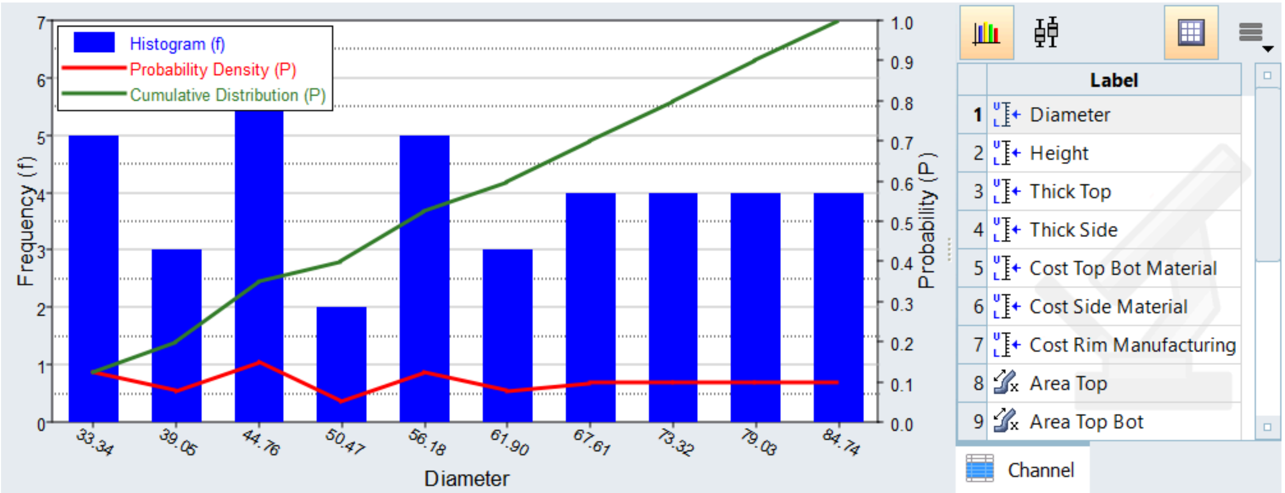


Figure 122:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processing tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Post Processing step, click the **Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

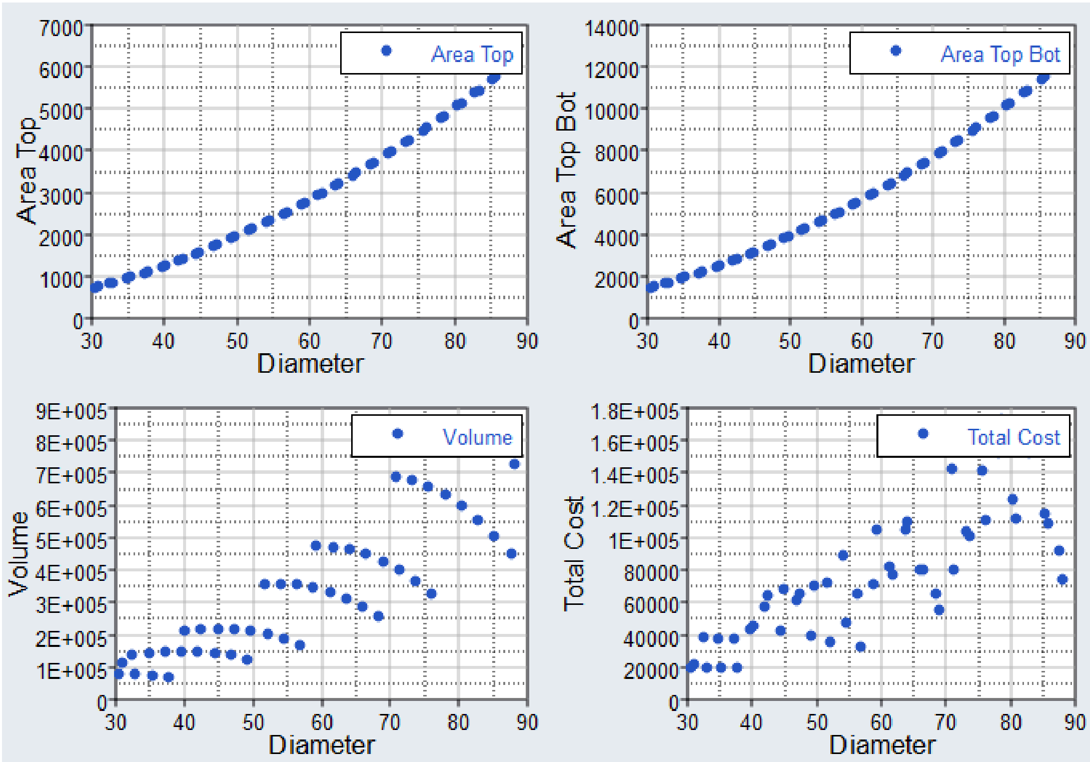


Figure 123:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

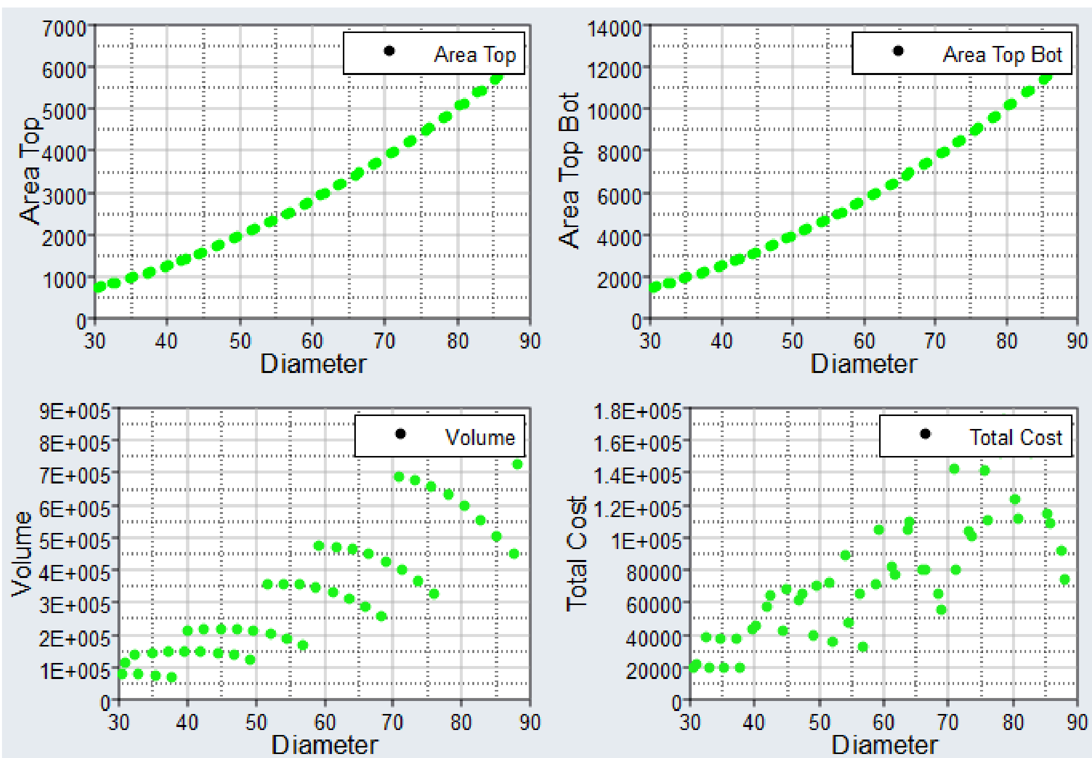


Figure 124:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

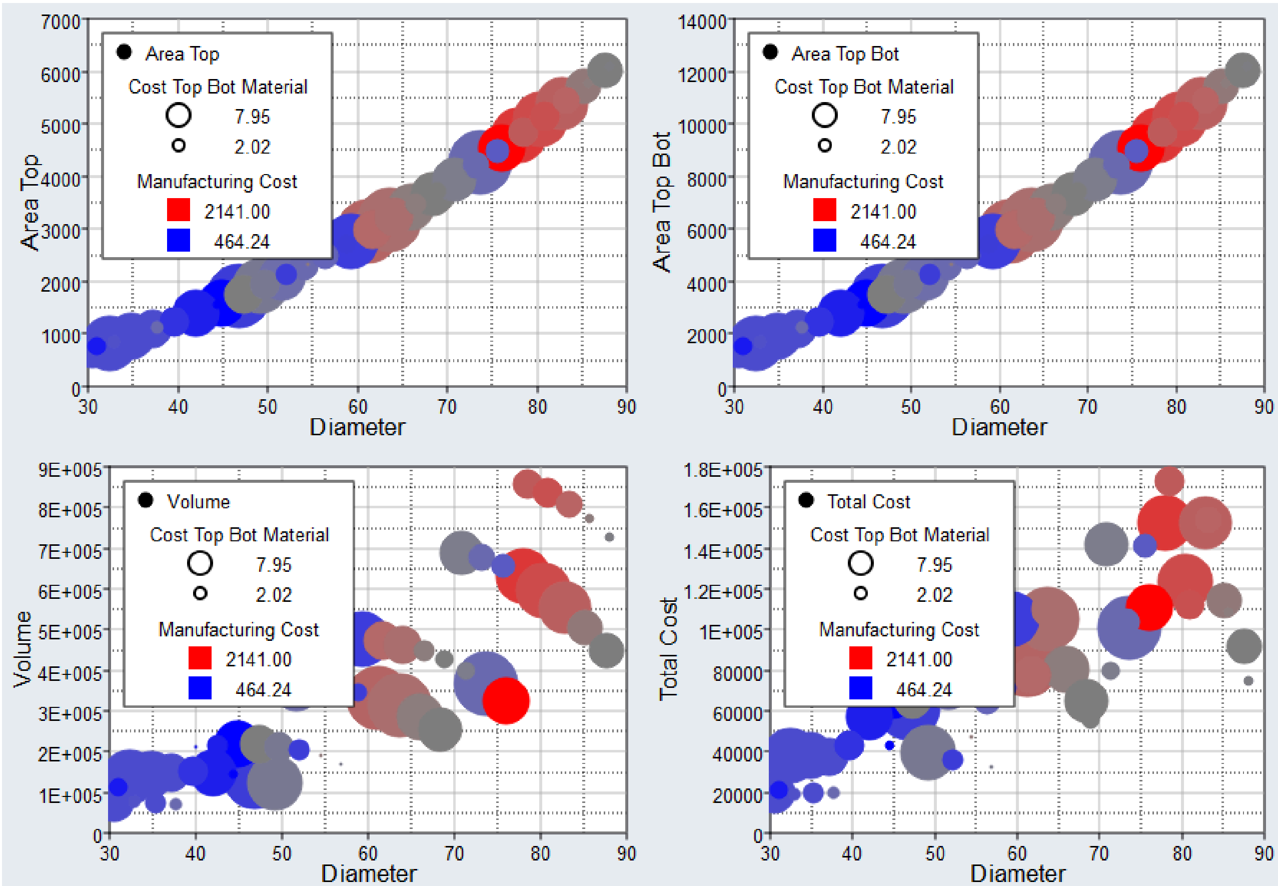


Figure 125:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Scatter Tab Settings

Settings to configure the plots displayed in the Scatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds

Display the X bounds in the plot.

Y-Bounds

Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
: Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
: Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 126: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 127: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥

Show only the column/rows with cells over the specified threshold.

Show Variables and Responses

Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

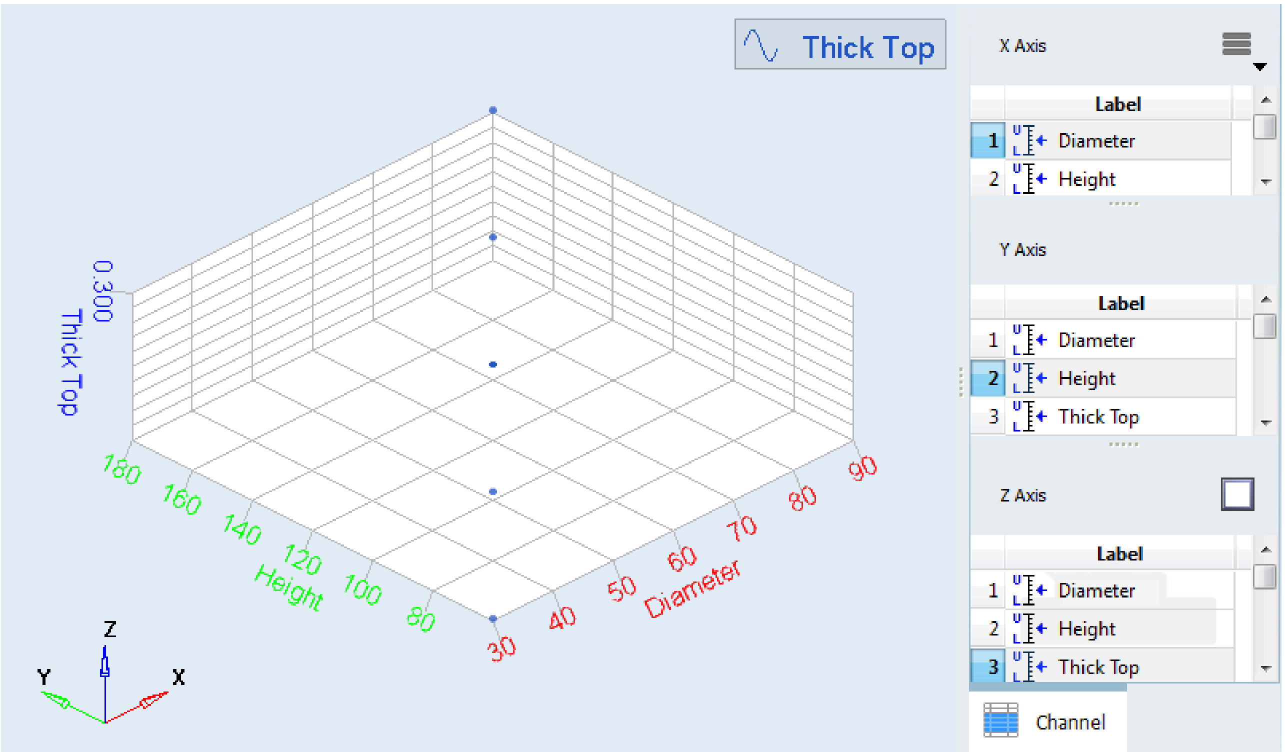


Figure 128:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

> **Tip:** For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the biplot.

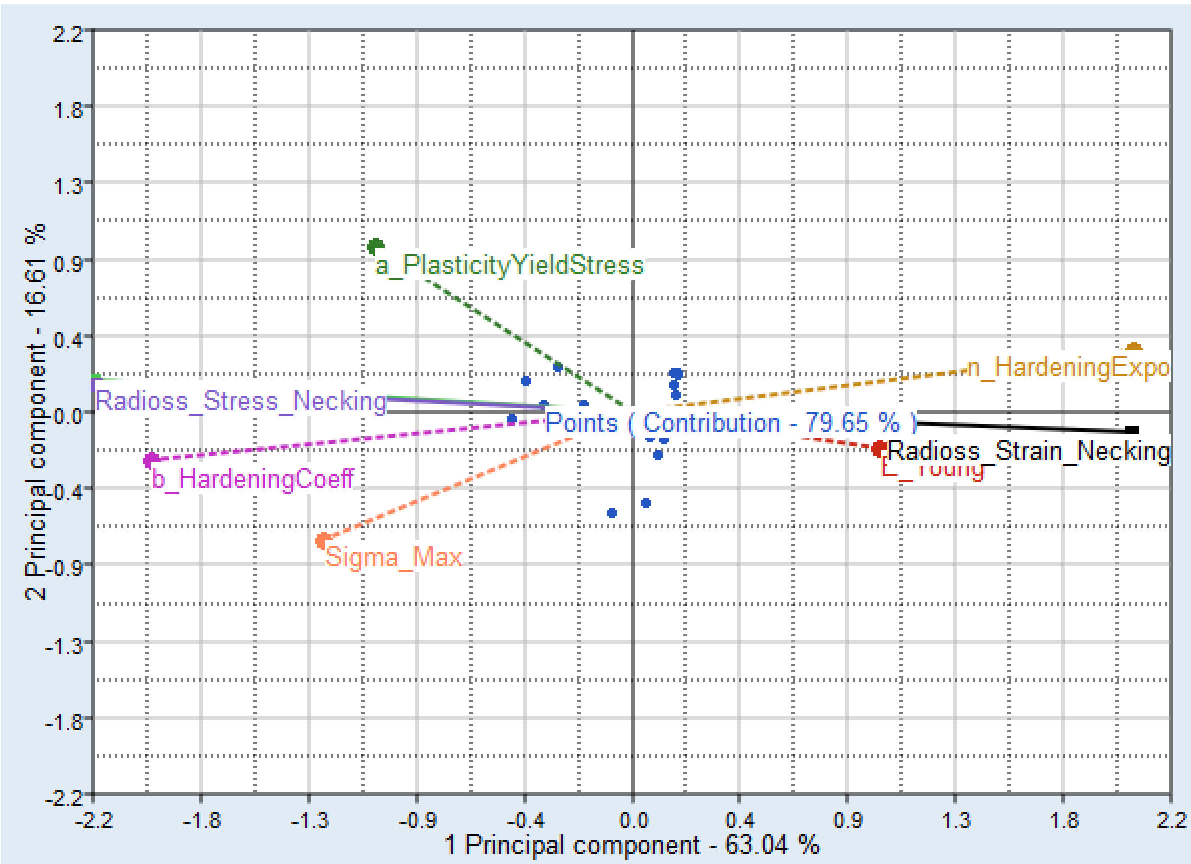


Figure 129:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
   a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

   > **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

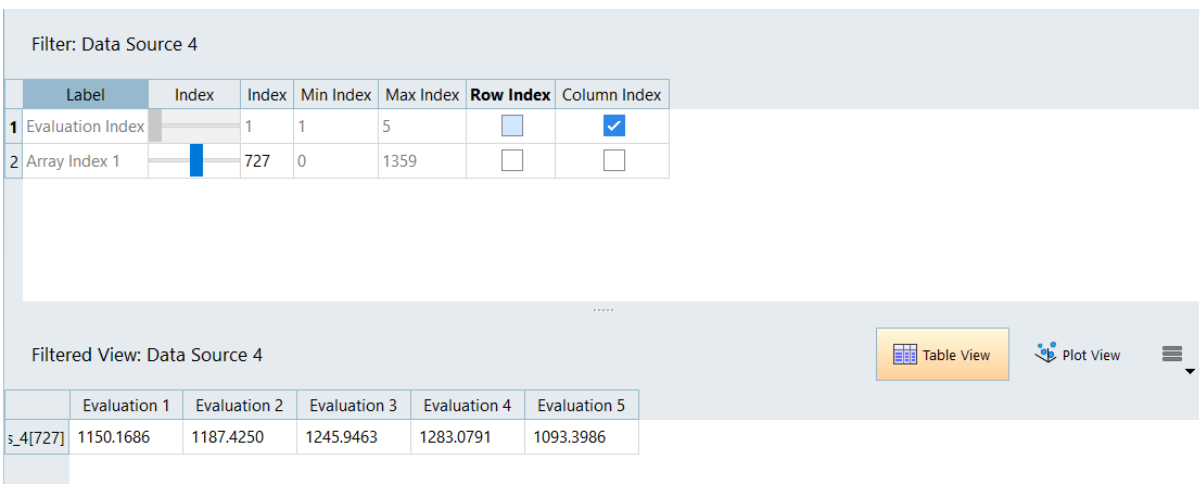


Figure 130:

5. Analyze the table.

Gradient Post Processing

Visualize gradients using vectors.

Analyze Vector

Analyze the vector in a gradient plot from the Gradient tab. Representing gradients as a vector field is an effective way to see gradients in space.

1. From the Post-Processing step, click the **Gradient** tab.

2. Use the Inputs and Output tabs of the Channel selector to select three dimensions of data to plot.
3. Analyze the direction and intensity of the vector created using the selected data sets.

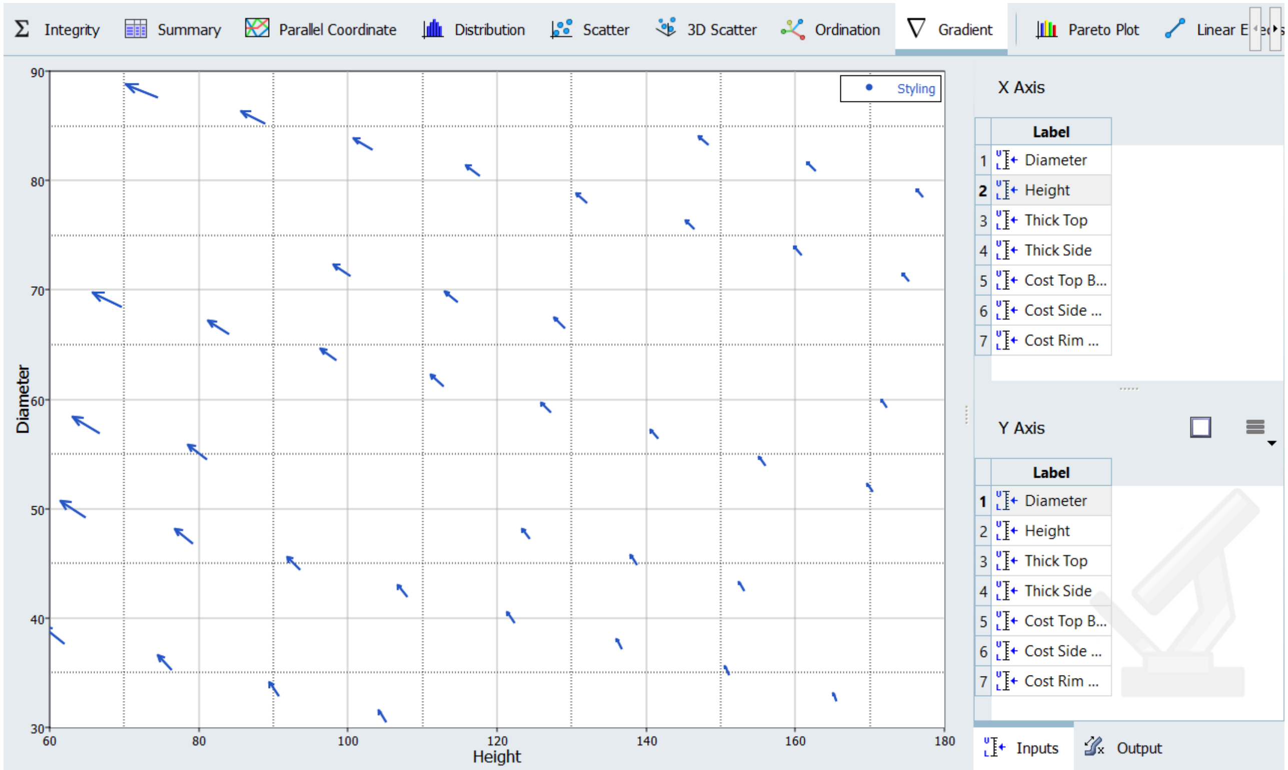


Figure 131:

Gradient Tab Settings

Settings to configure the plots displayed in the Gradient tab.

In the Gradient tab, use the Channel selector to select data to display in the gradient plot.

Channel Settings

Inputs	X-Bounds: display the X bounds in the plot. Y-Bounds: display the Y bounds in the plot.
Output	Gradient: display the vector in the plot.

Diagnostics Post Processing

Analyze the response surface quality.

Analyze the Predictive Model Quality

Analyze the Fit quality.

1. From the Post Processing step, click the **Diagnostics** tab.
2. In the work area, select the output response to analyze.
3. Click the tabs, below the output responses, to change the diagnostics used to analyze the selected output response.
    - **Detailed Diagnostics** displays diagnostic information for the Input matrix, Cross-Validation matrix, and Testing Matrix.
    - **Regression Terms** displays the confidence intervals which consist of an upper and lower bound on the coefficients of the regression equation.

      Bounds represent the confidence that the true value of the coefficient lies within the bounds, based on the given sample.

      Change the confidence value from the % Confidence settings. A higher confidence value will result in wider bounds; a 95% confidence interval is typically used.

      > **Note:** Only available for Least Squares Regression.

    - **Regression Equation** displays the complete formula for the predictive model as a function of the input variables.

      > **Note:** Only available for Least Squares Regression.

    - **ANOVA** estimates the error variance and determines the relative importance of various factors.

      Often used to identify which variables are explaining the variance in the data. This is done by examining the resulting increase in the unexplained error when variables are removed.

      > **Note:** Only available for Least Squares Regression.

    - **Confusion Matrix** summarizes the performance of a classifier. Correctly identified data is listed on the diagonal, and misclassifications are presented on the off-diagonals.

      > **Tip:** Click ≡ to toggle the confusion display from absolute count to percentages. Also, click ≡ to control the display of the confusion matrix between the input, cross-validation, and testing data set.

Configure the Diagnostics tab display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Diagnostic Tab Settings.

Diagnostic Tab Settings

Settings to configure the diagnostics displayed in the Diagnostic post processing tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the pane that displays the different types of diagnostics).

% Confidence	Change the confidence value. **Note:** Only available for Regression Terms diagnostics.

Diagnostic Definitions

Definitions used to describe diagnostic concepts.

For a given set of n input values, denoted as y_i, the Fit predictions at the same points are denoted as $\bar{y}_i$. The mean of the input values is expressed $\bar{y}$. For a Least Squares Regression, p is the number of unknown coefficients in the regression.

The following values are defined as follows:

Total Sum of Squares

$$SS_{tot} = \sum_{i=1}^{n} \left[y_i - \bar{y} \right]^2$$

Explained Sum of Squares

$$SS_{\exp} = \sum_{i=1}^{n} \left[\bar{y}_i - \bar{y} \right]^2$$

Residual Sum of Squares

$$SS_{eer} = \sum_{i=1}^{n} \left[y_i - \bar{y}_i \right]^2$$

Average Absolute Error

$$\frac{1}{n} \sum_{i=1}^{n} \left[abs(y_i - \bar{y}_i) \right]$$

Standard Deviation

$$\sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} \left[y_i - \bar{y}_i \right]^2 \right)}$$

Detailed Diagnostic

Data displayed in the Detailed Diagnostic tab of the Diagnostics post process tool.

Input Matrix

The Input Matrix column shows the diagnostic information using only the input matrix. For methods which go through the data points, such as HyperKriging or Radial Basis Functions, input matrix diagnostics are not useful.

Cross-Validation Matrix

The Cross-Validation Matrix column shows the diagnostic information using a k-fold scheme, which means input data is broken into k groups. For each group, the group's data is used as a validation set for a new approximate model using only the other k-1 group's data. This allows for diagnostic information without the need of a testing matrix.

Testing Matrix

The Testing Matrix column compares the approximate model, which was built using the input matrix, against a separate set of user supplied points. Using a Testing matrix is the best method to get accurate diagnostic information.

Criterion

R-Sqaure

Commonly called the coefficient of determination, is a measure of how well the Fit can reproduce known data points. Graphically, this can be visualized by scatter plotting the known values versus the predicted values. If the model perfectly predicts the known values, R-Square will have its maximum possible value of 1.0, and the scatter points will lie on a perfect diagonal line, as shown in the Figure 132.

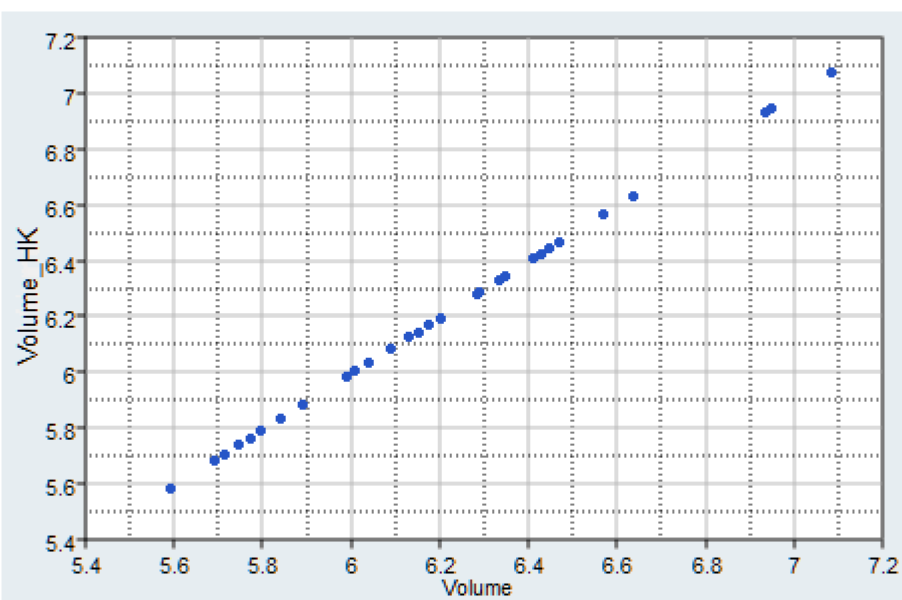


Figure 132:

More typically, the Fit introduces modeling error, and the scatter points will deviate from the straight diagonal line, as shown in the Figure 133.

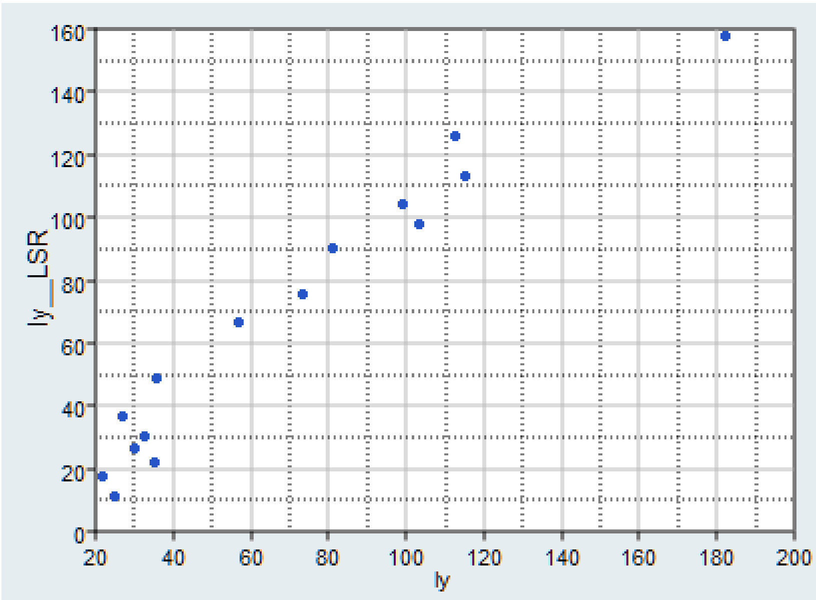


Figure 133:

The value of R-Square decreases as errors increase and the scatter plot deviates more from a straight line. The main interpretation of R-Square is that it represents the proportion of variance within the data which is explained by the Fit. For example, if R-Square = 0.84, then 84% of the variance in the data is predictable by the Fit. The higher the value of R-Square, the better the quality of the Fit. In practice, a value above 0.92 is often very good and a value lower than 0.7 necessitates investigation using other metrics. If R-Square is 1.0, you should be skeptical of this result unless the data was expected to be perfectly predicted by the Fit. There are some cases in which R-Square can be negative. A negative R-Square value indicates that using the raw mean would be a better predictor than the Fit itself; the Fit is very poor quality.
In the work area, these numbers are presented with a spark line to indicate the relative value of the number (values typically vary between 0 and 1). Values are color coded based on the following:

Red	When R^2 is less than 0.65 ($R^2 < 0.65$) it is displayed red, which indicates the value is not good.
Green	When R^2 is between 0.8 and 0.995 ($0.8 < R^2 < 0.995$) it is displayed green, which indicates the value is good.
Black	Indicates that you should apply judgment when determining whether the value is or is not good.

R-Sqaure is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left[y_i - \bar{y}_i\right]^2}{\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2}$$

R-Sqaure Adjusted

Due to its formulation, adding a variable to the model will always increase R-Square. R-Square Adjusted is a modification of R-Square that adjusts for the explanatory terms in the model. Unlike R-Square, R-Square Adjusted increases only if the new term improves the model more than would be expected by chance. The adjusted R-Square can be negative, and will always be less than or equal to R-Square. If R-Square and R-Square Adjusted differ dramatically, it indicates that non-significant terms may have been included in the model.

R-Sqaure Adjusted is defined as:

$$R^2\ adjusted = 1 - \frac{n-1}{n-p-1}\left(1 - R^2\right)$$

In the work area, these numbers are presented with a spark line to indicate the relative value of the number (values typically vary between 0 and 1). Values are color coded based on the following:

Red	When R^2 adjusted is less than 0.65 (R^2 adjusted < 0.65) it is displayed red, which indicates the value is not good.
Green	When R^2 adjusted is between 0.8 and 0.995 (0.8 < R^2 adjusted < 0.995) it is displayed green, which indicates the value is good.
Black	Indicates that you should apply judgment when determining whether the value is or is not good.

Multiple R

The multiple correlation coefficient between actual and predicted values, and in most cases it is the square root of R-Square. It is an indication of the relationship between two variables.

Note: Only available for Least Squares Regression.

Relative Average Absolute Error

The ratio of the average absolute error to the standard deviation. A low ratio is more desirable as it indicates that the variance in the Fit's predicted value are dominated by the actual variance in the data and not by modeling error.

Relative Average Absolute Error is defined as:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left[abs(y_i - \bar{y}_i)\right]}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2\right)}}$$

Maximum Absolute Error

The maximum difference, in absolute value, between the observed and predicted values. For the input and validation matrices, this value can also be observed in the Residuals tab.

Maximum Absolute Error is defined as:

$$\max\left(abs\left(y_i - \bar{y}_i\right)\right)$$

Root Mean Square Error

A measure of weighted average error. A higher quality Fit will have a lower value.

Root Mean Square Error is defined as:

$$\sqrt{\frac{\sum_{i=1}^{n}\left[y_i - \bar{y}_i\right]^2}{n}}$$

Number of Samples

The number of data points used in the diagnostic computations.

Regression Terms

Data displayed in the Regression Terms tab of the Diagnostics post process tool.

t-value is defined as:

$$t_j = \frac{\beta_j}{\sqrt{\sigma^2 c_{jj}}}$$

where β_j is the corresponding regression coefficient (the Values column) and SE is the standard error. The standard error is defined as:

$$SE = \sqrt{\sigma^2 c_{jj}}$$

and

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left[y_i - \bar{y}_i\right]^2}{n - p}$$

where c_{jj} is the diagonal coefficient of the information matrix used during the regression calculation.

p-values are computed using the standard error and t-value to perform a student's t-test. The p-value indicates the statistical probability that the quantity in the Value column could have resulted from a random sample and that the real value of the coefficient is actually zero (the null hypothesis). A low value, typically less than 0.05, leads to a rejection of the null-hypothesis, meaning the term is statistically significant.

ANOVA

Data displayed in the ANOVA (Analysis of Variance) tab of the Diagnostics post process tool.

Degrees of Freedom

Number of terms in the regression associated with the variable. All degrees of freedom not associated with a variable are retained in the Error assessment. More degrees of freedom associated with the error increases the statistical certainty of the results: the p-values. Higher order terms have more degrees of freedom; for example a second order polynomial will have two degrees of freedom for a variable: one for both the linear and quadratic terms.

Sum of Squares

For each variable, the quantity shown is the increase in unexplained variance if the variable were to be removed from the regression. A variable which has a small value is less critical in explaining the data variance than a variable which has a larger value.

The row Error, represents the variance not explained by the model, which is SS_{err}.

The row Total, which is SS_{tot}, will generally not equal to the sum of the others rows.

Mean Squares

The ratio between unexplained error increase and degrees of freedom, computed as the Sum of Squares divided by the associated degrees of freedom.

Mean Squares Percent

Interpreted as the relative contribution of the variables to the Fit quality, computed as the ratio of the Mean Square to the summed total of the Mean Squares. A variable with a higher percentage is more critical to explaining the variance in the given data than a variable with a lower percentage.

F-value

The quotient of the mean squares from the variable to the mean squares from the error. This is a relative measure of the variable's explanatory variance to overall unexplained variance.

p-value

The result of an F-test on the corresponding F-value. The p-value indicates the statistical probability that the same pattern of relative variable importance could have resulted from a random sample and that the variable actually has no effect at all (the null hypothesis). A low value, typically less than 0.05, leads to a rejection of the null-hypothesis, meaning the variable is statistically significant.

Residuals Post Processing

Identify design errors.

Identify Design Errors

Identify design errors in the Residuals post processing tab.

1. From the Post Processing step, click the **Residuals** tab.
2. From the Channel selector, select one output and one or more inputs to investigate.

3. Optional: Switch the view from residuals table to residuals plot by clicking , located next to the Channel selector.

> **Note:** The plot shows the residuals versus the original output response. This plot is useful to visually assess the performance of a Fit. It is desirable to not have any visual pattern to the residuals; unbiased and homoscedatic residuals appear similar to randomness.

4. Click ≡ (located above the Matrix Residuals table) to select the type of residuals displayed in the table.

By default, Input Matrix residuals are displayed.

The error (and percentage) between the original output response and the approximation is listed for each run of the Input, Cross-Validation, or Testing matrices.

> **CAUTION:**
>
> View the Percent Error column with caution when the values of the output response approach zero. In this situation, the Percent Error can be very high and potentially misleading.

	Cost Side Material	Cost Rim Manufacturing	Area Top	Area Top	Error	Percent Error
1	2.7614467	2.2016189	1385.4424	1385.4424	-9.09e-13	-6.56e-14
2	2.5228935	2.9032379	2290.2210	2290.2210	4.55e-13	1.99e-14
3	2.1054261	3.4036476	3421.1944	3421.1944	0.0000000	0.0000000
4	2.5527131	3.9518238	4778.3624	4778.3624	0.0000000	0.0000000
5	1.5192770	3.8582212	824.47958	824.47958	1.36e-12	1.65e-13
6	2.5948835	1.7132828	1548.3025	1548.3025	-1.59e-12	-1.03e-13
7	1.6704901	2.5683444	2498.3201	2498.3201	9.09e-13	3.64e-14
8	1.2530227	3.0687541	3674.5324	3674.5324	4.55e-13	1.24e-14

Figure 134: Cross-Validation Matrix Residuals

> **Tip:** Search for specific cases using the Find and Sort options, which can be accessed from the context menu that opens when you right-click in the work area.

Configure the Residuals tab display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Residuals Tab Settings.

Residuals Tab Settings

Settings to configure the results displayed in the Residuals post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Matrix Residuals table).

Input Matrix Residuals	Display Input matrix residuals.
Cross-Validation Matrix Residuals	Display Cross-Validation matrix residuals.
Testing Matrix Residuals	Display Testing matrix residuals.

Trade Off Post Processing

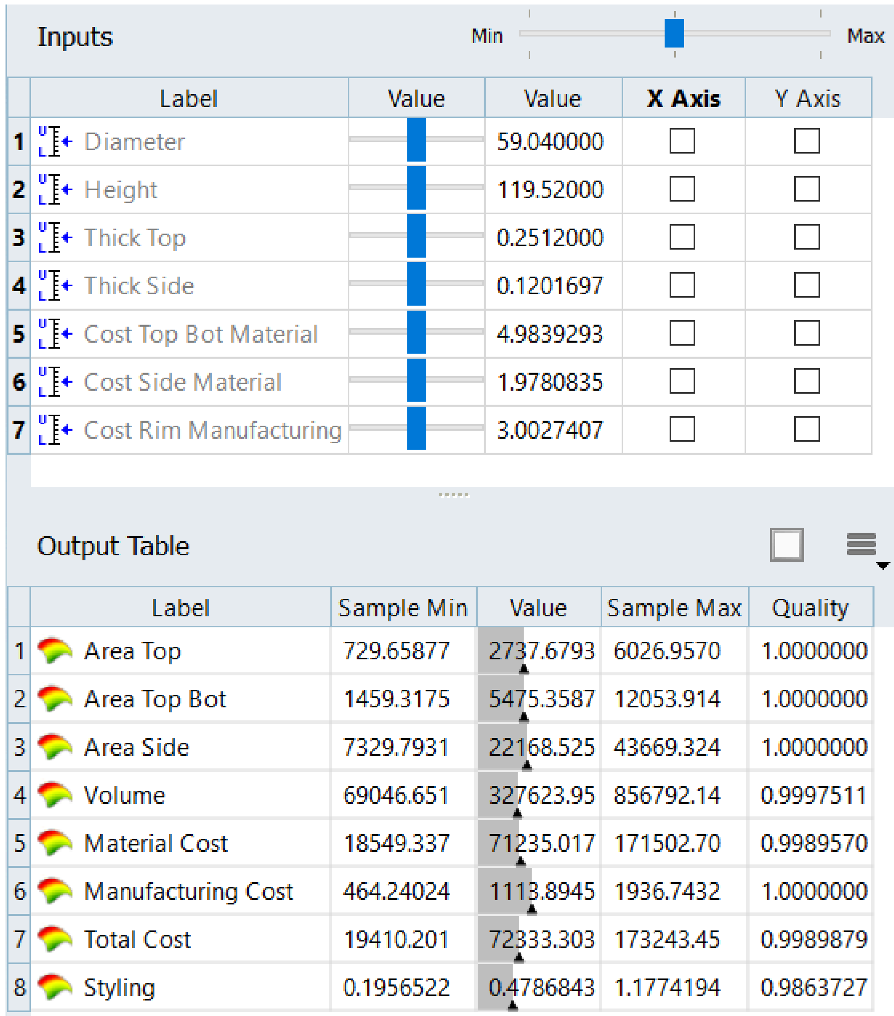
Perform "What If" scenarios.

Perform "What If" Scenarios

Perform "What If" scenarios with interactive response surface tools in the Trade-Off post process tab.

1. From the Post Processing step, click the **Trade-Off** tab.
2. From the Channel selector, select the output response(s) to analyze in the Output Table.
3. Analyze the effect on inputs vs. outputs.

Option	Description
Modify the values of input variables to see their effect on output response approximations	In the Inputs pane, change each input variable by moving the slider in the first Value column, or by entering a value into the second Value column. Set input variables to their initial, minimum, or maximum values by moving the slider in the upper right-hand corner of the Inputs frame.

Option	**Description**
	 *Figure 135:*
Plot the effect of input variables on output response approximations	In the Inputs pane, select an input variable to plot by selecting its corresponding X Axis and/or Y Axis checkbox. • Create a 2D trade-off by enabling the X Axis checkbox.

Option	Description

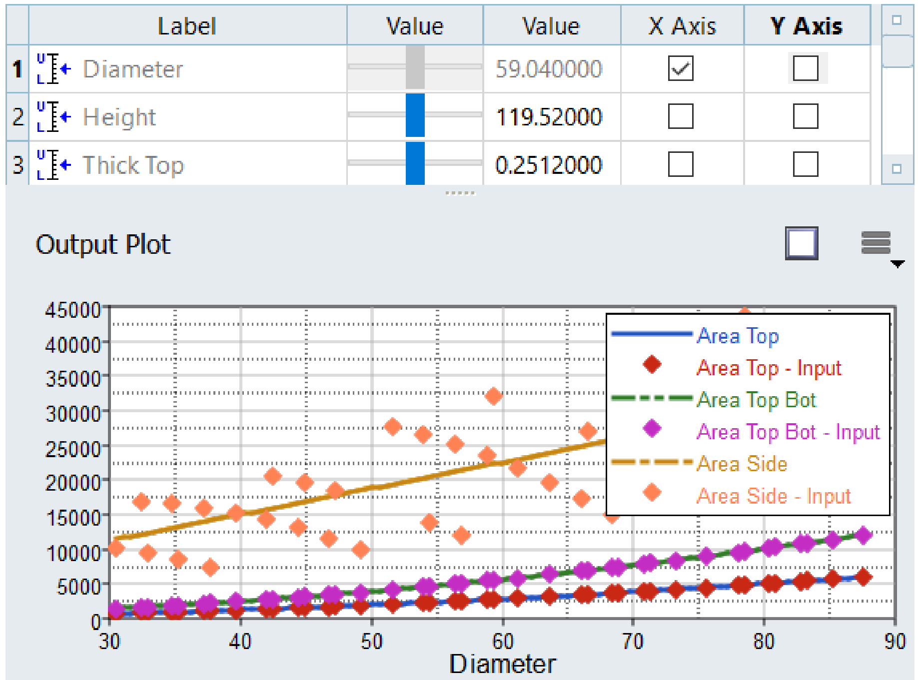


Figure 136:

- Create a 3D trade off by enabling X Axis and Y Axis checkboxes.

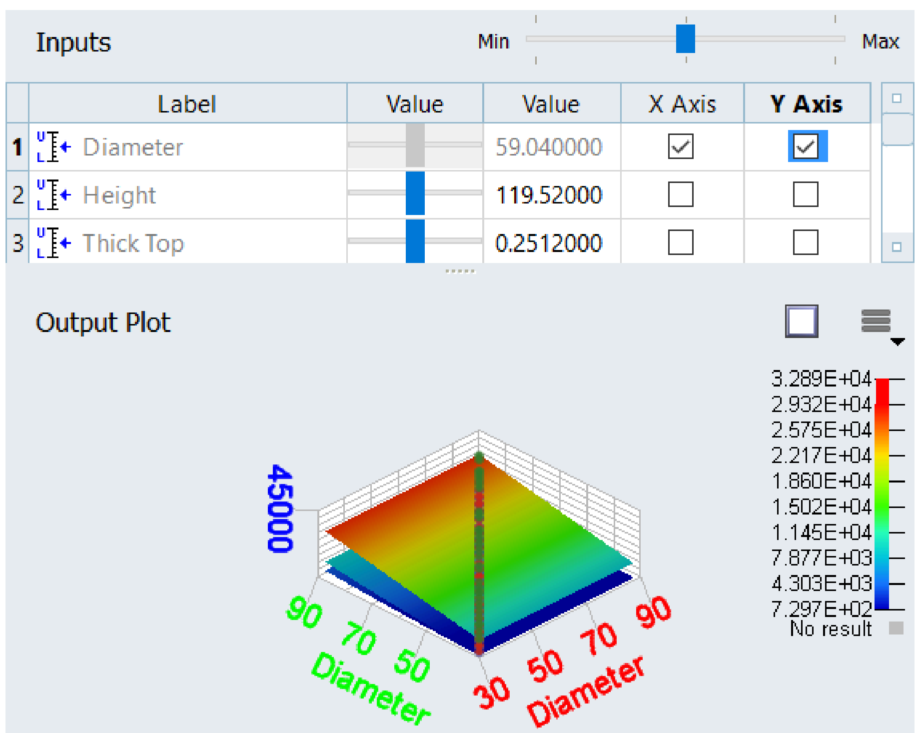


Figure 137:

Option	Description
	The output responses selected with the Channel selector are plotted. The values for the input variables which are not plotted can be modified by moving the sliders in the Value column to modify the other input variables, while studying the output response throughout the design space.

For the given values of the input variables, the output responses' predictions are calculated by the Fit, and displayed in the Output Plot pane. Table shading is used to indicate the output response's value between the minimum and maximum values contained in the input design matrix. When shading extends into either the Sample Min or Sample Max column, this indicates that the predicted value is beyond the bounds contained in the input matrix. If the shading extends significantly into these regions, it is suggested that you asses the validity of this value based on experience and knowledge of the modeled problem.

The Quality column is provided as a measure to assess both the accuracy and trust in the Fit at a specific point in the design space. Both global and local metrics are combined to create a metric that runs between 0 and an Upper Bound limited by the Fit's R^2 value. The quality will be highest at points inside the convex hull formed around the Fit's input points, where the predictive model has been trained to explain variance in the data. The quality metric decreases proportional to the distance outside the convex hull as the predictions at these points becomes less reliable, partially due to the values increasing based on an extrapolation of the data.

> **Tip:** In a 2D trade-off the metrics shown in the Quality column can be plotted alongside the output response curve by selecting **Fit Quality** from the menu that displays when you click ≡ (located in the Output Plot pane).

Configure the Trade Off tab's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Trade-Off Tab Settings.

Trade-Off Tab Settings

Settings to configure the results displayed in the Trade-Off post processing tab.

Access settings from the menu that displays when you click ≡ (located in the Output Plot pane).

Fit	Display the predicated curve (2D plot) or surface (3D plot) of the Fit.
Fit Quality	Plot the estimated quality (value shown in the Quality column) alongside the output response curve.
Input Matrix	Display the scatter points of the Input matrix.
Testing Matrix	Display the scatter points of the Testing matrix.

# samples	Change the number of discretized points used when drawing the trade-off (2D plot) or surface (3D plot). **Note:** Increasing this number will result in a smoother representation, which could be at the cost of interface responsiveness.
Discrete Surface Contour	Display a discrete color profile of the surface. Disable this checkbox to display a blended color profile of the surface (3D plot).
Mesh lines	Display a visual projection of the samples' mesh grid lines onto the surface (3D plot).

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	Description
HyperStudy Data	Generates a data report (`*.data`).
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel. In the Excel spreadsheet, click the **Trade-Off** sheet to perform trade-off studies in Excel, independent of HyperStudy. In a Fit Excel report, the Trade-Off 1D tab is a reflection of the corresponding Trade-Off 1D tab within HyperStudy. From the Trade-Off 1D tab in Excel, you can adjust the input variable values on the right-hand side to change the predicted output responses values displayed on the left-hand side.

Report Type	Description
	Restriction: To use this feature, you must add the HstAddinFit add-in to Excel. To install the Excel Plug-in for HyperStudy Fit engines, go to the `<ALTAIR_HOME>\hst\plugins\externals\hstfitaddin\` directory and double-click `hstfitaddin_install.vbs`. The Excel Plug-in requires access to a valid HyperWorks license for all features to work properly.

3. Click **Create Report**.

4.2.3 Setup Optimization Studies

An Optimization is a mathematical procedure used to determine the best design for a set of given constraints, by changing the input variables in an automatic manner.

Add a Optimization Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
   a) In the Label field, enter a name for the Optimization.
   b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.
      By default, the Definition defined in the study Setup is selected.
   c) Under Select Type, select **Optimization**.
   d) Click **OK**.

A new Optimization is added to the Explorer.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.
3. Test Models.

4. Define Output Responses.
5. Review definitions in the following ways:

To: **Do this:**

Review status

Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

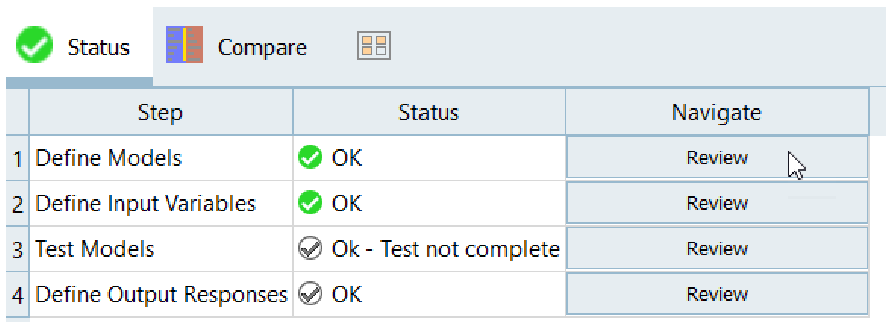


Figure 138:

Compare definitions

Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

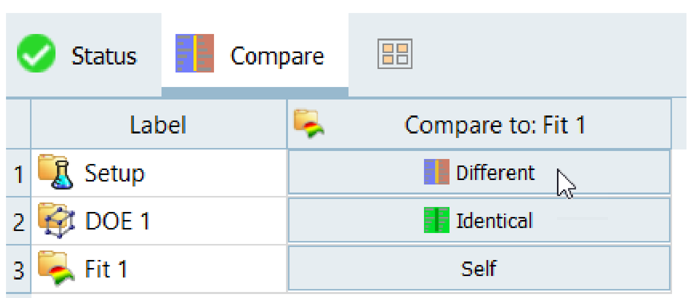
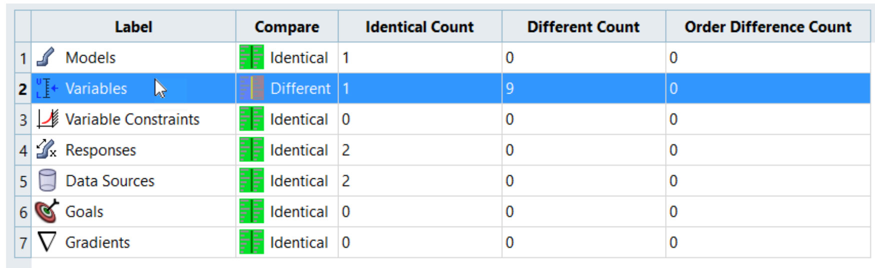


Figure 139:

The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.

To:	**Do this:**
	4. Click a channel to display a detailed comparison.

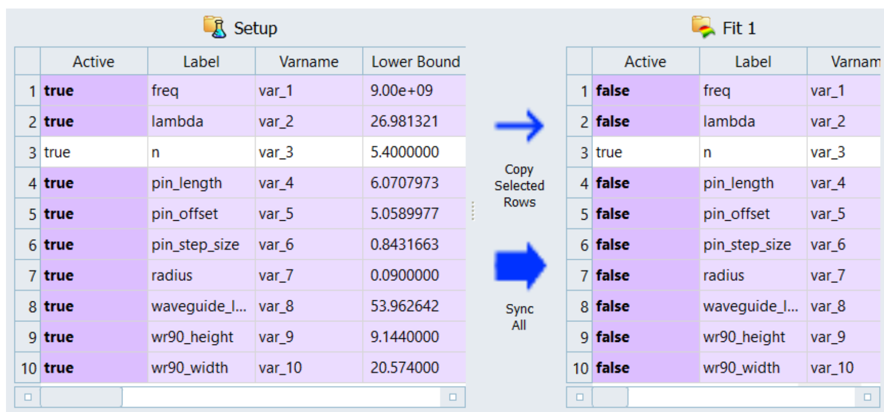


	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 140:

5. Sync data.

- Click **Copy Selected Rows** to sync the single row.
- Click **Sync All** to sync all rows.



	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 141:

Select a Numerical Method

Select a numerical method to use when evaluating the Optimization.

1. In the Specifications step, Mode column, select a numerical method.
2. In the Settings and More tabs, change settings as needed.
3. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog. For more information, see Edit the Run Matrix.

Optimization Methods

Numerical methods available for an Optimization approach.

Constraint violation tolerance and constraint threshold are set in the Objectives/Constraints - Goals tab of the Define Output Responses step within the Definition. For more information, visit Constraints.

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributio Role	# of Objectives	Exploratio Type	Accuracy	Efficiency	Comments
Adaptive Response Surface Method (ARSM)	None	None	Determinis	Single	Local	#	###	Default method for single objective problems.
ARSM-Based Sequential Optimizatio and Reliability Assessmen (SORA_ARS	Continuous only	Input variable constraints are not allowed	Probabilisti	Single	Local	#	###	More efficient than Sequential Optimization and Reliability Assessment, but not as accurate. It is not recommended to use ARSM-Based Sequential Optimization and Reliability Assessment with a Fit.
Genetic Algorithm (GA)	None	None	Determinis	Single	Global	##	#	Significantly expensive. Use Genetic Algorithm if the simulation

Method	Input Variable Model Restriction	Input Variable Constraint Restriction	Distribution Role	# of Objectives	Exploration Type	Accuracy	Efficiency	Comments
								is affordable or if you have a good Fit.
Global Response Search Method (GRSM)	None	None	Deterministic	Single or Multiple	Global	###	##	Default method for multi objective problems. Preferred method when the number of design variables is large. Optimizing can start with just a few number of points independent of the number of design variables.
Method of Feasible Directions (MFD)	Continuous only	Input variable constraints are not allowed	Deterministic	Single	Local	##	##	May work more efficiently for problems with a large number of constraints.

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributio Role	# of Objectives	Exploratio Type	Accuracy	Efficiency	Comments
Multi - Objective Genetic Algorithm (MOGA)	None	None	Deterministic	Multiple	Global	##	#	Significantly more expensive. Use Multi - Objective Genetic Algorithm if the simulation is affordable or if you have a good Fit.
Sequential Optimizatio and Reliability Assessmen (SORA)	Continuous only	Input variable constraints are not allowed	Probabilisti	Single	Local	###	#	Use if the simulation is affordable or if you have a good Fit.
Sequential Quadratic Programmi (SQP)	Continuous only	Input variable constraints are not allowed	Deterministic	Single	Local	###	##	Use if the simulation is affordable or if you have a good Fit.
System Reliability Optimizatio (SRO)	None	None	Probabilisti	Single	Global	###	##	Default method for probabilistic problems. In robust optimization, it provides the

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributio Role	# of Objectives	Exploratio Type	Accuracy	Efficiency	Comments
								trade-off between the nominal value and variance of the objective.
Xopt (User-Defined Optimizatio Engine)	None	None	Determinis	Single				

Adaptive Response Surface Method (ARSM)

Internally builds response surfaces and adaptively updates them as new evaluations are available.

The first response surface the Adaptive Response Surface Method builds is a linear regression polynomial, then it finds the optimum on this surface and validates it with the exact simulation. If the output response values from the response surface and the exact simulation are not close; the Adaptive Response Surface Method updates the surface with the new evaluation and finds the optimum in this updated surface. This loops is repeated until one of the convergence criteria is met.

Usability Characteristics

- Adaptive Response Surface Method is the default method. However, if the number of input variables is large, or if a global optima is required, then it is suggested that you use Global Response Search Method instead.
- For Revisions A-multi and B-multi, Adaptive Response Surface Method can take advantage of parallel execution. The number of runs in the iterative stages after N+1 evaluations (N is the number of variables) can be controlled by the setting parameter Points per iteration.
- It is an efficient optimization method because it utilizes response surfaces. It is recommend to use Adaptive Response Surface Method directly on a solver and not on a Fit.
- In the case of a failed run, it is possible to ignore a failed analysis or terminate an optimization. When omitting failed runs, the optimizer will back up half of a step between the failed run and the previous design.
- Adaptive Response Surface Method terminates when one of the following conditions are met:
    - One of the convergence criteria is satisfied.

- The maximum number of allowable analysis (Maximum Iterations) is reached.
- An analysis fails and the Terminate optimization option is the default (On Failed Evaluation).

- Supports input variable constraints.
- The algorithm begins with N+1 evaluations, where N is the number of design variables. When SRSM is set to Response Surface, control the number of initial evaluations using the Sample Points parameter. The number of evaluations in subsequent iterations is controlled by the Points per Iteration setting. Evaluations are created sequentially by default. Adjust the Revision setting to use muli variants to execute evaluations in parallel.

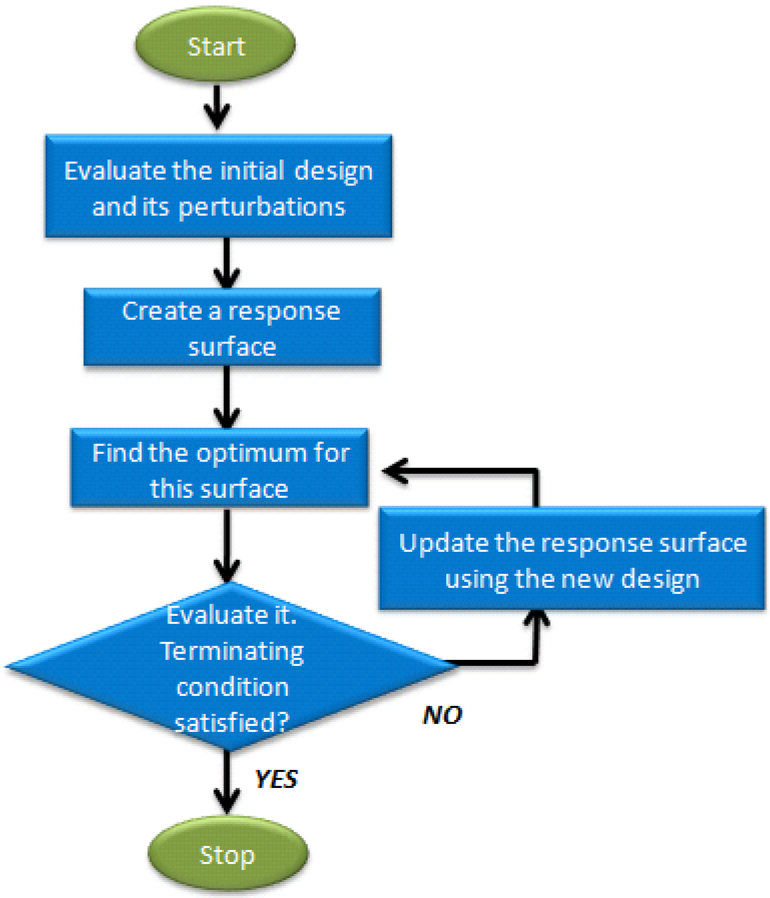


Figure 142: Adaptive Response Surface Method Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Number of Evaluations and On Failed Evaluation.

Table 7: Settings Tab

Parameter	Default	Range	Description
Number of Evaluations	25	>0	Maximum number of analyses (only for Adaptive Response Surface Method number of analysis is equal to number of iterations) allowed.
Absolute Convergence	0.001	>0.0	Determines an absolute convergence tolerance, which is constant and equal to Absolute Convergence, times the initial objective function value. The design has converged when there are two consecutive designs for which the absolute change in the objective function is less than this tolerance. There also must not be any constraint whose allowable violation is exceeded in the last design. **Note:** A larger value allows for faster convergence, but worse results could be achieved. $\begin{cases} c_{\max}^{k} \le g_{\max} \\ \lvert f^{i} - f^{i-1} \rvert < \max\left(\varepsilon \lvert f^{0} \rvert, 10^{-19}\right) \\ i = k, k-1 \end{cases}$ Where f is the objective value; $f^{0}$is the objective value of the initial design; k is the current iteration number; ε is the absolute convergence parameter; $c_{\max}$ is the maximum constraint violation; $g_{\max}$ is the allowable constraint violation.
Relative Convergence (%)	1.0	>0.0	The design has converged if the relative (percent) change in the objective function is less than this value for two consecutive designs. There also must not be any constraint whose allowable violation is exceeded in the last design.

Parameter	**Default**	**Range**	**Description**
			Note: A larger value allows for faster convergence, but worse results could be achieved. $$\begin{cases} c_{\max}^{k} \le g_{\max} \\ \frac{\lvert f^{i}-f^{i-1}\rvert}{\lvert f^{i-1}\rvert+10^{-6}} < \varepsilon \\ i = k,\ k-1 \end{cases}$$ Where, f is the objective value; k is the current iteration number; ε is the relative convergence parameter; $c_{\max}$ is the maximum constraint violation; $g_{\max}$ is the allowable constraint violation.
Design Variable Convergence	0.001	>0.0	Input variable convergence parameter. Design has converged when there are two consecutive designs for which the change in each input variable is less than both (1) Design Variable Convergence times the difference between its bounds, and (2) Design Variable Convergence times the absolute value of its initial value (simply Design Variable Convergence if its initial value is zero). There also must not be any constraint whose allowable violation is exceeded in the last design. **Note:** A larger value allows for faster convergence, but worse results could be achieved. $$\begin{cases} \lvert x_j^{i} - x_j^{i-1}\rvert < \gamma \cdot (x_j^{U} - x_j^{L}) \\ \begin{cases} \lvert x_j^{i} - x_j^{i-1}\rvert < \gamma \cdot \lvert x_j^{0}\rvert,\ \text{if}(x_j^{0} \neq 0) \\ \lvert x_j^{i} - x_j^{i-1}\rvert < \gamma,\ \text{if}(x_j^{0} = 0) \end{cases} \\ c_{\max}^{k} \le g_{\max} \end{cases}$$ $$i = k,\ k-1;\ j = 1, 2, \ldots, n$$

Parameter	Default	Range	Description
			Where, x is input variable; x^0 is the initial design; x^L, x^U are lower bound and upper bound of input variables respectively; k is the current iteration number; n is the number of input variables; y is the input variable convergence parameter.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	**Terminate optimization** Terminates with an error message when an analysis run fails. **Ignore failed evaluations** Ignores the failed analysis run, reduces the preceding step size by 50%, and attempts the analysis again.

Table 8: More Tab

Parameter	Default	Range	Description
Initial Linear Move	By DV Initial	By DV Initial By DV Bounds	**By DV initial** Initial move = Initial Input Perturbation * Move Limit Fraction * abs(INI). Default when initial value of input variable is non-zero. An exception is that initial move will be set to minimum move if it is less than minimum move. Minimum move = Minimal Move Factor * (UB-LB) if (UB-LB) is less than 1. Minimum move = Minimal Move Factor if (UB-LB) is not less than

Parameter	Default	Range	Description
			1 and absolute value of INI is less than 1. Minimum move = Minimal Move Factor * min((UB-LB),abs(INI)) if (UB-LB) is not less than 1 and absolute value of INI is not less than 1. **By DV bounds** Initial move = Initial Input Perturbation * Move Limit Fraction * (UB-LB). Default when initial value of input variable is zero. INI Initial input variable value LB, UB Lower and upper bounds on input variable
Move Limit Fraction	0.15	0.0 < Move Limit Fraction < 1.0	Move limit fraction. **Note:** Smaller values allow for a more steady convergence (smaller fluctuation of the output response values), but more computational effort could be consumed. The value will be adaptively updated during optimization process.
Initial Input Perturbation	1.1	≠0.0	Initial input variable perturbation value. Larger value result in wider spread of the initial N designs (n is the number of input variables; the $n$designs together with the start design can determine a linear response surface). Adaptive Response Surface Method will search the design space more widely.

Parameter	Default	Range	Description
Constraint Screening (%)	50.0	real value	Constraint screening. **> 0.0** Constraint is retained (not screened out) if it is violated or within the given percentage of its critical value (bound). **< 0.0** As many constraints are retained as memory permits.
Max Failed Evaluations	20,000	>=0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.
Minimal Move Factor	0.1	0.0 < Minimal Move Factor < Move Limit Fraction	Minimal move factor. It is to avoid too small of the step size. It is used in the initial sampling step (See Minimal Move Factor in Initial Linear Move) and also in the preceding move limit strategy.
Response Surface	SORS	SORS SRSM	**SORS** Uses the second order response surface (SORS). **SRSM** Uses the scalable response surface method (SRSM). When SRSM is used, the limit on Maximum Iterations >= N+2 should be deleted, where N is number of input variables. When there are a lot of input variables and the computational effort is limited, SRSM is a good choice.
Solver	SQP	MFD SQP	The method Adaptive Response Surface Method uses to solve the response surface based optimization problem.

Parameter	Default	Range	Description
		Hybrid	**Tip:** It is recommended to use 2 when there are a lot of discrete variables.
Points per Iteration	1	>0	Controls the number of points used in an iteration after the first iteration. The number of points used per iteration can result in different iteration histories.
Sample Points	0	>=0	**0** Automatically determined; in SRSM, Sample Points is set to n. n is the number of input variables. **>0** Use the user defined value. Sample Points is useful only if Response Surface = 1.
Use SVD	No	No or Yes	Useful in case of soft convergence In case of soft convergence: **No** Adaptive Response Surface Method is terminated. **Yes** Singular Value Decomposition is activated to re-build the response surfaces, and the optimization process is continued.
Revision	A-multi	A B A-multi B-multi	Assists when there is a convergence difficulty. The B revision is less likely to become stuck if iterations do not exhibit successive improvement. By default, "A" is selected meaning the legacy algorithm.

Parameter	Default	Range	Description
			Note: A-multi and B-multi are new versions of A and B that support multi-execution. The classification of iteration points is different between A and A-multi (and B and B-multi).
Use Inclusion Matrix	No	No With Initial Without Initial	**No** Ignores the Inclusion matrix. **With Initial** Runs the initial point. The best point of the inclusion or the initial point is used as the starting point. **Without Initial** Does not run the initial point. The best point of the inclusion is used as the starting point. **Restart ARSM** Used with existing data from an Adaptive Response Surface Method run with the same settings. Mainly used if the Optimization is terminated early or reached the limit on the number of evaluations. If the included data is not from an existing Adaptive Response Surface Method run, for example from a DOE, it can negatively effect the performance

Parameter	Default	Range	Description
			of Adaptive Response Surface Method.

ARSM-Based Sequential Optimization and Reliability Assessment (SORA_ARSM)

Reliability and robustness based optimization methods require many design evaluations, therefore improving their efficiency is one of the issues. ARSM-Based Sequential Optimization and Reliability Assessment attempts to address this issue by using Adaptive Response Surface Method (ARSM).

In this process, response surfaces are created and an optimization is carried out on the surfaces. During deterministic optimization and reliability analysis, response surfaces are adaptively updated for increased accuracy.

Usability Characteristics

- ARSM-Based Sequential Optimization and Reliability Assessment is the most efficient of the three RBDO methods available in HyperStudy, but it is also the least accurate.
- An extension of ARSM-Based Sequential Optimization and Reliability Assessment is implemented in HyperStudy to allow for robust design optimization. Robust design optimization attempts to minimize the objective variance in order to reduce its sensitivity to design variations and consequently increase the design's robustness. The implementation in HyperStudy is based on the use of percentiles for the objective function and is turned on via the Robust Optimization setting in the Specification step.
- ARSM-Based Sequential Optimization and Reliability Assessment terminates if one of the conditions are met.
    - One of the three convergence criteria are met.
        - The absolute objective change is less than a convergence tolerance value (Termination Criteria) and there is no constraint violation (Constraint Violation Tol. (%)) in the last design.
        - The relative objective change is less than a convergence tolerance value (Termination Criteria) and there is no constraint violation (Constraint Violation Tol. (%)) in the last design.
        - The absolute change and relative change of each input variable (Termination Criteria) is less than a convergence tolerance value. Also, there is no constraint violation (Constraint Violation Tol. (%)) in the last design.
    - The maximum number of allowable iterations (Maximum Iterations) is reached.

  An exception is when the absolute change and relative change of each input variable is less than this value in the last two consecutive designs or when we have found feasible designs and the best feasible design has not been improved during the last two consecutive iterations. When this occurs, ARSM-Based Sequential Optimization and Reliability Assessment will be terminated.

- The reliability analysis is carried out by searching for the most probable point (MPP). Issues such as non-uniqueness of the MPP and highly non-linear output response functions can reduce the accuracy of the reliability calculation.
- The algorithm begins with N+1 evaluations, where N is the number of design variables. Evaluations in subsequent iterations are generated sequentially.

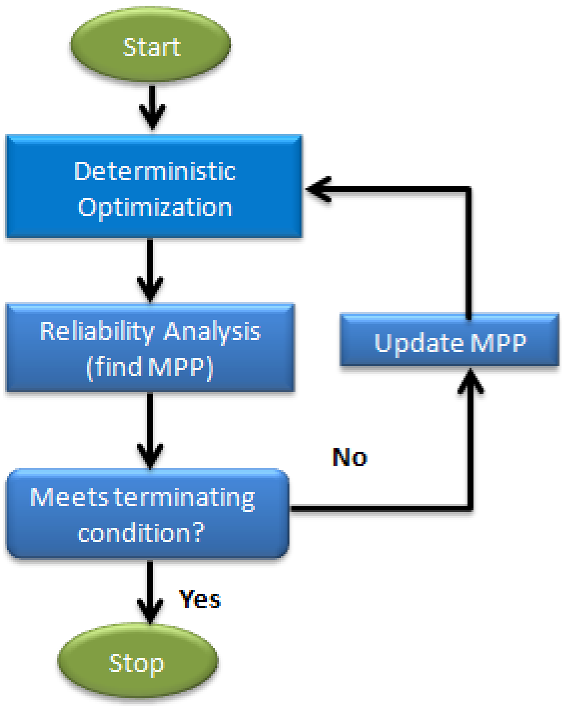


Figure 143: ARSM-Based Sequential Optimization and Reliability Assessment Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Maximum Iterations and Robust Optimization.

Table 9: Settings Tab

Parameter	Default	Range	Description
Maximum Iterations	25	> 0	Maximum number of iterations allowed.
Angle Convergence Tol.	0.25	> 0.0	Angle convergence tolerance for inverse MPP search, in unit of degrees. If the angle between the vector of $\overline{u}$ (design point in standard normal distribution space) and the negative gradient falls

Parameter	Default	Range	Description
			within the tolerance, then inverse MPP search is regarded as converged. **Tip:** A smaller value favors a higher precision of reliability analysis, but more computational effort is needed.
Robust Optimization	No	No or Yes	Defines whether this is a robust optimization or not. **No** Do not use robust optimization. **Yes** Use robust optimization.
Robust Min %	95.0	> 50 < 100	Defines the percentile value of robust optimization for minimization objective.
Robust Max %	5.0	> 0 < 50	Defines the percentile value of robust optimization for maximization objective.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	**Terminate optimization** Terminates with an error message when an analysis run fails. **Ignore failed evaluations** Ignores the failed analysis run.

Table 10: More Tab

Parameter	Default	Range	Description
Termination Criteria	1.0e-4	> 0.0	If the absolute or relative change of the objective value is less than this value, or the absolute or relative change of the input variables is less than this value, and the constraint violation is not larger than this value, then ARSM-Based

Parameter	Default	Range	Description
			Sequential Optimization and Reliability Assessment will be terminated. Also, there must not be any constraints with an allowable violation that has been exceeded in the last design. An exception is when the absolute change and relative change of each input variable is less than this value in the last two consecutive designs or when we have found feasible designs and the best feasible design has not been improved during the last two consecutive iterations, ARSM-Based Sequential Optimization and Reliability Assessment will be terminated. $\begin{cases} c_{\max}^{k} \le g_{\max} \\ \left[\left\lvert \frac{f^{k}-f^{k-1}}{\lvert f^{k-1}\rvert+10^{-10}} \right\rvert < \varepsilon\right] or \left[\lvert f^{k}-f^{k-1}\rvert < \varepsilon\right] or \left[\begin{cases} \left\lvert \frac{x_j^{k}-x_j^{k-1}}{\lvert x_j^{k-1}\rvert+10^{-10}} \right\rvert < \varepsilon, j=1,\ldots,n \\ \lvert x_j^{k}-x_j^{k-1}\rvert < \varepsilon \end{cases}\right] \end{cases}$ Where, x is the input variable; f is the objective; n is the number of input variables; k is the current iteration number; $c_{\max}$ is the maximum constraint violation; $g_{\max}$ is the allowable constraint violation; ε is the value of the termination criteria. $if\left(\begin{cases} \left\lvert \frac{x_j^{i}-x_j^{i-1}}{\lvert x_j^{i-1}\rvert+10^{-10}} \right\rvert < \varepsilon, j=1,\ldots,n \\ \lvert x_j^{i}-x_j^{i-1}\rvert < \varepsilon \end{cases}\right)$ $.or.\left(\begin{cases} c_{\max}^{k-2} \le g_{\max} \\ \left(c_{\max}^{i} > g_{\max}\right).or.\begin{pmatrix} f^{i} \ge f^{k-2}, \text{ minimization} \\ f^{i} \le f^{k-2}, \text{ maximization} \end{pmatrix} \end{cases}\right)$ $i=k, k-1$
Move Limit Fraction	0.1	0.0 < Move Limit Fraction < 1.0	Move limit fraction. Smaller values allow more steady convergence (smaller fluctuation of the output response values), but more computational effort could be consumed. The value will be adaptively

<table>
<tr><th>Parameter</th><th>Default</th><th>Range</th><th>Description</th></tr>
<tr><td></td><td></td><td></td><td>updated during the optimization process.</td></tr>
<tr><td>Initial Linear Move</td><td>By DV Initial</td><td>By DV Initial
By DV Bounds</td><td>By DV Initial: Initial move = Initial Input Perturbation * Move Limit Fraction * abs(INI)
(Default when initial value of input variable is non-zero)
An exception is that initial move will be set to minimum move if it is less than minimum move.
Minimum move = Minimal Move Factor * (UB-LB) if (UB-LB) is less than 1.
Minimum move = Minimal Move Factor if (UB-LB) is not less than 1 and absolute value of INI is less than 1.
Minimum move = Minimal Move Factor * min((UB-LB),abs(INI)) if (UB-LB) is not less than 1 and absolute value of INI is not less than 1.
By DV Bounds: Initial move = Initial Input Perturbation * Move Limit Fraction * (UB-LB)
(Default when initial value of input variable is zero)
INI: Initial input variable value.</td></tr>
</table>

Parameter	Default	Range	Description
			LB, UB: Lower and upper bounds on input variable.
Minimal Move Factor	0.1	0.0 < Minimal Move Factor < Move Limit Fraction	Minimal move factor. See the usage of Minimal Move Factor in Initial Linear Move.
Initial Input Perturbation	1.1	≠ 0.0	Initial input variable perturbation value. Larger values result in a wider spread of the initial N designs (N is the number of input variables; the N designs together with the start design can determine a linear response surface). Adaptive Response Surface Method will search the design space more widely.
Use SVD	No	No or Yes	*No* Adaptive Response Surface Method is terminated in case of soft convergence (when the current design is the same or little change (1.0e-15) to one of the history designs). *Yes* Singular Value Decomposition is activated to build the response surfaces and the optimization process will be continued. $\left\|x_i^k - x_i^j\right\| \le 10^{-15}, i = 1, .., n$ $\forall\ j, j \in \{1 \le j < k\}$ Where, k is the current iteration number; n is the number of input variables; xj is one of the history designs.
Revision	A	A or B	Used to help when there is a convergence difficulty. By default, "A" is selected meaning the legacy algorithm.

Genetic Algorithm (GA)

Modeled after the evolutionary process theory.

Genetic Algorithm starts with the creation of a population of designs (a generation). These designs are then ranked with respect to their fitness. Fitness is a measure of how good a design is and it is calculated as a function of constraint violation and objective function values. Selected designs are then reproduced through the application of genetic operators, typically crossover and mutation. The individuals that result from this process (the children) become members of the next generation. This process is repeated for many generations until the evolution of a population converges to the optimal solution.

Usability Characteristics

- Genetic Algorithm differs from conventional optimization techniques in the following ways:
    - They are classified as exploratory methods.
    - They work on a population of designs at once.
    - The design population can be run in parallel.
    - They do not show the typical convergence of other optimization algorithms. You will typically select the maximum number of iterations (generations) to be evaluated. A number of solver runs is executed in each generation, with each run representing a member of the population.
- Genetic Algorithm does a global search.
- They are well suited for discrete problems.
- Genetic Algorithm is computationally expensive as it requires a large number of runs. In HyperStudy, a local search algorithm (Hooke-Jeeves or a response surface method) is utilized to improve the efficiency of Genetic Algorithm.
- In HyperStudy, population size is calculated automatically according to the optimization problem that you defined. It can also be modified manually.
- In HyperStudy, both a binary and a real coded Genetic Algorithm exists. Default is the real coded Genetic Algorithm as it is more efficient than the binary coded Genetic Algorithm.
- Genetic Algorithm terminates if one of the conditions below are met:
    - The change of the objective function during several successive iterations (as controlled by the Global search setting) is less than 0.1%.
    - The maximum number of allowable iterations (Maximum Iterations) is reached.
    - An analysis fails and the Terminate optimization option is the default (On Failed Evaluation).
- Supports input variable constraints.
- Although the number of evaluations per iteration is a combination of multiple settings, it is primarily affected by the Population Size setting. All evaluations within an iteration may be executed in parallel. To take advantage of parallel computation and muli-execution, set the hybrid algorithm to the meta model based method.

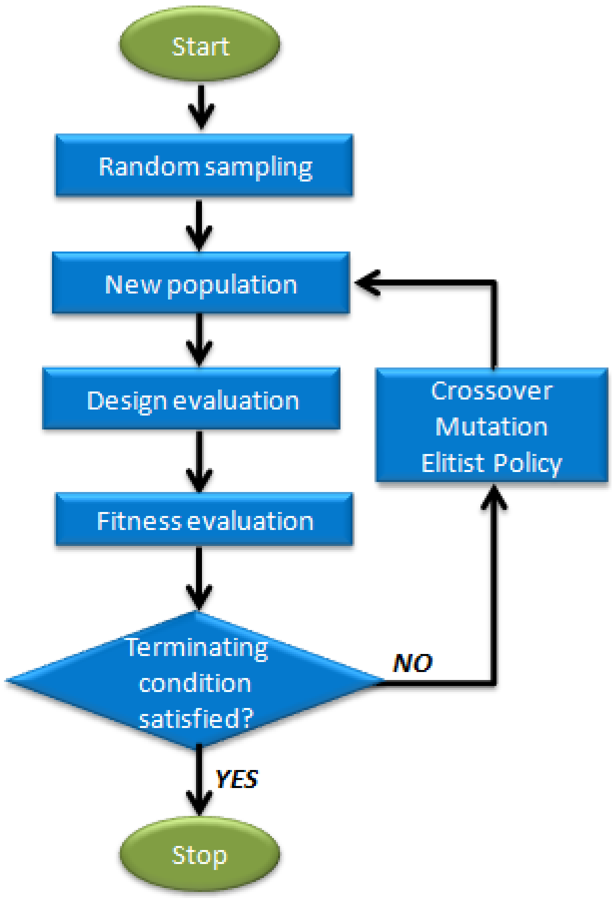


Figure 144: Genetic Algorithm Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Maximum Iterations and On Failed Evaluation.

Table 11: Settings Tab

Parameter	Default	Range	Description
Maximum Iterations	200	>0	Maximum number of iterations allowed.
Minimum Iterations	25	>0 <=Maximum Iterations	Processes at least Minimum Iterations iteration steps. Use this setting to prevent pre-mature convergence.

Parameter	Default	Range	Description
			By setting Minimum Iterations to be the same as Maximum Iterations, the defined number of iteration steps will be run.
Population Size	0	Integer > 1	If Population Size is 0, then population size is calculated according to the following equation, where N is the number of input variables. $Population\ Size = \left[3 + 37e^{-\left(\frac{N-1}{5.806}\right)^{0.5}}\right]N$ If Population Size is greater than 0, then population size uses the user defined value. If the allowable computational effort is limited, set your own value. **Tip:** In general, it is better to process at least 25 iteration steps.
Global Search	2	Integer 1 2 3	Controls the global search ability. Requires ast least 0.1% improvement in the objective of the most recent M iterations or terminates when M is 5 times this setting.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	**Terminate optimization**: Terminates with an error message when an analysis run fails. **Ignore failed evaluations**: Ignores the failed analysis run.

Table 12: More Tab

Parameter	Default	Range	Description
Type	Real	Real or Binary	**Real**: Real coded Genetic Algorithm is used.

Parameter	Default	Range	Description
			Note: When Type is Real, Discrete States, Number of Contenders, Penalty Multiplier, and Penalty Power are grayed out. **Binary** Binary coded Genetic Algorithm is used. **Note:** When Type is Binary, then Distribution Index is grayed out. In general, real coded Genetic Algorithm performs better than binary coded Genetic Algorithm. For discrete optimization problems, binary coded Genetic Algorithm could be better.
Discrete States	1024	Integer > 1	Number of discrete values uniformly covering the range of continuous variables including upper and lower bound. **Tip:** Select as a power of 2, for example 64 = 2^6, 1024 =2^10, and so on. A larger value allows for higher solution precision, but more computational effort is needed to find the optima.
Mutation Rate	0.01	0.0-1.0	Mutation rate (probability). Larger values introduce a more random effect.

Parameter	Default	Range	Description
			As a result, the algorithm can explore more globally but the convergence could be slower. **Tip:** Recommended range: 0.001 – 0.05
Elite Population (%)	10	1.0-50.0	Percentage of population that belongs to elite. The one with highest fitness value is directly passed to the next generation. This is a very important strategy, as it ensures the quality of solutions be non-decreasing. A larger value means that more individuals will be directly passed to the next generation, therefore new gene has less chance to be introduced. The convergence speed could be increased. The drawback is that too large of values could cause premature convergence. **Tip:** Recommended range: 1.0 – 20.0.
Random Seed	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Number of Contenders	2	Integer 2 to 5	Number of contenders in a tournament selection. For larger values, individuals with lower fitness value have less chance to be selected. Thus, the good individuals have more chance to produce offspring. The bad effect is that, diversity of the population is

Parameter	Default	Range	Description
			reduced. The algorithm could converge prematurely.
Penalty Multiplier	2.0	>0.0	Initial penalty multiplier in the formulation of the fitness function as exterior penalty function. Penalty multiplier will be increased gradually with iterating steps going on. In general, larger values allow the solution to become feasible with less iteration steps; but too large of a value could result in a worse solution. **Tip:** Recommended range: 1.0 – 5.0.
Penalty Power	1	>0.0 <10.0	Penalty power in the formulation of the fitness function as exterior penalty function. **Tip:** Recommended range: 1.0 – 2.0.
Distribution Index	5	Integer 1 to 100	Distribution index used by real coded Genetic Algorithm. Controls offspring individuals to be close to or far away from the parent individuals. Increasing the value will result in offspring individuals being closer to the parents. **Tip:** Recommended range: 3.0 – 10.0.
Max Failed Evaluations	20,000	>=0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.

Parameter	Default	Range	Description
Hybrid Algorithm	Hooke-Jeeves method	Hooke-Jeeves method Meta-model based method No hybrid	Hybrid algorithm used in Genetic Algorithm. **Note:** This parameter is used in Genetic Algorithm real type. It is not available for binary type.
Use Inclusion Matrix	No	No With Initial Without Initial	**No** Ignores the Inclusion matrix **With Initial** Runs the initial point. The inclusion set and initial point are used to build the initial response surface. **Without Initial** Does not run the initial point. The inclusion set is used to build the initial response surface.

Global Response Search Method (GRSM)

A response surface based approach. During each iteration, the response surface based optimization generates a few designs. Additional designs are generated globally to ensure a good balance on local search capability and global search capability. Response surface is adaptively updated with the newly generated designs to have a better fit of the model.

Usability Characteristics

- Global Response Search Method can disposes of both single objective problems and multi-objective problems.
- Default method for multi-objective optimization problems. Global Response Search Method is also the suggested method when you are solving a single objective optimization problem with a large number of input variables and/or when a global optima is required.
- It is recommend to use Global Response Search Method directly on a solver and not on a Fit. If you have a Fit, consider using an Inclusion matrix with your data.
- Consists of a global search capability.
- Supports discrete optimization.
- All the designs generated in one iteration can be solved in parallel.

- If the model analysis is time consuming, then Global Response Search Method is a good choice. If the model analysis is quite cheap and a thorough search of the design space is needed (for example, 100000 model evaluations are needed), then Genetic Algorithm or Multi - Objective Genetic Algorithm are recommended.
- In the case of a failed run, it is possible to ignore a failed analysis or terminate an optimization. When omitting failed runs, the optimizer randomly generates designs in under-sampled region in order to explore the whole design space effectively.
- Terminates when the maximum number of evaluations (Number of Evaluations) is reached.
- Supports input variable constraints.
- The size of the first iteration is controlled by the Initial Sampling Points setting. The number of evaluations in subsequent iterations is controlled by the Points per Iteration setting. All designs generated within one iteration can be executed in parallel.

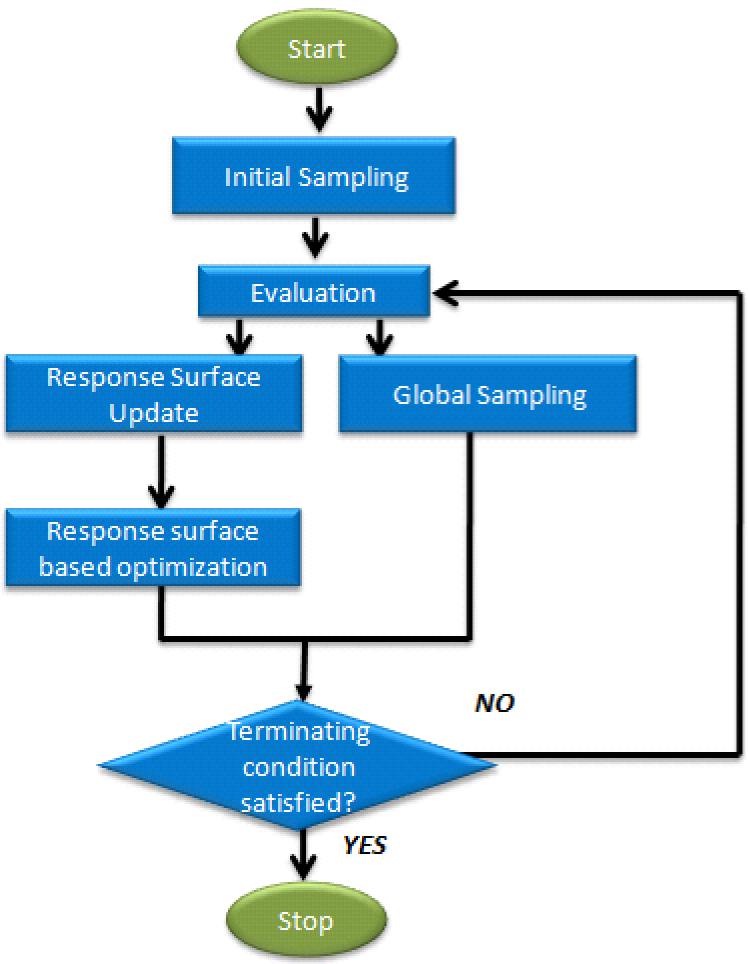


Figure 145: Global Response Search Method Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Number of Evaluations and On Failed Evaluation.

Table 13: Settings Tab

Parameter	Default	Range	Description
Number of Evaluations	Single objective: 50 Multiple objectives: 200	>0	Number of evaluations allowed.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	*Terminate optimization* Optimizer terminates with an error message when an analysis run fails. *Ignore failed evaluations* Optimizer ignores the failed analysis run, randomly generates a design, and re-tries the analysis.

Table 14: More Tab

Parameter	Default	Range	Description
Initial Sampling Points	min(20,n+2)	Integer >=0	The number of initial sample points. Default is min(20,n+2) initial sample points; n is the number of input variables; > 0 use the user defined value.
Random Seed	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable) **>0** Triggers a new sequence of pseudo-random

Parameter	Default	Range	Description
			numbers, repeatable if the same number is specified.
Points per Iteration	2	1 to Initial Sampling Points	Number of designs to be evaluated for optimum design search and response surface update. After the initial Initial Sampling Points designs, Points per Iteration designs are generated from response surface based optimization and/or incremental sampling. In the iteration that follows, these designs are used to adaptively update the response surface to have a better fit of the model. **Note:** These designs can be evaluated in parallel.
Max Failed Evaluations	20,000	>=0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.
Stop after no Improvement	1000	> 0.0	Terminates the optimization if the number of iterations without improvement exceeds this value.
Use Inclusion Matrix	No	No With Initial Without Initial	**No** Ignores the Inclusion matrix. **With Initial** Runs the initial point. The inclusion set and initial point are used to build the initial response surface. **Without Initial** Does not run the initial point. The inclusion

Parameter	Default	Range	Description
			set is used to build the initial response surface.

Method of Feasible Directions (MFD)

The fundamental principle behind the Method of Feasible Directions is to move from one feasible design to an improved feasible design, therefore, the objective function must be reduced and the constraints at the new design point should not be violated.

Used for solving constrained optimization problems.

Usability Characteristics

- A gradient-based method, which will most likely find the local optima.
- May be efficient with a large number of constraints, but in general it is less accurate than Sequential Quadratic Programming and less efficient than Adaptive Response Surface Method.
- One iteration of Method of Feasible Directions will require a number of simulations. The number of simulations required is a function of the number of input variables since finite difference method is used for gradient evaluation. As a result, it may be an expensive method for applications with a large number of input variables.
- Method of Feasible Directions terminates if one of the conditions below are met:
    - One of the two convergence criterias are satisfied.
        - Absolute convergence (Absolute Convergence)
        - Relative convergence (Relative Convergence (%))
    - The maximum number of allowable iterations (Number of Evaluations) is reached.
    - An analysis fails and the “Terminate optimization” option is the default (On Failed Evaluation).
- The number of evaluations in each iteration is automatically set and varies due to the finite difference calculations used in the sensitivity calculation. The number of evaluations in each iteration is dependent of the number of variables and the Sensitivity setting. The evaluations required for the finite difference are executed in parallel. The evaluations required for the line search are executed sequentially.

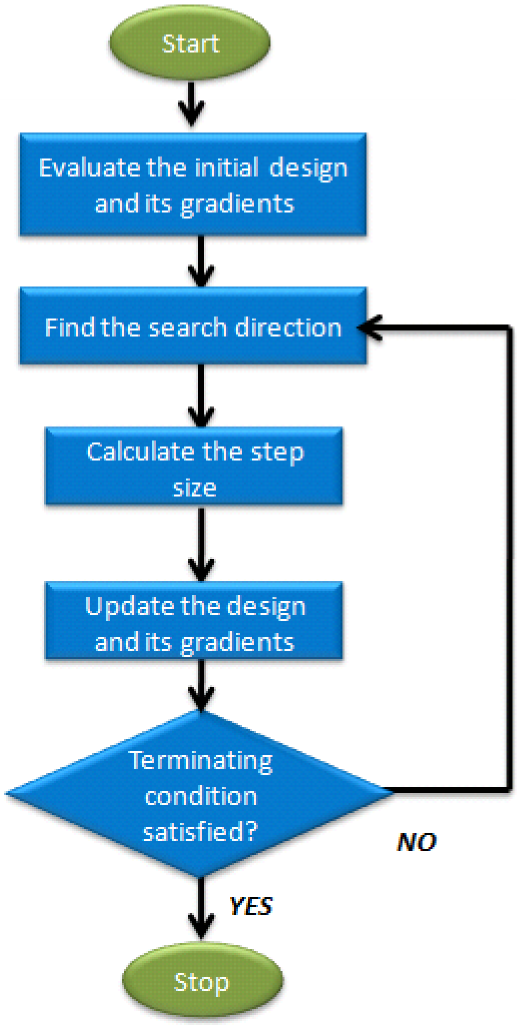


Figure 146: Method of Feasible Directions Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Maximum Iterations and On Failed Evaluation.

Table 15: Settings Tab

Parameter	**Default**	**Range**	**Description**
Maximum Iterations	25	>0	Maximum number of iterations allowed.
Absolute Convergence	0.001	>0.0	Determines an absolute convergence tolerance, which is constant and equal to Absolute Convergence, times the

Parameter	Default	Range	Description
			initial objective function value. The design has converged when there are three consecutive designs for which the absolute change in the objective function is less than this tolerance. There also must not be any constraint whose allowable violation is exceeded in these three consecutive designs. **Note:** A larger value allows for faster convergence, but worse results could be achieved. $\begin{cases} c_{\max}^{i} \le g_{\max} \\ \lvert f^{i} - f^{i-1} \rvert < \varepsilon \\ i = k, k-1, k-2 \end{cases}$ where f is the objective value; k is the current iteration number; ε is the absolute convergence parameter; $c_{\max}$ is the maximum constraint violation; $g_{\max}$ is the allowable constraint violation.
Relative Convergence (%)	1.0	>0.0	The design has converged if the relative (percent) change in the objective function is less than this value for three consecutive designs. There also must not be any constraint whose allowable violation is exceeded in these three consecutive designs. **Note:** A larger value allows for faster convergence, but worse results could be achieved. $\begin{cases} c_{\max}^{i} \le g_{\max} \\ \begin{cases} \left\lvert \frac{f^{i} - f^{i-1}}{f^{i-1}} \right\rvert < \varepsilon, \ \text{if}(\lvert f^{i-1} \rvert > 10^{-10}) \\ \lvert f^{i} - f^{i-1} \rvert < \varepsilon, \ \text{if}(\lvert f^{i-1} \rvert \le 10^{-10}) \end{cases} \\ i = k, k-1, k-2 \end{cases}$

Parameter	Default	Range	Description
			where f is the objective value; k is the current iteration number; ε is the relative convergence parameter, $c_{\max}$ is the maximum constraint violation; $g_{\max}$ is the allowable constraint violation.
On Failed Evaluation	Terminate optimization	Terminate optimization (default) Ignore failed evaluations	**Terminate optimization** Terminates with an error message when an analysis run fails. **Ignore failed evaluations** Ignores the failed analysis run, and tries different step sizes to do line search.

Table 16: More Tab

Parameter	Default	Range	Description
Max Failed Evaluations	20,000	>=0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.
Use Perturbation size	No	No or Yes	Enables the use of Perturbation Size, otherwise an internal automatic perturbation size is set.
Perturbation Size	0.0001	>0.0	Defines the size of the finite difference perturbation. For a variable x, with upper and lower bounds (xu and xl, respectively), the following logic is used to preserve reasonable perturbation sizes across a range of variables magnitudes: • If abs(x) >= 1.0 then perturbation = Perturbation Size * abs(x) • If (xu - xl) < 1.0 then perturbation = Perturbation Size * (xu – xl)

Parameter	Default	Range	Description
			• Otherwise perturbation = Perturbation Size
Use Inclusion Matrix	No	No With Initial Without Initial	**No** Ignores the Inclusion matrix **With Initial** Runs the initial point. The inclusion set and initial point are used to build the initial response surface. **Without Initial** Does not run the initial point. The inclusion set is used to build the initial response surface.

Multi - Objective Genetic Algorithm (MOGA)

An extension of Genetic Algorithm that solves multi-objective optimization (MOO) problems.

In MOO problems, there is more than one objective function to be minimized or maximized and as such the goal is not to find an optimum but to find the Pareto front instead. Pareto front is a collection of non-dominated designs. Non-dominated designs are better than other designs because at least one of the objective functions are considered.

Usability Characteristics

- Multi - Objective Genetic Algorithm uses a crowding distance metric to create a homogeneous distribution of the non-dominated points on the Pareto front.
- Multi - Objective Genetic Algorithm terminates if one of the conditions below are met:
  - The convergence criteria is satisfied. This occurs when the minimum number of allowable iterations (Minimum Iterations) are run, feasible designs are found (Constraint Violation Tol. (%)), and the non-dominated designs did not change in the last iteration.
  - The maximum number of allowable iterations (Maximum Iterations) is reached.
  - An analysis fails and the Terminate optimization option is the default (On Failed Evaluation).
- Supports input variable constraints.
- Although the number of evaluations per iteration is a combination of multiple settings, it is primarily affected by the Population Size setting. All evaluations within an iteration may be executed in parallel. If parallel computing is required, it is recommended to use the Meta-Model or No Hybrid method.

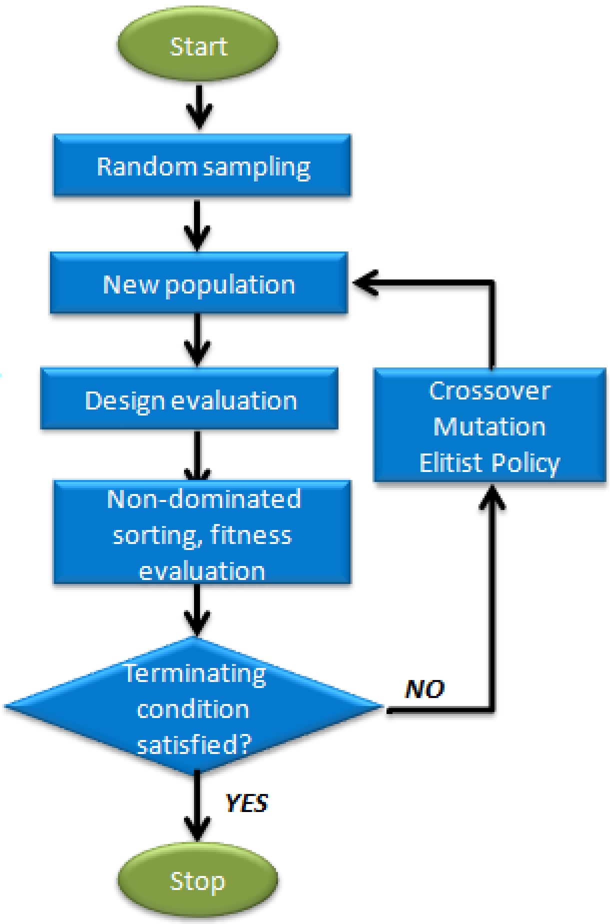


Figure 147: Multi - Objective Genetic Algorithm Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Maximum Iterations and On Failed Evaluation.

Table 17: Settings Tab

Parameter	Default	Range	Description
Maximum Iterations	50	>0	Maximum number of iterations allowed.
Minimum Iterations	25	>0 <=Maximum Iterations	Processes at least Minimum Iterations iteration steps. Use this setting to prevent pre-mature convergence.

Parameter	Default	Range	Description
			By setting Minimum Iterations to be the same as Maximum Iterations, the defined number of iteration steps will be run. Multi - Objective Genetic Algorithm will be terminated if it has iterated the minimum iteration steps and feasible designs are found and the non-dominated designs did not change in the last iteration.
Population Size	0	Integer > 1	If Population Size is 0, then population size is calculated according to the following equation, where N is the number of input variables. $$Population\ Size = \left[3 + 37e^{-\left(\frac{N-1}{5.806}\right)^{0.5}}\right]N$$ If Population Size is greater than 0, then population size uses the user defined value. If the allowable computational effort is limited, set your own value. **Tip:** In general, it is better to process at least 25 iteration steps.
On Failed Evaluation	Ignore failed evaluations	Terminate optimization Ignore failed evaluations	**Terminate optimization** Terminates with an error message when an analysis run fails. **Ignore failed evaluations** Ignores the failed analysis run.

Table 18: More Tab

Parameter	Default	Range	Description
Crowding Distance	Design Space	Design Space Solution Space	Determines in which space the crowding distance is evaluated.

Parameter	Default	Range	Description
		Design/ Solution Space	The crowding distance evaluation strategy allows users to get solutions more uniformly distributed in the selected space. **Design Space** Crowding distance is evaluated in design space. **Solution Space** Crowding distance is evaluated in solution space. **Design/ Solution Space** Crowding distance is evaluated in both of the two spaces.
Discrete States	1024	Integer > 1	Number of discrete values uniformly covering the range of continuous variables including upper and lower bound. **Tip:** Select as a power of 2, for example 64 = 2^6, 1024 =2^10, and so on. A larger value allows for higher solution precision, but more computational effort is needed to find the optima.
Mutation Rate	0.01	0.0 - 1.0	Mutation rate (probability). Larger values introduce a more random effect. As a result, the algorithm can explore more globally but the convergence could be slower. **Tip:** Recommended range: 0.001 – 0.05
Elite Population (%)	10	1.0 - 50.0	Percentage of population that belongs to elite. The one with highest fitness value is directly passed to the next generation. This is a very important strategy, as it ensures the quality of

Parameter	Default	Range	Description
			solutions be non-decreasing. A larger value means that more individuals will be directly passed to the next generation, therefore new gene has less chance to be introduced. The convergence speed could be increased. The drawback is that too large of values could cause premature convergence. **Tip:** Recommended range: 1.0 – 20.0.
Random Seed	0	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Number of Contenders	2	Integer 2 to 5	Number of contenders in a tournament selection. For larger values, individuals with lower fitness value have less chance to be selected. Thus, the good individuals have more chance to produce offspring. The bad effect is that, diversity of the population is reduced. The algorithm could converge prematurely.
Penalty Power	1	> 0.0 < 10.0	Penalty power in the formulation of the fitness function as exterior penalty function. **Tip:** Recommended range: 1.0 – 2.0.
Penalty Multiplier	2.0	> 0.0	Initial penalty multiplier in the formulation of the fitness function

Parameter	Default	Range	Description
			as exterior penalty function. Penalty multiplier will be increased gradually with iterating steps going on. In general, larger values allow the solution to become feasible with less iteration steps; but too large of a value could result in a worse solution. **Tip:** Recommended range: 1.0 – 5.0.
Distribution Index	5	Integer 1 to 100	Distribution index used by real coded Multi - Objective Genetic Algorithm. Controls offspring individuals to be close to or far away from the parent individuals. Increasing the value will result in offspring individuals being closer to the parents. **Tip:** Recommended range: 3.0 – 10.0.
Type	Real	Real or Binary	**Real** Use real coded Multi - Objective Genetic Algorithm. **Binary** Use binary coded Multi - Objective Genetic Algorithm. In general, real coded Multi - Objective Genetic Algorithm performs better than binary coded Multi - Objective Genetic Algorithm. For discrete optimization problem, binary coded Multi - Objective Genetic Algorithm could be better.
Max Failed Evaluations	20,000	>=0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.

Parameter	Default	Range	Description
Hybrid Algorithm	No hybrid	No hybrid Meta-model based method	
Use Inclusion Matrix	No	No With Initial Without Initial	**No** Ignores the Inclusion matrix **With Initial** Runs the initial point. The inclusion set and initial point are used to build the initial response surface. **Without Initial** Does not run the initial point. The inclusion set is used to build the initial response surface.

Sequential Optimization and Reliability Assessment (SORA)

A reliability-based design optimization method. Reliability-based design optimization (RBDO) methods take uncertainties in the design into account and search for designs that satisfy the design requirements with a required probability of success.

A reliability-based design problem is formulated as:

Objective	min f(x, r, p)
Constraints	P(g(x, r, p ≤ 0.0) > PS

where,

x	Deterministic input variables
r	Random input variables (affect the design but are subject to uncertainties)
p	Pure random parameters (variables we have no control over but affect the design, such as humidity and temperature)

Usability Characteristics

- An extension of Sequential Optimization and Reliability Assessment is implemented in HyperStudy to allow for robust design optimization. Robust design optimization attempts to minimize the objective variance in order to reduce its sensitivity to design variations and consequently increase the design's robustness. The implementation in HyperStudy is based on the use of percentiles for the objective function and is turned on via the Robust Optimization setting in the Specification step.

- Sequential Optimization and Reliability Assessment is the most accurate of the three RBDO methods available in HyperStudy. It is also the most expensive.
- Sequential Optimization and Reliability Assessment terminates if one of the conditions below are met:
    - One of the two convergence criterias are met.
        - The absolute objective change is less than a convergence tolerance value (Termination Criteria) and there is no constraint violation (Constraint Violation Tol. (%)).
        - The relative objective change is less than a convergence tolerance value (Termination Criteria) and there is no constraint violation (Constraint Violation Tol. (%)) in the last design.
    - The maximum number of allowable iterations (Maximum Iterations) is reached.

  An exception is when the current objective is worse than the previous objective and the constraint violation of the previous design is within allowable violation. When this occurs, Sequential Optimization and Reliability Assessment will be terminated.
- The reliability analysis is carried out by searching for the most probable point (MPP). Issues such as non-uniqueness of the MPP and highly non-linear output response functions can reduce the accuracy of the reliability calculation.
- The number of evaluations in each iteration is automatically set and varies due to the finite difference calculations used in the sensitivity calculation. The number of evaluations in each iteration is dependent of the number of variables. The evaluations required for the finite difference are executed in parallel. The evaluations required for the line search are executed sequentially.

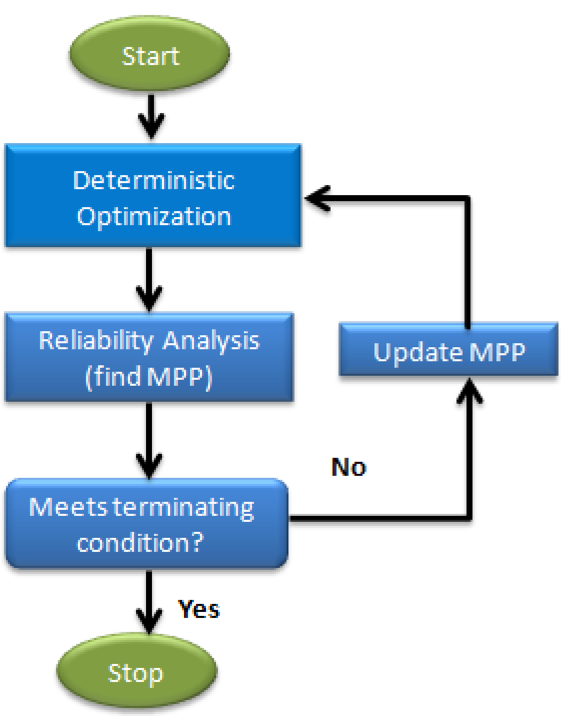


Figure 148: Sequential Optimization and Reliability Assessment Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Maximum Iterations and Robust Optimization.

Table 19: Settings Tab

Parameter	Default	Range	Description
Maximum Iterations	25	> 0	Maximum number of iterations allowed.
Robust Optimization	No	No or Yes	Defines whether this is a robust optimization or not. **No** Do not use robust optimization. **Yes** Use robust optimization.
Robust Min %	95.0	> 50 < 100	Defines the percentile value of robust optimization for minimization objective.
Robust Max %	5.0	> 0 < 50	Defines the percentile value of robust optimization for maximization objective.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	**Terminate optimization** Terminates with an error message when an analysis run fails. **Ignore failed evaluations** Ignores the failed analysis run.

Table 20: More Tab

Parameter	Default	Range	Description
Angle Convergence Tol.	0.25	> 0.0	Angle convergence tolerance for inverse MPP search, in unit of degrees. If the angle between the vector of $\overline{u}$ (design point in standard normal distribution

Parameter	Default	Range	Description
			space) and the negative gradient falls within the tolerance, then inverse MPP search is regarded as converged. **Tip:** A smaller value favors a higher precision of reliability analysis, but more computational effort is needed.
Termination Criteria	1.0e-4	> 0.0	Termination tolerance. If the absolute or relative change of the objective value is less than this value, and the constraint violation is not larger than this value, then Sequential Optimization and Reliability Assessment will be terminated. There also must not be any constraint with an allowable violation that has been exceeded in the last design. $\begin{cases} c_{\max}^{k} \le g_{\max} \\ if \left\| f^{k} - f^{k-1} \right\| < \text{Termination Criteria} \\ or \frac{\left\| f^{k} - f^{k-1} \right\|}{\left\| f^{k-1} \right\| + 10^{-10}} < \text{Termination Criteria} \end{cases}$ Where, f is the objective; k is the current iteration number; $c_{\max}$ is the maximum constraint violation; $g_{\max}$ is the allowable constraint violation; Termination Criteria is the value of the termination criteria. An exception is when the current objective is worse than the previous objective and the constraint violation of the previous design is within allowable violation, Sequential Optimization and Reliability Assessment will be terminated.

Parameter	Default	Range	Description
			$\begin{cases} c_{\max}^{k-1} \le g_{\max} \\ \begin{cases} f^k > f^{k-1}, & \text{minimization} \\ f^k < f^{k-1}, & \text{maximization} \end{cases} \end{cases}$

Sequential Quadratic Programming (SQP)

A gradient-based iterative optimization method and is considered to be the best method for nonlinear problems by some theoreticians. In HyperStudy, Sequential Quadratic Programming has been further developed to suit engineering problems.

Usability Characteristics

- A gradient-based method, therefore it will most likely find the local optima.
- One iteration of Sequential Quadratic Programming will require a number of simulations. The number of simulations required is a function of the number of input variables since finite difference method is used for gradient evaluation. As a result, it may be an expensive method for applications with a large number of input variables.
- Sequential Quadratic Programming terminates if one of the conditions below are met:
    - One of the two convergence criteria is satisfied.
        - Termination Criteria is based on the Karush-Kuhn-Tucker Conditions.
        - Input variable convergence
    - The maximum number of allowable iterations (Maximum Iterations) is reached.
    - An analysis fails and the Terminate optimization option is the default (On Failed Evaluation).
- The number of evaluations in each iteration is automatically set and varies due to the finite difference calculations used in the sensitivity calculation. The number of evaluations in each iteration is dependent of the number of variables and the Sensitivity setting. The evaluations required for the finite difference are executed in parallel. The evaluations required for the line search are executed sequentially.

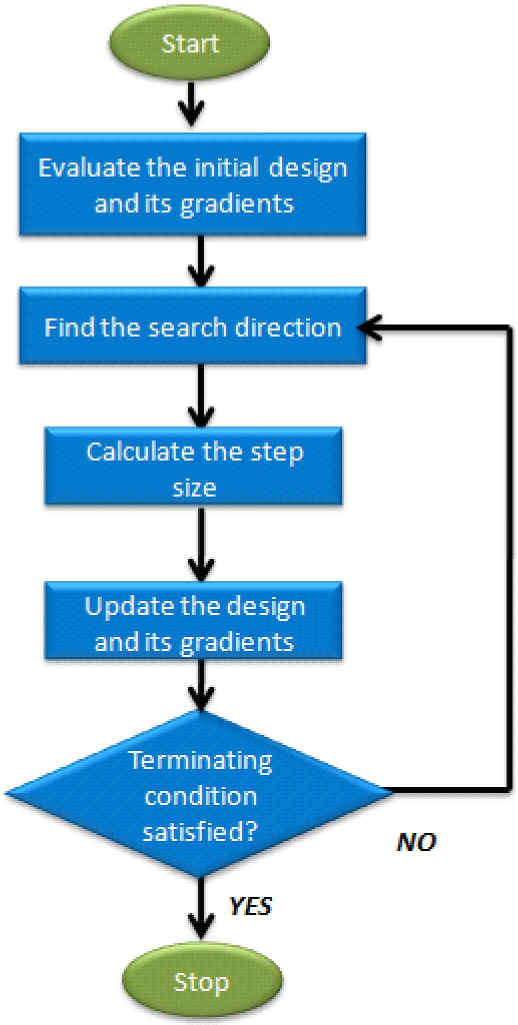


Figure 149: Sequential Quadratic Programming Process Phases

Settings

In the Specifications step, change method settings from the Settings and More tabs.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Maximum Iterations and On Failed Evaluation.

Table 21: Settings Tab

Parameter	Default	Range	Description
Maximum Iterations	25	> 0	Maximum number of iterations allowed.
Design Variable Convergence	0.0	>=0.0	Input variable convergence parameter.

Parameter	Default	Range	Description
			Design has converged when there are two consecutive designs for which the change in each input variable is less than both (1) Design Variable Convergence times the difference between its bounds, and (2) Design Variable Convergence times the absolute value of its initial value (simply Design Variable Convergence if its initial value is zero). There also must not be any constraint whose allowable violation is exceeded in the last design. **Note:** A larger value allows for faster convergence, but worse results could be achieved. $\begin{cases} \lvert x_j^i - x_j^{i-1} \rvert < \gamma \cdot (x_j^U - x_j^L) \\ \lvert x_j^i - x_j^{i-1} \rvert < \gamma \cdot \lvert x_j^0 \rvert, \ \text{if}(x_j^0 \neq 0) \\ \lvert x_j^i - x_j^{i-1} \rvert < \gamma, \ \text{if}(x_j^0 = 0) \\ c_{\max}^k \leq g_{\max} \end{cases}$ $i = k, k-1;\ j = 1, 2, ..., n$ Where, x is input variable; x^0 is the initial design; x^L, x^U are lower bound and upper bound of input variables respectively; k is the current iteration number; n is the number of input variables; y is the input variable convergence parameter.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	**Terminate optimization**: Terminates with an error message when an analysis run fails. **Ignore failed evaluations**: Ignores the failed analysis run. If analysis is failed at line search then the step size is reduced by half and the optimization is continued; if analysis

Parameter	Default	Range	Description
			is failed at gradient calculation then the corresponding gradient is set to zero (an exception is that if the gradient calculation is failed on all of the input variables, then Sequential Quadratic Programming will be terminated).

Table 22: More Tab

Parameter	Default	Range	Description
Termination Criteria	1.0e-4	>0.0	Defines the termination criterion, relates to satisfaction of Kuhn-Tucker condition of optimality. Recommended range: 1.0E-3 to 1.0E-10. In general, smaller values result in higher solution precision, but more computational effort is needed. For the nonlinear optimization problem: $$\begin{array}{ll} \min & f(x) \\ & g_i(x) \leq 0 \\ s.t. & h_j(x) = 0 \\ & i = 1, \ldots, m; \; j = 1, \ldots, 1 \end{array}$$ Sequential Quadratic Programming is converged if: $$\left|S^T \cdot \nabla f\right| + \sum_{i=1}^{m} \left|\mu_i \cdot g_i\right| + \sum_{j=1}^{l} \left|\lambda_j \cdot h_j\right| \leq \Delta$$ Where S is the search direction generated by Sequential Quadratic Programming; ∇f is objective function gradient; μ, λ are Lagrange multipliers;

Parameter	Default	Range	Description
			Δ is the value of the termination criteria parameter.
Sensitivity	Forward FD	Forward FD Central FD Asymmetric FD Analytical	Defines the way the derivatives of output responses with respect to input variables are calculated. **Forward FD** For approximation by one step forward finite difference scheme. $df/dx = (f(x+dx) - f(x))/(dx)$ **Central FD** For approximation by two step central (one step forward, one step back) finite difference scheme. $df/dx = (f(x+dx) - f(x-dx))/(2*dx)$ **Asymmetric FD** For approximation by two step non-symmetric (one step forward, half step back) finite difference scheme. $df/dx = ((f(x) + f(x+dx))/3 - 4/3 * f(x-0.5dx))/dx$ **Tip:** For higher solution precision, 2 or 3 can be used, but more computational effort is consumed.
Max Failed Evaluations	20,000	>=0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.

Parameter	Default	Range	Description
Use Perturbation size	No	No or Yes	Enables the use of Perturbation Size, otherwise an internal automatic perturbation size is set.
Perturbation Size	0.0001	> 0.0	Defines the size of the finite difference perturbation. For a variable x, with upper and lower bounds (xu and xl, respectively), the following logic is used to preserve reasonable perturbation sizes across a range of variables magnitudes: • If abs(x) >= 1.0 then perturbation = Perturbation Size * abs(x) • If (xu - xl) < 1.0 then perturbation = Perturbation Size * (xu – xl) • Otherwise perturbation = Perturbation Size
Use Inclusion Matrix	No	No With Initial Without Initial	**No**: Ignores the Inclusion matrix **With Initial**: Runs the initial point. The inclusion set and initial point are used to build the initial response surface. **Without Initial**: Does not run the initial point. The inclusion set is used to build the initial response surface.

System Reliability Optimization (SRO)

Searches for designs that satisfy design requirements with a required probability of success for the system as a whole.

When there are multiple reliability constraints, it becomes important to account for the system reliability as a whole, rather than the reliability of individual constraints.

As an example, consider a design with two probabilistic constraints that are each individually 50% reliable. In Figure 150, the system is 25% reliable since the four runs do not include failures.

	A	B	C
1	Constraint 1	Constraint 2	System Level
2	Failed	Okay	Failed
3	Failed	Failed	Failed
4	Okay	Failed	Failed
5	Okay	Okay	Okay
6	50%	50%	25%

Figure 150:

Usability Characteristics

- The reliability assessments are not based on a Most Probable Point (MPP) formulation, but instead is based on Monte Carlo simulations using advanced response surface techniques.
- System Reliability Optimization requires fewer runs than MPP optimizers, such as Sequential Optimization and Reliability Assessment, ARSM-Based Sequential Optimization and Reliability Assessment , and Single Loop Approach.
- Consists of a global search capability.
- Terminates when there are not enough remaining evaluations to complete the next iteration.
- Supports input variable constraints.
- All defined constraints are part of the system level constraint. The reliability of an individual constraint will be imposed when the constraint is defined as random. If the constraint is deterministic, it is only considered at the system level.
- For robust optimizations, the optimization problem is formulated as a multi-objective problem between the nominal objective goal and minimizing the objective's standard deviation. This results presents a family of optimal designs that explore the trade-off between performance and robustness. Use the Optima tab to visualize the trade-off.
- Supports input variable constraints.
- The size of the first iteration is controlled by the sum of the Initial Sampling Points and Local Sampling Points settings. The number of evaluations in subsequent iterations is controlled by sum of the Local Sampling Points and Global Sampling Points settings. All the designs generated in one iteration can be executed in parallel.

Settings

In the Specifications step, change method settings from the Settings and More tabs.

Table 23: Settings Tab

Parameter	**Default**	**Range**	**Description**
Number of Evaluations	200	> 0 Integer	Maximum number of iterations allowed.
System Reliability (%)	0.98	Numeric > 0	Defines the system level reliability constraint.

Parameter	Default	Range	Description
		< 100	
Robust Optimization	No	No or Yes	Defines whether this is a robust optimization or not. **No** Do not use robust optimization. **Yes** Use robust optimization.
System Reliability Tol.	0.1	> =0	The allowable percentage violation on system reliability. The design is acceptable if system reliability is not less than System Reliability (%) minus this value.
On Failed Evaluation	Terminate optimization	Terminate optimization Ignore failed evaluations	Determines how to react to evaluation failures.

Table 24: More Tab

Parameter	Default	Range	Description
Max Failed Evaluations	20000	Integer > 0	When On Failed Evaluations is set to Ignore failed evaluations (1), the optimizer will tolerate failures until this threshold for Max Failed Evaluations. This option is intended to allow the optimizer to stop after an excessive amount of failures.
Initial Sampling Points	50	Integer > 0	Number of initial sample points. **Tip:** Recommended range: 20 - 100
Global Sampling Points	2	Integer > 0	Number of samples required per iteration spread throughout the design space.

Parameter	Default	Range	Description
			Used to estimate global effects, and avoid a local minima. **Tip:** Recommended: 1-10
Local Sampling Points	5	Integer > 0	Number of samples required per iteration localized around current iterate. Used to build an accurate response surface in the vicinity of the iterate. **Tip:** Recommended: 3-10
Monte Carlo Points	1000	Integer > 0	Number of internal Monte Carlo runs used to calculate the reliability. **Tip:** Recommended: 500-10000
Random Seed	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Stop after no Improvement	1000	Integer > 0	Terminate the optimization if the number of iterations without improvement exceeds this value.
Use Inclusion Matrix	No	No With Initial Without Initial	**No** Ignores the Inclusion matrix **With Initial** Runs the initial point. The inclusion set and initial point are used to

Parameter	Default	Range	Description	
				build the initial response surface.
			Without Initial	Does not run the initial point. The inclusion set is used to build the initial response surface.

Xopt (User-Defined Optimization Engine)

In this section, you will find detailed information on how to setup and use the "User" optimization method in HyperStudy.

In HyperStudy, you can link to an external optimizer to use in the optimization approach. This is called a User method.

When you select User method in HyperStudy, HyperOpt is invoked to manage the optimization process using the external optimizer. HyperOpt manages the optimization by providing information on problem definition to the external optimizer, initializing solver runs using input variable values obtained from the external optimizer, and providing results from analyses to the external optimizer. The external optimizer performs the optimization using problem definition and analysis result information provided by HyperOpt. It determines when the optimization has converged and relates this to HyperOpt, which then terminates the study.

In order for a user-defined optimizer to interface with HyperStudy/HyperOpt, four things are required:

1. The user-defined optimizer must be able to interpret the problem definition as exported by HyperStudy, to the `extrnopt.dat` file.

   The format of the `extrnopt.dat` file is:

   ```
   Number_of_variables
   Number_of_constraints
   Variable_number,initial_value,lower_bound,upper_bound
   Variable_number,initial_value,lower_bound,upper_bound
   ...
   Constraint_number,ibound,constraint_value
   Constraint_number,ibound,constraint_value
   ...
   ```

   where:

   Number_of_variables
   : Total number of input variables.
     There must be a "`Variable_number,initial_value,lower_bound,upper_bound`" line for each input variable.

   Number_of_constraints
   : Total number of constraints.
     There must be a "`Constraint_number,ibound,constraint_value`" line for each constraint.

Variable_number
: An integer value.
: Each input variable is assigned a variable number, starting from 1. This number is used to identify the variable.

initial_value
: Initial value for the input variable.

lower_bound
: Lower limit for the input variable.

upper_bound
: Upper limit for the input variable.

Constraint_number
: An integer value.
: Each constraint is assigned a constraint number, starting from 1. This number is used to identify the constraint.

ibound
: Either 1 or -1, this specifies whether the constraint is an upper bound or lower bound constraint, respectively.

constraint_value
: Bound value for the constraint.

2. The user-defined optimizer must provide design point information in a format readable by HyperStudy, in a file named `extrnopt.des`.

The format of the `extrnopt.des` file is:

```
keyword
Variable_number,value
Variable_number,value
...
```

where:

Keyword
: One of design, intermediate and stop. These stand for design point, intermediate point, and converged result, respectively. Intermediate points are design points that are not stored in the iteration history (points which may be required for finite difference calculations).

Variable_number
: An integer value corresponding to the value assigned in the `extrnopt.des` file. This identifies the input variables.

value
: Value of the input variable for the next solver run.

> **Note:** The format of the `extrnopt.des` file can be repeated to handle multiple run points.

3. The user-defined optimizer must be able to read results output by HyperStudy to the `extrnopt.rsp` file.

   The format of the `extrnopt.rsp` file is:

   ```
   Number_of_constraints
   Constraint_number,value
   Constraint_number,value
   ...
   Objective_value
   ```

   where:

   Number_of_constraints
   : Total number of constraints. There must be a "`Constraint_number,value`" line for each constraint.

   Constraint_number
   : An integer value corresponding to the value assigned in the `extrnopt.dat` file. This identifies the constraint.

   value
   : Value of the constraint from the latest solver run.

   Objective_value
   : Value of the objective function from the latest solver run.

   > **Remember:** The optimization is to minimize this value.

   > **Note:** The format of the `extrnopt.rsp` file can be repeated to handle multiple run points.

4. The user-defined optimizer must always try to minimize the objective value given in the `extrnopt.rsp` file.

   Minimize
   : If Minimize is chosen from the HyperStudy interface, the objective is written out as the value of the objective output response.

   Maximize
   : If Maximize is chosen from the HyperStudy interface, the objective is written out with the opposite sign. (i.e, an objective output response value of 100.0 would be written to the `extrnopt.rsp` file as -100.0, and an objective output response of -45.0 would be written to the `extrnopt.rsp` file as 45.0).

   System Identification
   : If System Identification is chosen from the HyperStudy interface, the objective is calculated as:
   Objective=$((\text{value}-\text{target})/\text{target})^2$

An optimizer will only appear in the user-defined pull-down menu if it is registered in the current preference file.

An example of a user-defined optimizer is provided with HyperStudy. This optimizer is called Xopt.

How Xopt Interfaces with HyperStudy

The files `hopt_lock` and `extopt_lock` are opened alternately to switch between the operations of HyperOpt and Xopt. In addition, the files `hopt_run` and `extopt_run` identify if either process is still active.

> **Note:** Since the HyperOpt process is fixed, the user-defined optimizer must be written to work with this process.

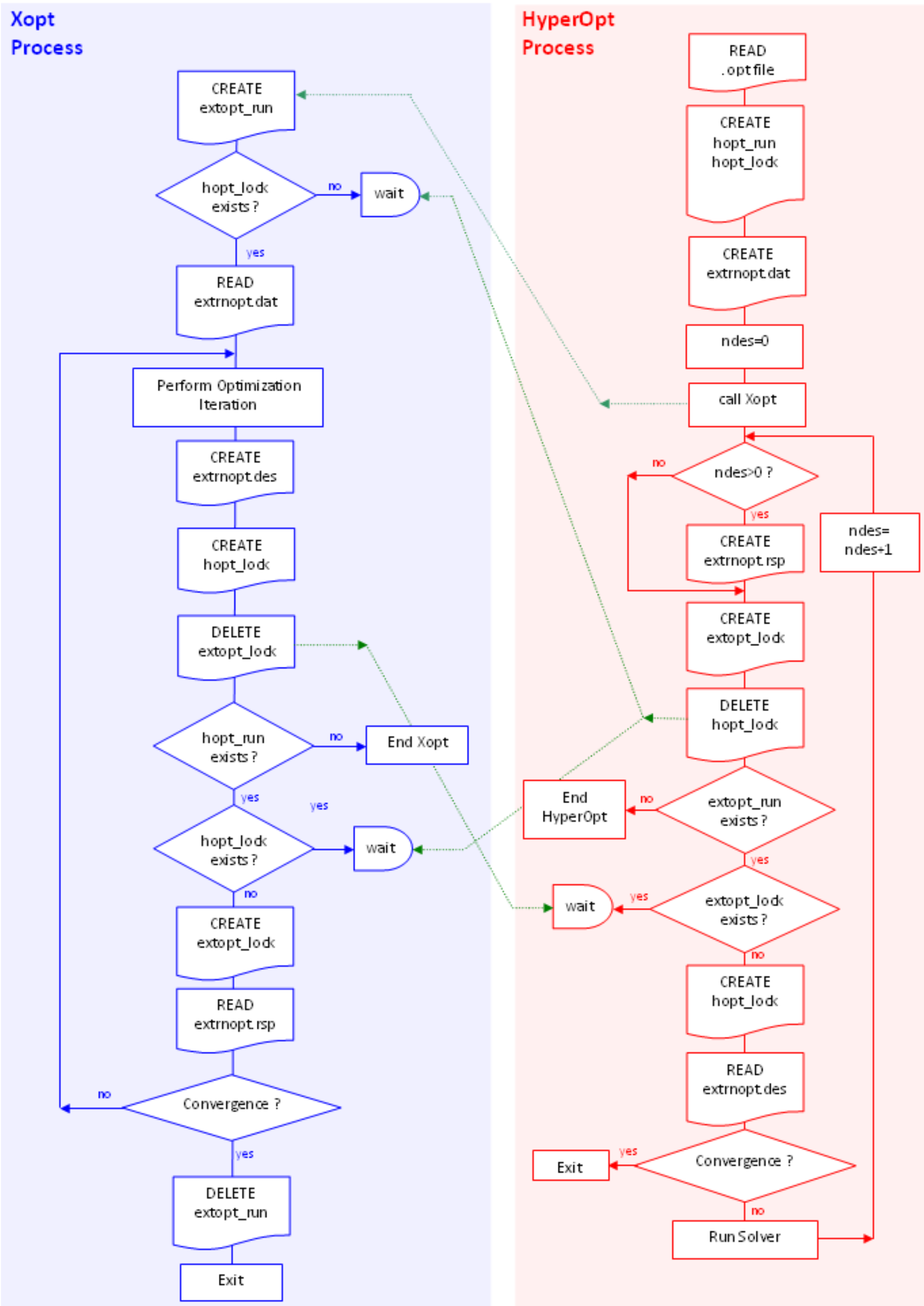


Figure 151: HyperOpt and Xopt Processes

HyperOpt Process

1. HyperStudy invokes HyperOpt.

   Information is provided to HyperOpt on input variables, constraints, and the objective through the `.opt` file.
2. HyperOpt creates the files `hopt_run` and `hopt_lock`.

- hopt_run indicates that HyperOpt is running. This file is automatically deleted if HyperOpt terminates.
- extopt_lock indicates that HyperOpt must wait for Xopt.

3. HyperOpt creates the file extrnopt.dat.

   Xopt will read some key data (number of variables, variable bounds, constraints, etc.) from extrnopt.dat.
4. Set ndes = 0.
5. HyperOpt invokes Xopt.
6. If ndes > 0, create extrnopt.rsp which contains the output response values.
7. HyperOpt creates the file extopt_lock and deletes the file hopt_lock.
8. HyperOpt checks for the existence of extopt_run; if extopt_run does not exist, go to step 14.
9. HyperOpt checks for the existence of extopt_lock.

   If extopt_lock exists then HyperOpt waits, continually checking for the existence of extopt_lock. It continues on to step 10 when extopt_lock no longer exists.
10. HyperOpt creates the file hopt_lock.
11. HyperOpt reads the file extrnopt.des; if the keyword "stop" is found in the file, go to step 14.
12. HyperOpt invokes the solver defined in the HyperStudy interface.
13. ndes = ndes + 1.

    If ndes < MAXDES (maximum number of iterations), go to step 6.
14. The process ends here.

Xopt Process

1. Xopt creates the file extopt_run.

   extopt_run indicates that Xopt is running.
2. Xopt checks for the existence of hopt_lock.

   If hopt_lock exists then Xopt sleeps for 0.1 seconds and repeats step 2.
3. Xopt reads extrnopt.dat which is essential for optimization (number of variables, variable bounds, constraints, etc.).
4. Xopt performs optimization iteration.
5. Xopt writes out the next design point that needs to be analyzed to the file extrnopt.des.
6. Xopt creates the file hopt_lock.

   hopt_lock indicates that Xopt must wait for HyperOpt (see step 9).
7. Xopt deletes extopt_lock.

   This allows HyperOpt to continue.
8. Xopt checks for the existence of hopt_run; if hopt_run does not exist, then Xopt terminates as HyperOpt has been terminated irregularly.
9. Xopt checks for the existence of hopt_lock; if hopt_lock exists then Xopt sleeps for 0.1 seconds and repeats step 9.
10. Xopt creates the file extopt_lock.

`extopt_lock` indicates that HyperOpt must wait for Xopt.

Xopt must remove the file `extopt_lock` in case of process termination. For example, if Xopt is written using Fortran, use of the statement `'dispose = "delete"'` is required.

11. Xopt reads in output responses from the file `extrnopt.rsp`.

12. If the optimization process has not reached convergence, go to step 4.

13. The process ends here.

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

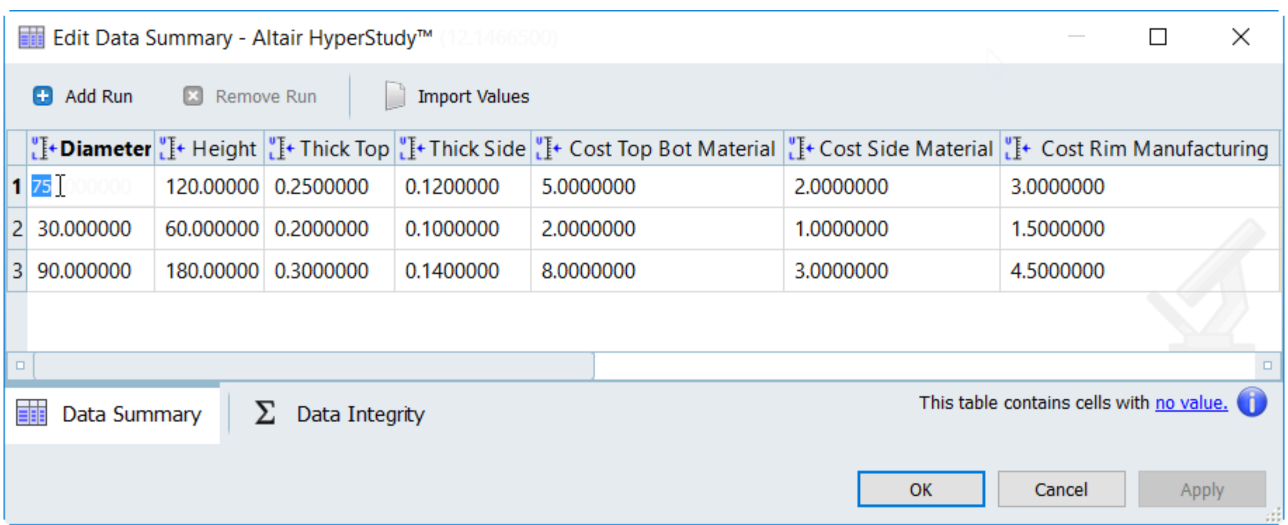


Figure 152:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.

- Manually enter run data.
- Copy and paste run data into the run matrix.

  Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

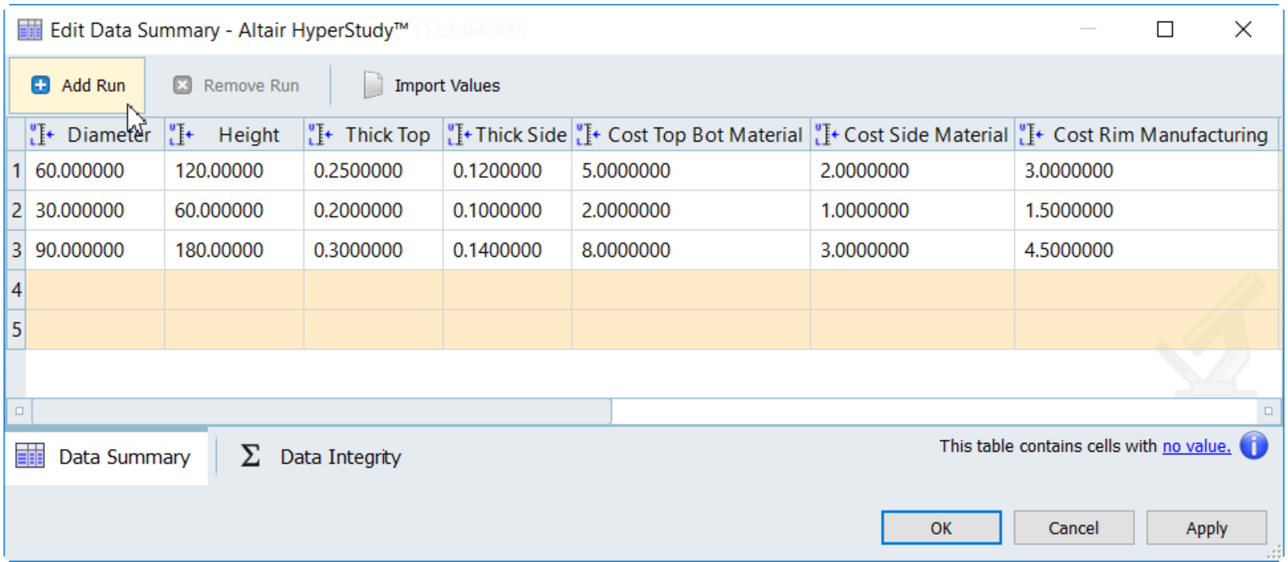


Figure 153:

Tip: Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.

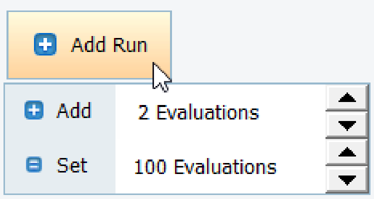


Figure 154:

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.

5. Select the source that contains run data.
    - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
    - For Approach evaluation data, select the approach that contains run data.
    - For HyperStudy post processing file, select the source file.
6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.

Reuse Run Data

An Inclusion matrix contains existing data that will be appended into the newly created approach as known data points. This data typically comes from other approaches, such as DOEs or previously run Optimizations.

In an Optimization, the inclusion can recycle data and in some cases can act as a restart.

1. Go to the **Specifications** step for the Optimization.
2. In the top, right of the work area, click **Edit Matrix** > **Inclusion Matrix**.
3. In the **Edit Inclusion Matrix** dialog, click **Import Values**.
4. In the **Import Values** dialog, select **Approach evaluation data** and click **Next**.
5. For Approach evaluation data, select the approach that contains run data.
6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to ColumnAssignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.
9. Review the imported run data.
10. Click **OK**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.
2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.
3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

   By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

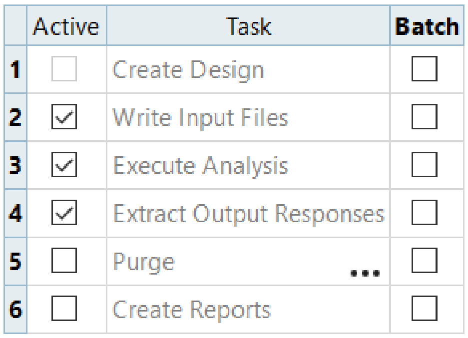


Figure 155:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to stop the submission of new jobs. Click ≡ to select an execution mode.

Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

For an Optimization, multi-execution is affected by your choice in method. To learn more, refer to each method listed in Optimization Methods.

Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

- Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
- Horizontal execution mode sequences the write, execute, and extract task for each run independently.

- Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

Optimization Output Files

Output files generated from an Optimization.

<opt_variable_name>.hyperopt

File Creation

This file is created during an Optimization study.

File Location

`<study_directory>/approaches/<opt_variable_name>/<opt_variable_name>.hyperopt`

File Contents

Result	Format	Description
Protocol file	ASCII	Contains all information about the optimization progress and convergence. You can open this file in the text editor of your choice.

Comments

1. The default `opt_variable_name` is `opt_i`, where `i` is the number of the respective optimization study.

<opt_variable_name>.opt

File Creation

This file is created during an Optimization study.

File Location

`<study_directory>/approaches/<opt_variable_name>/<opt_variable_name>.opt`

File Contents

Result	Format	Description
Definition of the optimization problem	ASCII	Contains HyperOpt input data.

Comments

1. The default `opt_variable_name` is `opt_i`, where i is the number of the respective optimization study.

<opt_variable_name>.hstds

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<opt_variable_name>/<opt_variable_name>.hstds`

File Contents

Result	**Format**	**Description**
Run Matrix Data	hstds, binary	Hstds files stores the retained data sources; direct access data using the `.hstds` file is not suggested.

<opt_variable_name>.hstdf

File Creation

This file is created when **Apply** is selected during the Specifications step.

File Location

`<study_directory>/approaches/<opt_variable_name>/<opt_variable_name>.hstdf`

File Contents

Result	**Format**	**Description**
Run Matrix Data	hstdf, binary	Hstdf files store the run data; however, direct access to the data using the hstdf files are not suggested.

<opt_variable_name>.status

File Creation

This file is created during an Optimization study.

File Location

`<study_directory>/approaches/<opt_variable_name>/<opt_variable_name>.status`

File Contents

Result	Format	Description
Status Report	ASCII	Contains the final status report on the optimization convergence.

Comments

1. The default `opt_variable_name` is `opt_i`, where i is the number of the respective optimization study.

Evaluation Parameters

Modify the run environment settings for the Evaluation tasks.

1. From the Evaluation step, click the **Evaluation Parameters** tab.
2. In the Value column, modify settings accordingly.

> **Note:** Review the Effectuation column to determine the scope at which each setting takes effect.

Review Evaluation Results

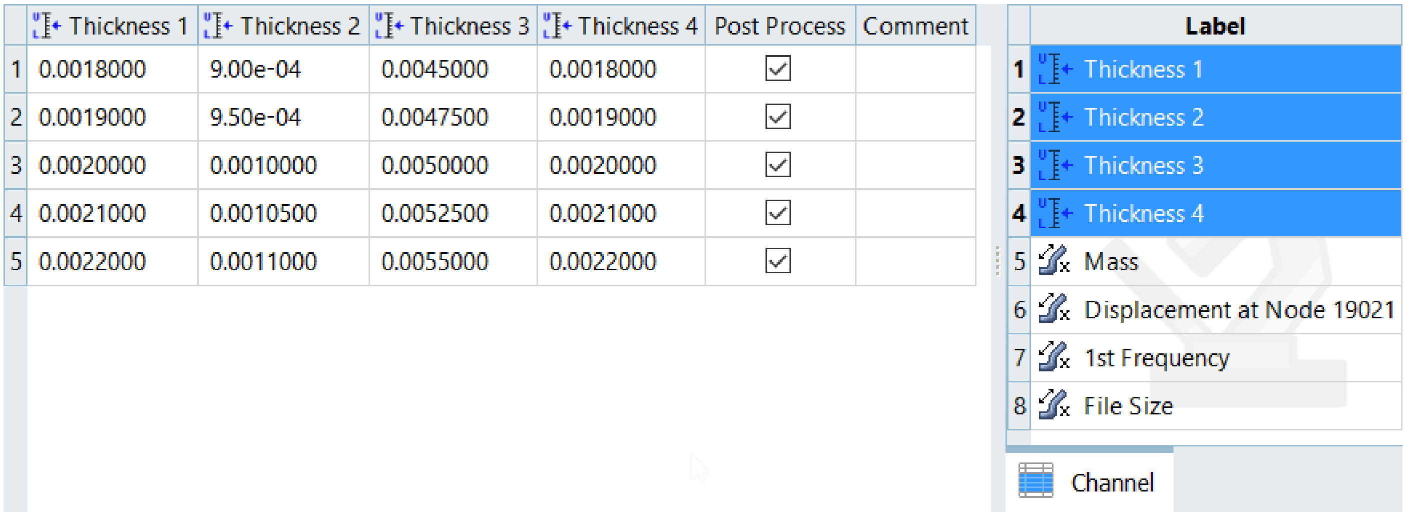
Review the input variable and output response values for each run, as well as review the run files.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Evaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.
4. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

Figure 156:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

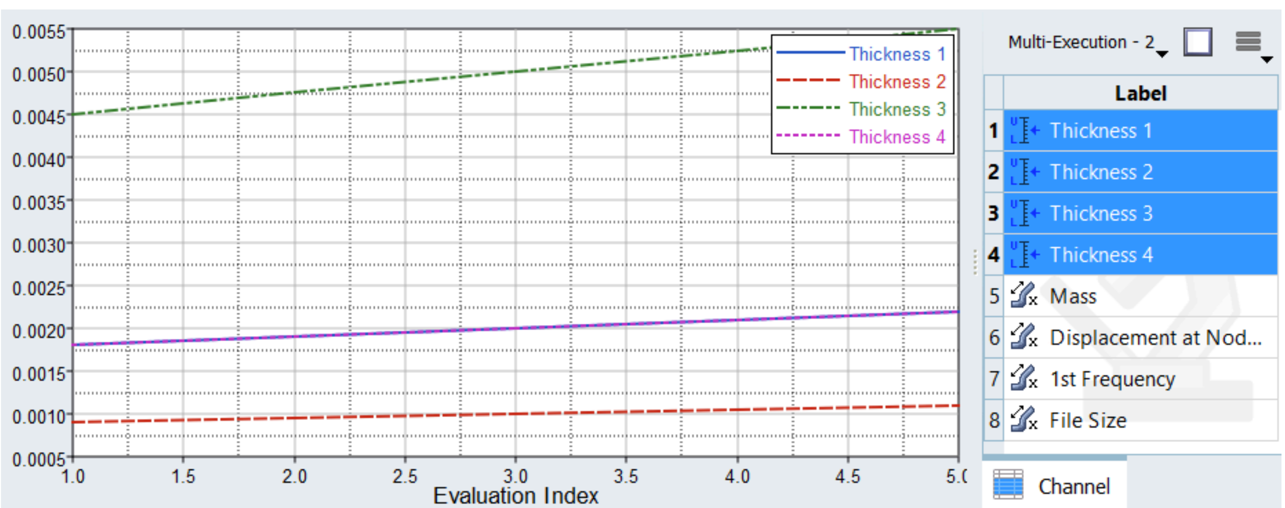


Figure 157:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Evaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

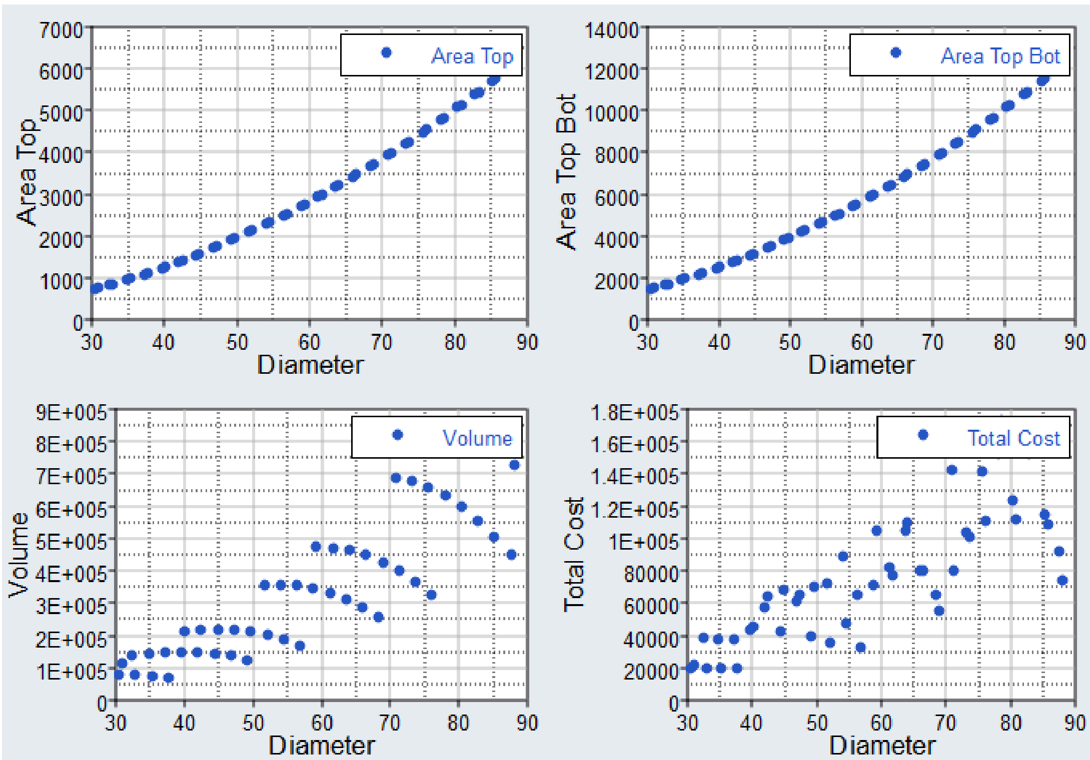


Figure 158:

    - Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

      The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

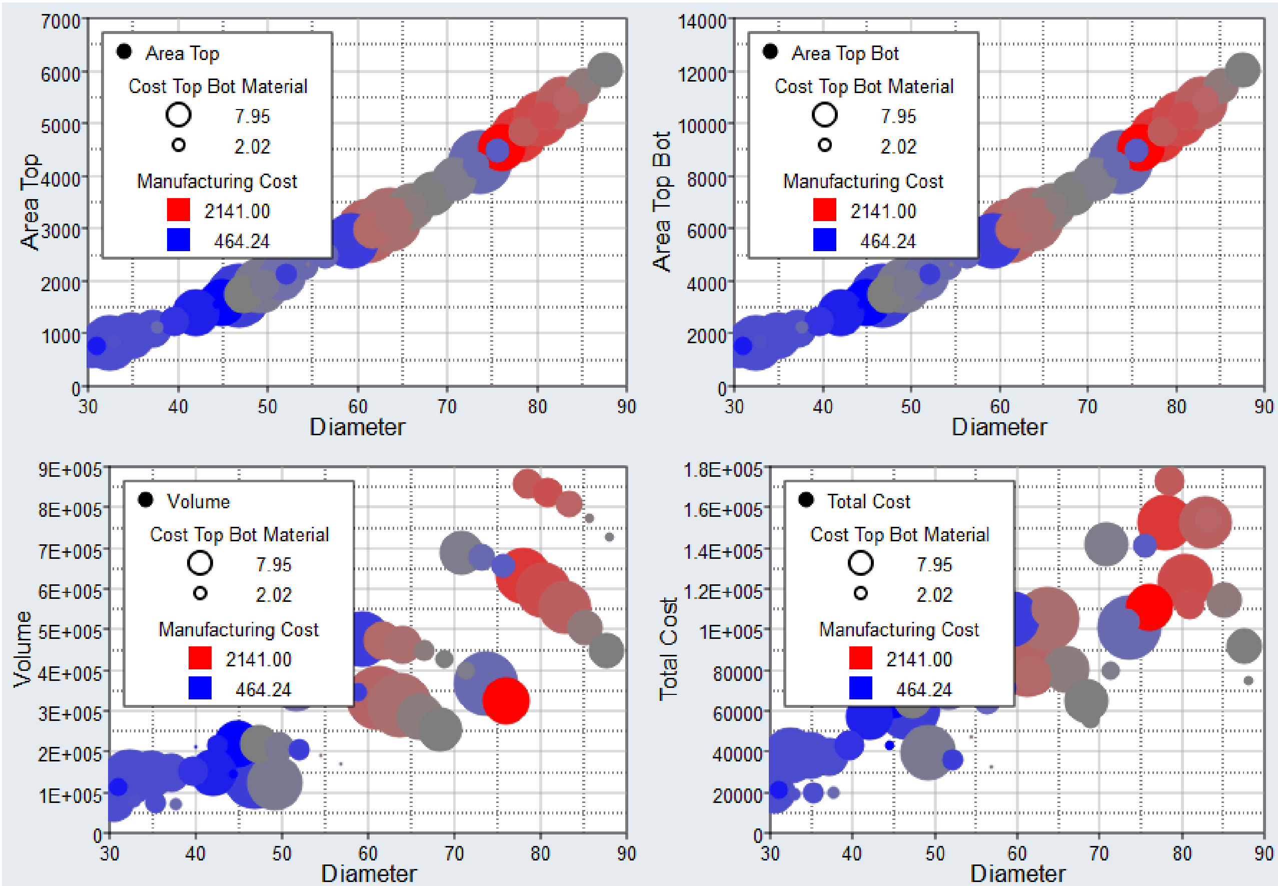


Figure 159:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation Scatter Tab Settings

Settings to configure the plots displayed in the Evaluation Scatter tab.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.

Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

View Iteration History Summary

View a detailed iteration history summary of all input variables and output responses in a tabular format using the Iteration History tool.

1. From the Evaluate step, click the **Iteration History** tab.
2. From the Channel selector, select the channels to display in the table.
3. Analyze the iteration history summary.

Iteration History Table Data

Data reported in the Iteration History table.

General Column Data

Iteration Index
Displays the current iteration number.

Evaluation Reference
Corresponds to the row number in the Evaluation Data table.

Iteration Reference
Displays the iteration number where the current optimal is located.

Condition
Identifies one of three states for the iterate:

- Violated. At least one constraint is violated.
- Feasible. No constraints are violated.
- Acceptable. At least one constraints violated, but only by a very small percentage.

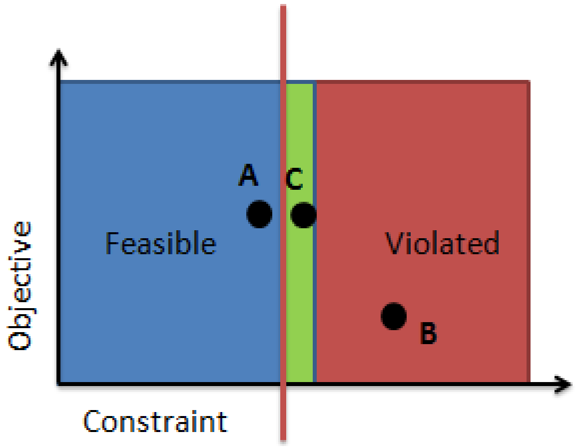


Figure 160: Violation Codes

A minimized objective versus an output response constrained to stay below a bound. Design A is feasible, design B is violated, and design C is acceptable. The plot exaggerates the size of the green zone for visualization, it is typically very small.

Best Iteration

Identifies the set of iterations on the non-dominated Pareto front; this is the optimal set of solutions in a multi-objective optimization.

> **Note:**
> - Only applicable to multi-objective problems.
> - This column is populated once the evaluation is complete

System Identification Column Data

Additional columns are added to the Iteration History table if the objective of your study is system identification.

Objective Function Value

Sum of normalized difference-squared between the objective function value and target value.

DTV

Delta between the target value and the objective value for each objective.

DTVN

Normalized DTV for each objective.

Probabilistic Method Column Data

When using a probabilistic method, additional columns are added to the Iteration History table. The probabilistic method channels are defined in Table 25.

Table 25:

Channel	SORA and SORA_ARSM	SRO
Objective label	The value of the objective at the given input variable values.	The value of the objective at the given input variable values.
Objective label (value at percentile)	The estimated objective value at the specified robust min/max percent CDF.	-
Standard Deviation of Objective	-	The measurement of the spread in the objective distribution.
Constraint label	The value of the constraint at the given input variable values.	The value of the constraint at the given input variable values.
Constraint label (value at percentile)	The estimated constraint value at the constrained CDF limit.	The reliability of design with respect to this constraint.
System reliability	-	The reliability of the system considering all constraints.

The value of the constraint that meets the required reliability.

Consider a case where a constraint value needs to be less than 75.0 with 98% reliability. In the first iteration, Sequential Optimization and Reliability Assessment finds a design with a constraint value of 75.025, but the PV value for 98% reliability is at 99.383. Hence, this design meets the constraint's upper bound, but does not meet the reliability requirement. In the sixth iteration, Sequential Optimization and Reliability Assessment finds a design with a constraint value of 57.412 and the PV value for 98% reliability is at 75.075. This design meets the reliability constraint, as 98% of the design will have a constraint value less than 75.075.

Column Color Coding

The Iteration History table uses color coding to help you determine which designs are feasible, optimal, and violated.

> **Note:** If an iteration contains a violated constraint, the violated constraint is displayed in a bold font.

White Background/Black Font
    Feasible design

White Background/Red Font
    Violated design

White Background/Orange Font
    Acceptable design, but at least one constraint is near violated

Green Background/White Font

Optimal design

Green Background/Orange Font

Optimal design, but at least one constraint is near violated

Analyze Iteration Plot

Plot the iteration history of a study's objectives, constraints, input variables and unused output responses in a 2D chart using the Iteration Plot tab.

1. From the Evaluate step, click the **Iteration Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the iteration history.
3. Analyze the plot.

When the objective history is plotted, infeasible designs are marked with larger markers.

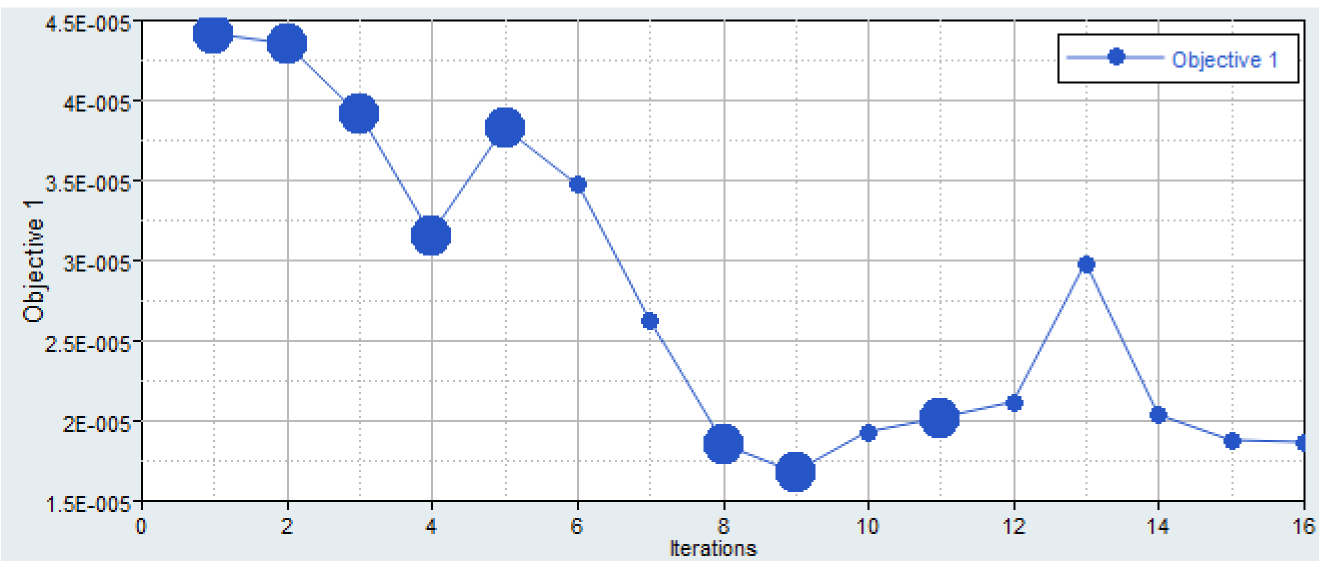


Figure 161: Objective History

When the constraint history is plotted, the constraint bounds can be marked with a datum line.

> **Tip:** To display datum lines, click ≡ (located in the top, right corner of the work area) and enable Bounds.

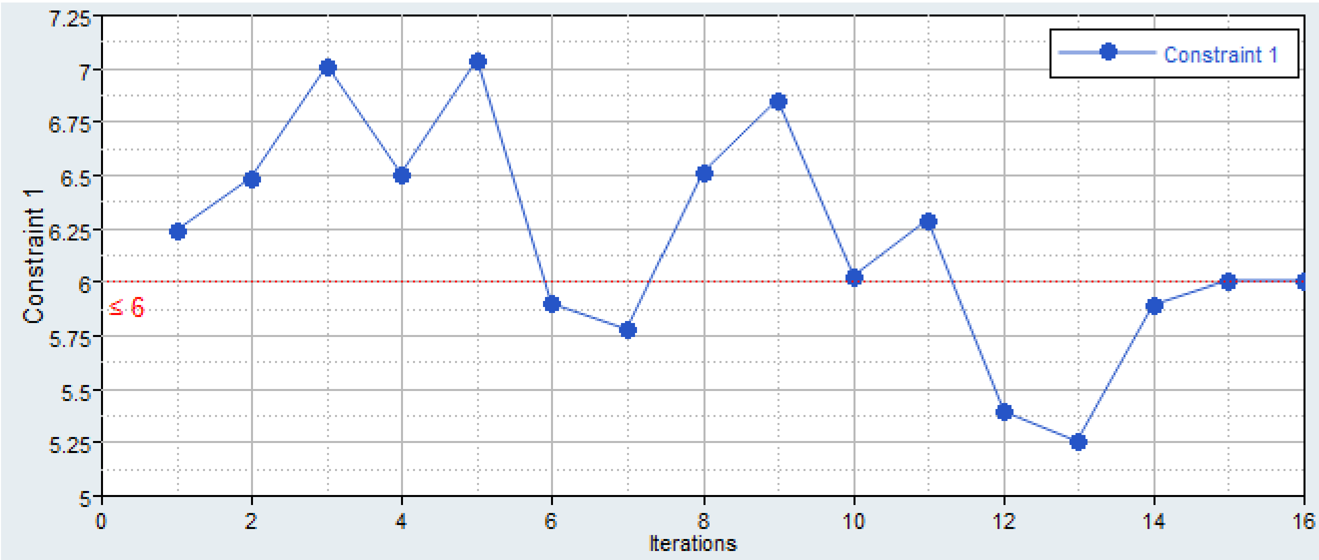


Figure 162: Constraint History

Review Evaluation Time

Inspect task wall-clock times.

Review the time spent in each task within the Evaluation Time tab. Identify bottlenecks in tabular or plot form.

1. From the Evaluate step, click the **Evaluation Time** tab.
2. Use the top channel selector to select the model(s) to review.
3. Use the bottom channel selector to identify the time categorises for review.

Option	Action
Write	Time spent in the write task.
Execute	Time spent in the execute task.
Extract	Time spent in the extract task.
Model Total	Total time of the write, execute, and extract tasks.
All Models Total	Summation of all Model Totals. **Note:** This category is independent of the selected models.

4. Switch the view between table and plot by clicking Table or Plot, located above the Channel selector.

Evaluation Time Settings

Settings to configure the plots and tables displayed in the Evaluation Time tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Cumulative Rows
: Each row entry is a summation of its own wall time and the wall times preceding it with a lower evaluation index.

Plot Time-Unit
: Controls the units of time when plotting the wall times.

Post Processing

View the computational results from the Optimization.

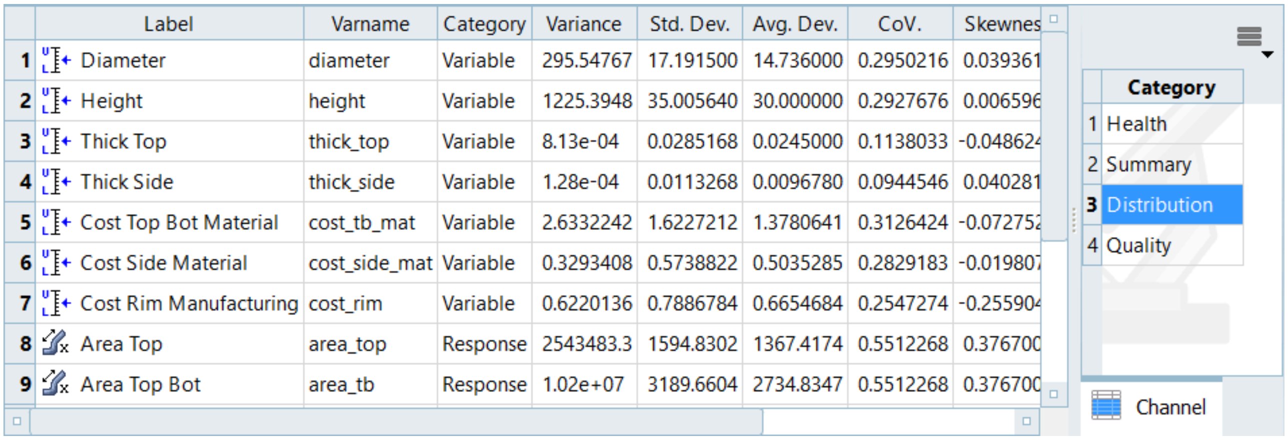
Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
   - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
   - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.
   - **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
   - **Quality** Values typically used in Quality Engineering.



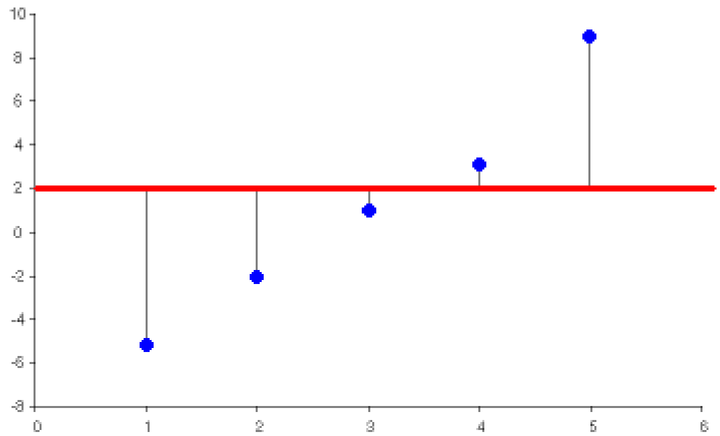
	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.07275
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.01980
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.25590
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

Figure 163:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	Description
Avg Dev (Average Deviation)	Average deviation is evaluated using: $$\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$$ In Figure 164, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 164:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $$CoV = \frac{Standard\ Deviation}{Mean}$$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $\sqrt{\frac{\sum x_i^2}{N}}$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N - 1}$

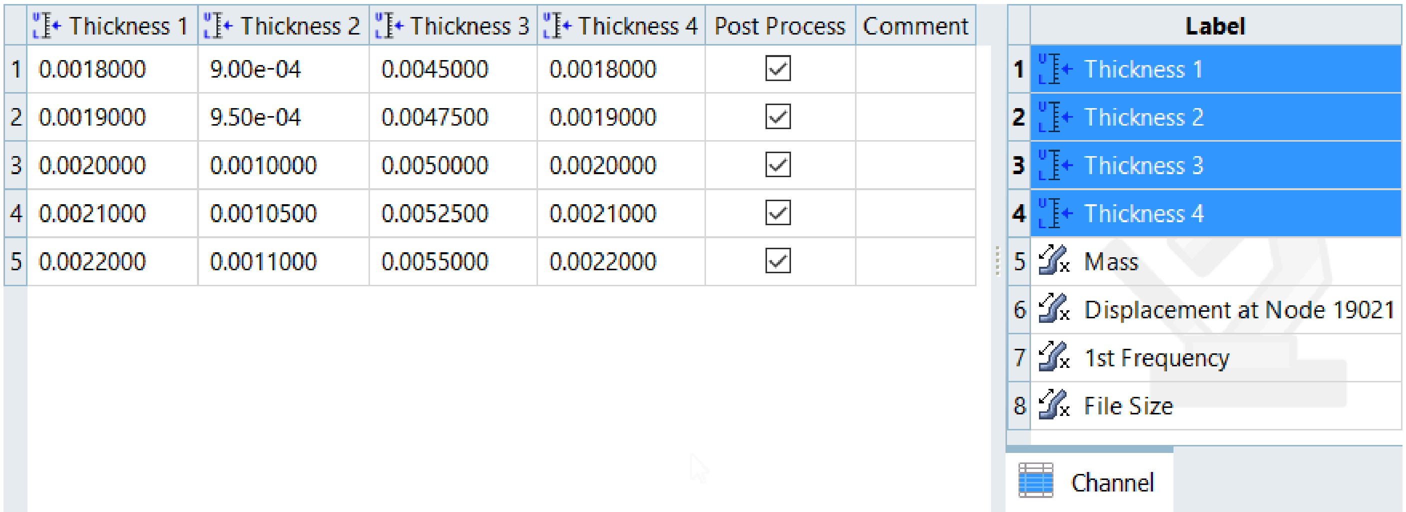
Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processing tab.

1. From the Post Processing step, click the **Summary** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Channel

Figure 165:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot.

Option	Description
	Note: Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 166 run data was selected for a set of lines. In Figure 167, you can see that when Styling is low, Height is high.

Option	**Description**

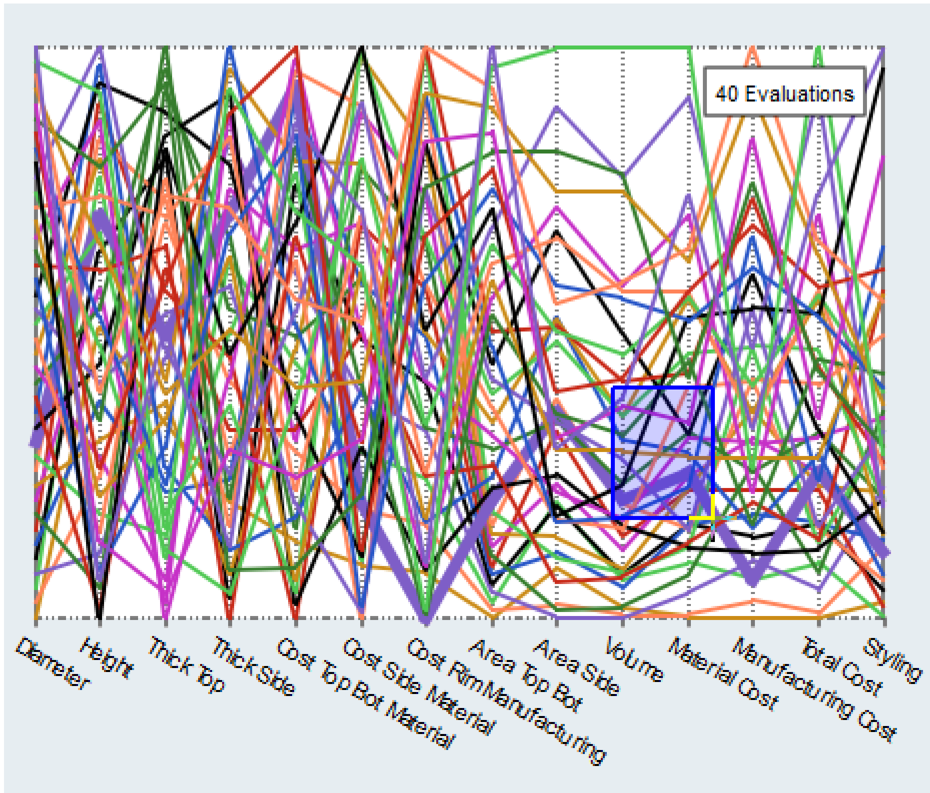


Figure 166:

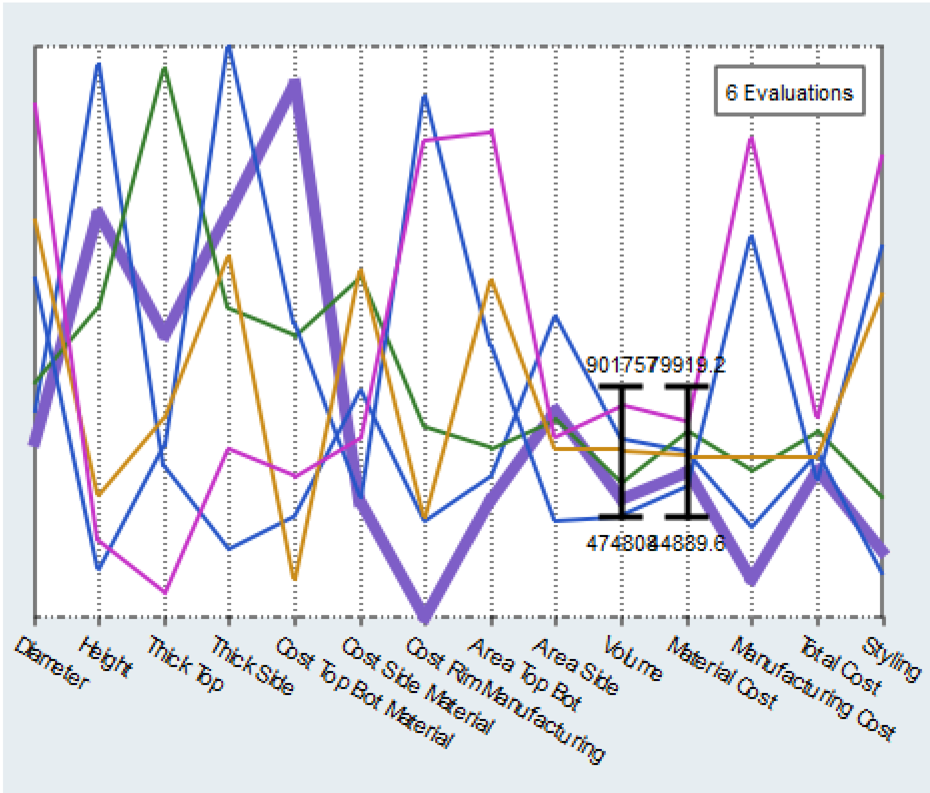


Figure 167:

Option	Description
Filter run data	Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog. • From the Filter column, select the input variables and output responses to plot. • From the Filter Min and Filter Max columns, enter values to filter. The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

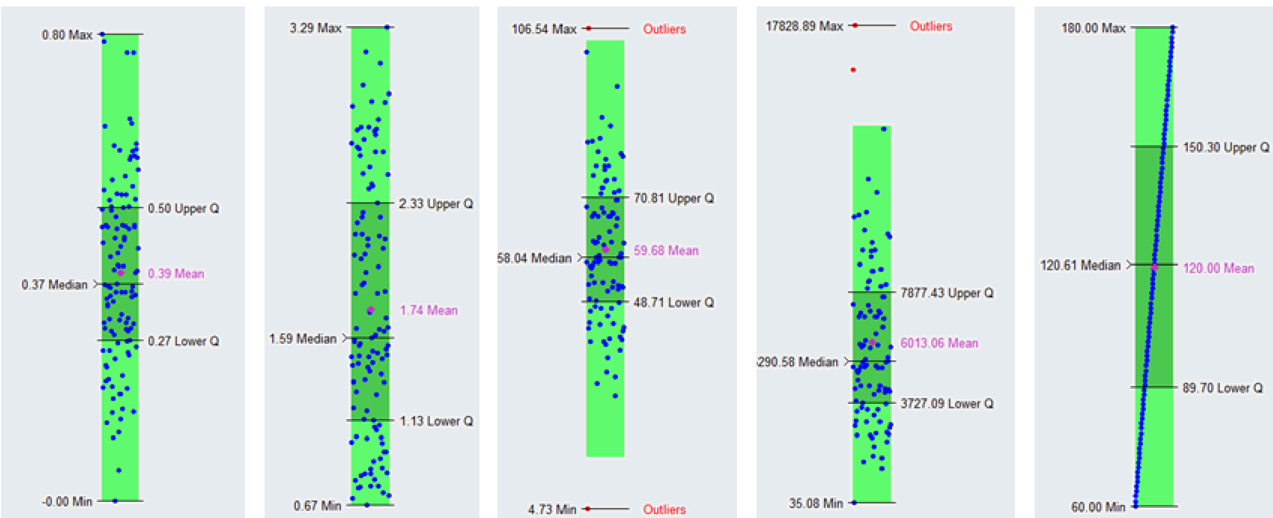


Figure 168:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 169 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

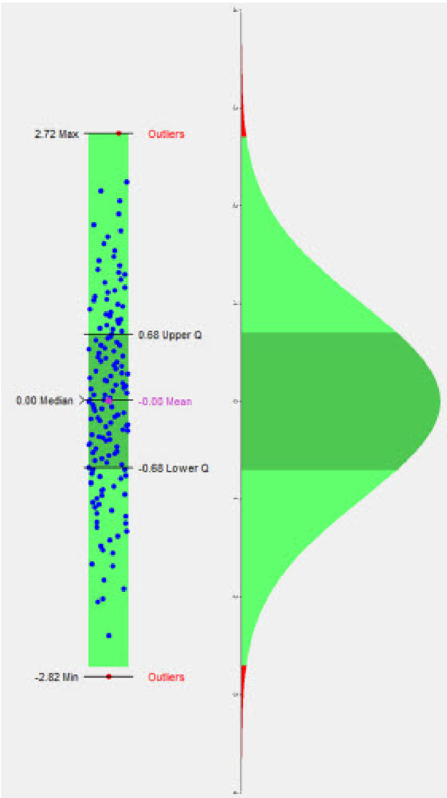


Figure 169:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

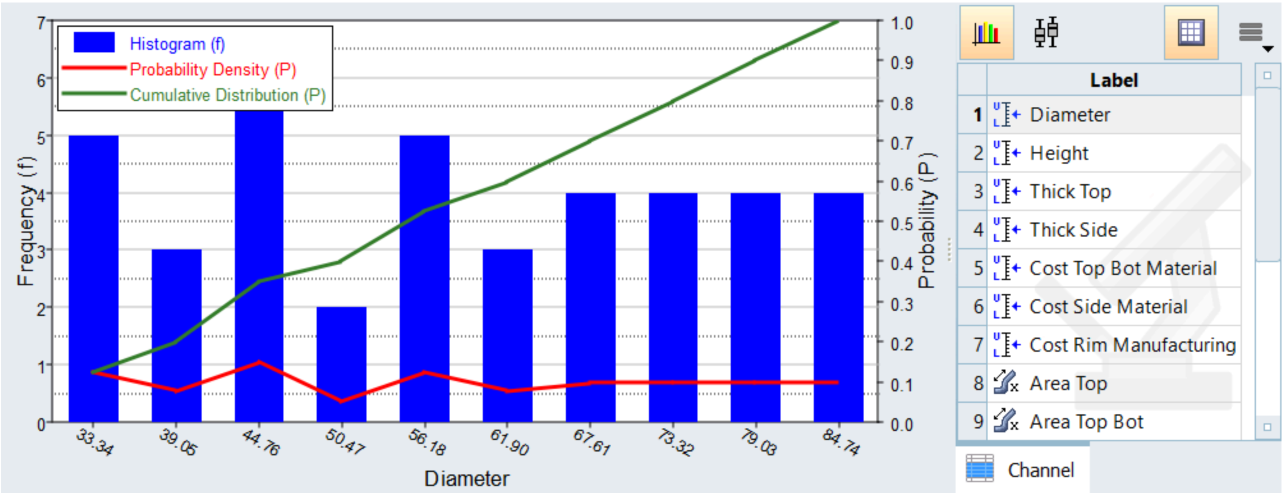


Figure 170:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processing tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Post Processing step, click the **Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

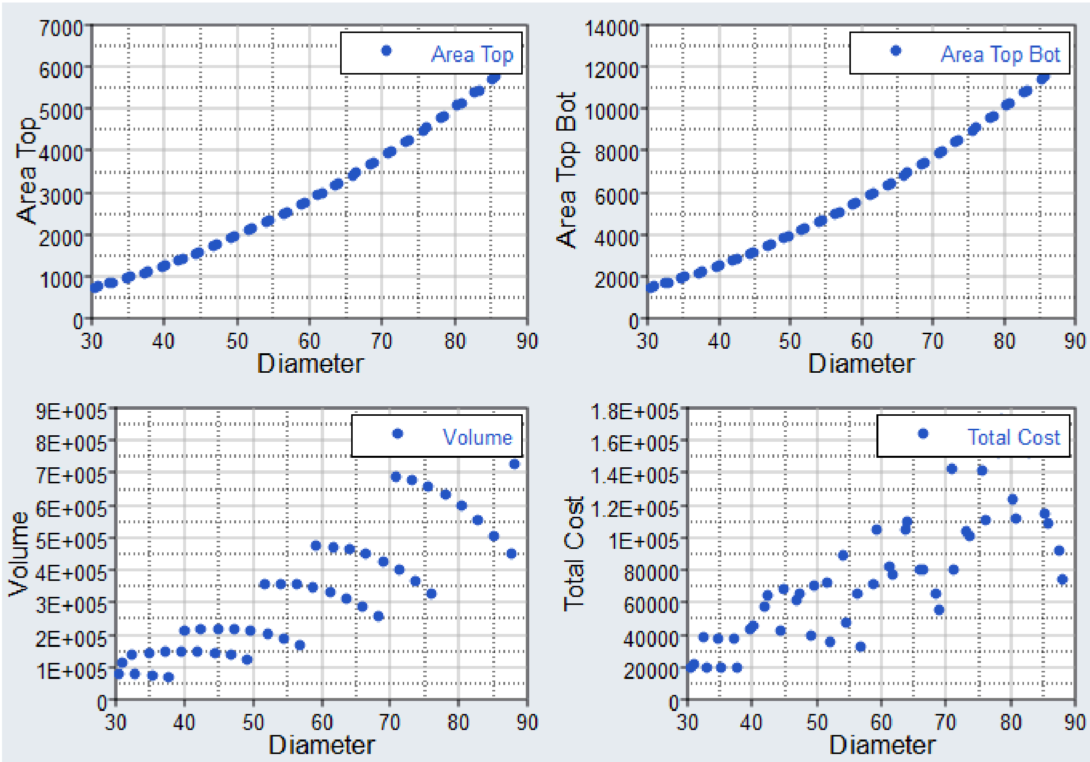


Figure 171:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

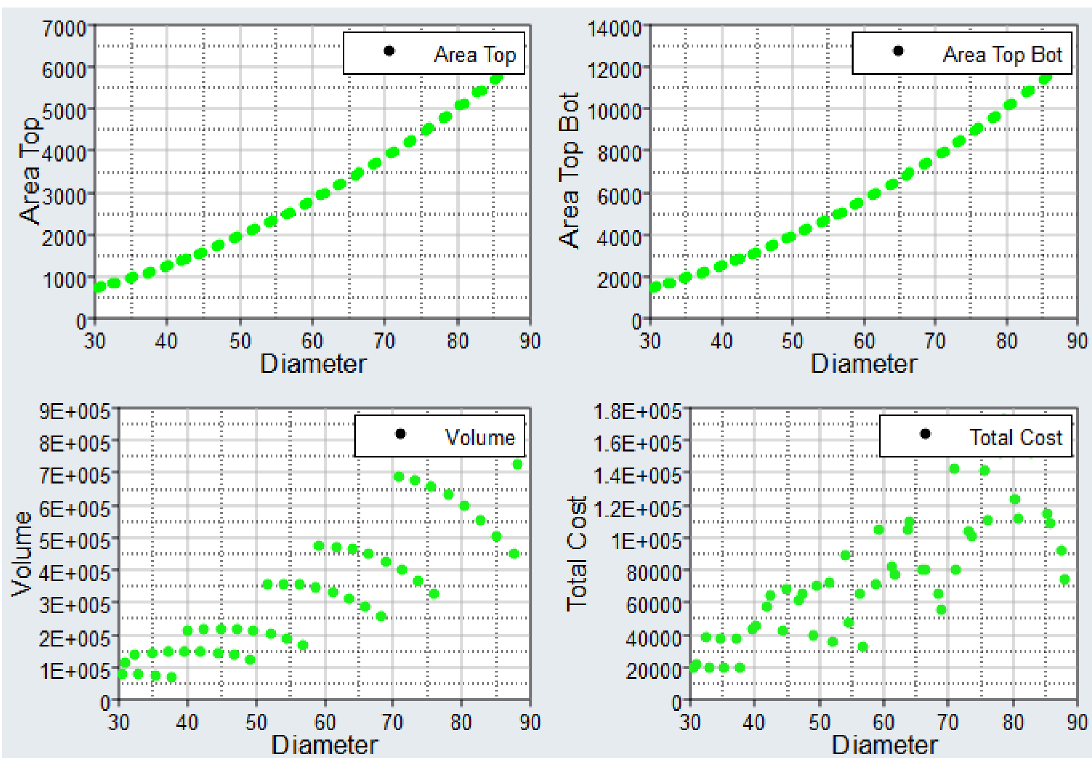


Figure 172:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

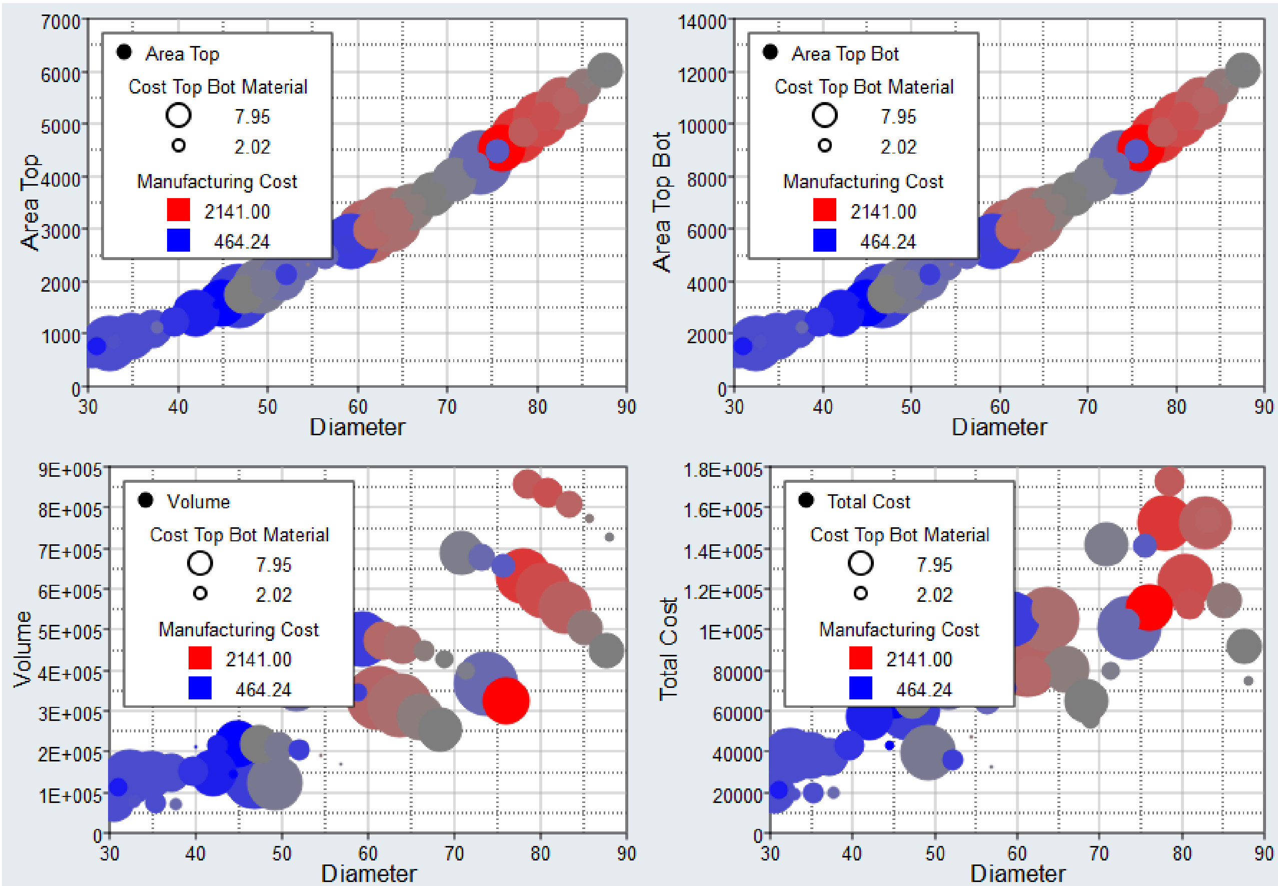


Figure 173:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Scatter Tab Settings

Settings to configure the plots displayed in the Scatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds
Display the X bounds in the plot.

Y-Bounds
Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 174: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 175: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥
Show only the column/rows with cells over the specified threshold.

Show Variables and Responses
Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

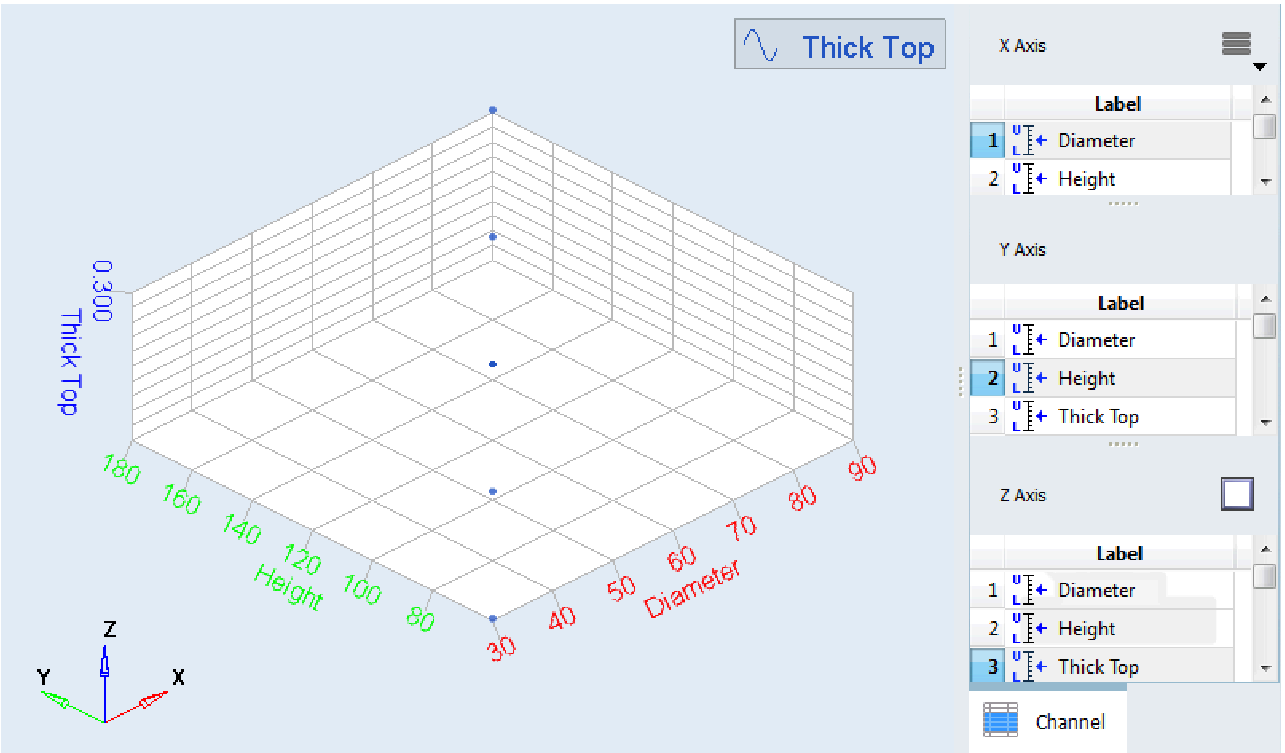


Figure 176:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

> **Tip:** For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the biplot.

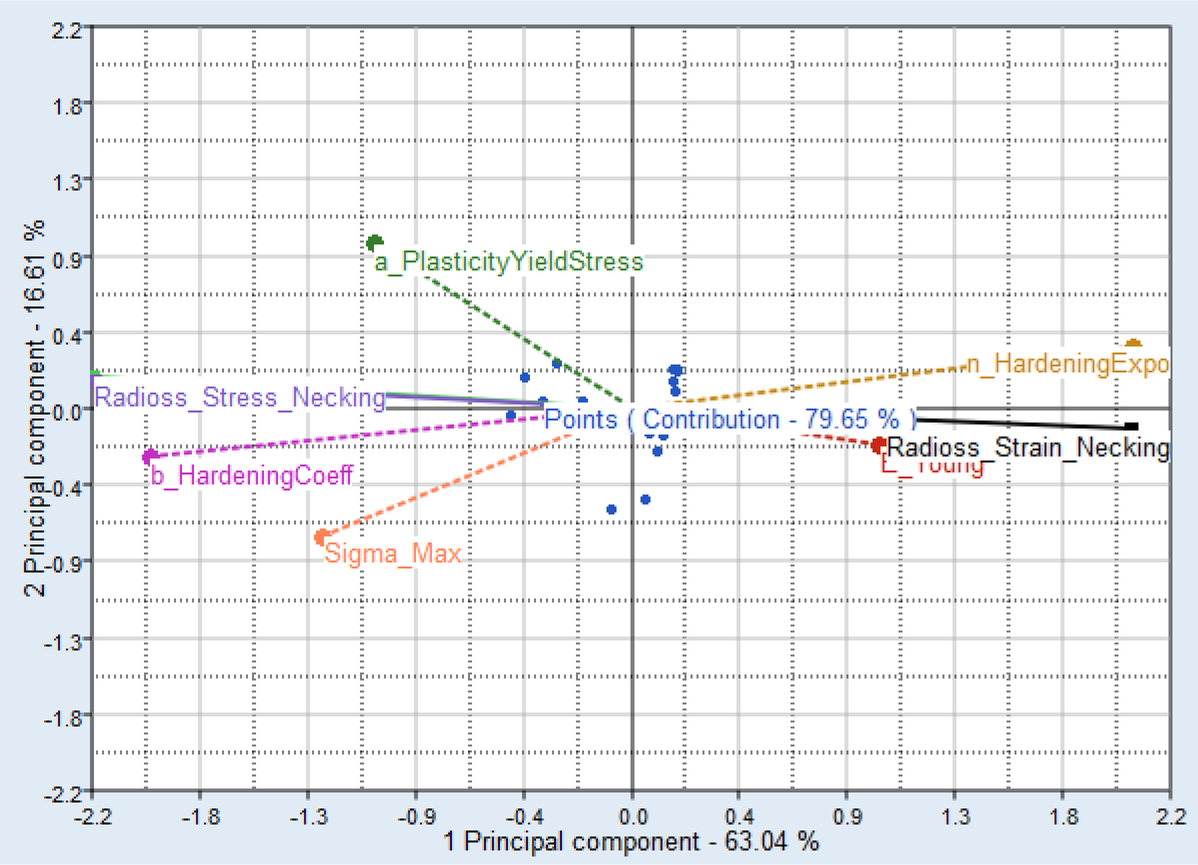


Figure 177:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
   a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

   > **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

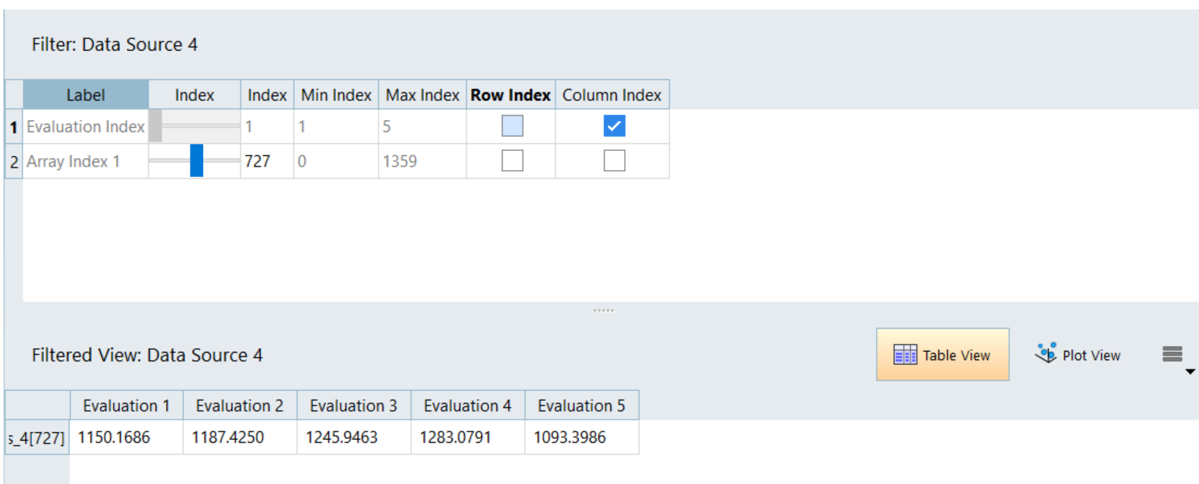


Figure 178:

5. Analyze the table.

Iterations Post Processing

Visualize Optimization scatter history.

Visualize Optimization Scatter History

Visualize the scatter history of an Optimization in the Iterations post processing tab.

1. From the Post Processing step, click the **Iterations** tab.
2. Using the Channel selector, select the data to plot.

a) Under Iteration History, select the iteration(s) of the Optimization.

> **Note:** If you select all the iterations of the optimization problem, the complete iteration history will be plotted accordingly, potentially including iterations with a violated constraint.

b) Under X Axis and Y Axis, select the channels to plot.

Any two channels can be selected for plotting on the x and y-axis.

> **Tip:** For the Y Axis, multiple channels can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the plot.

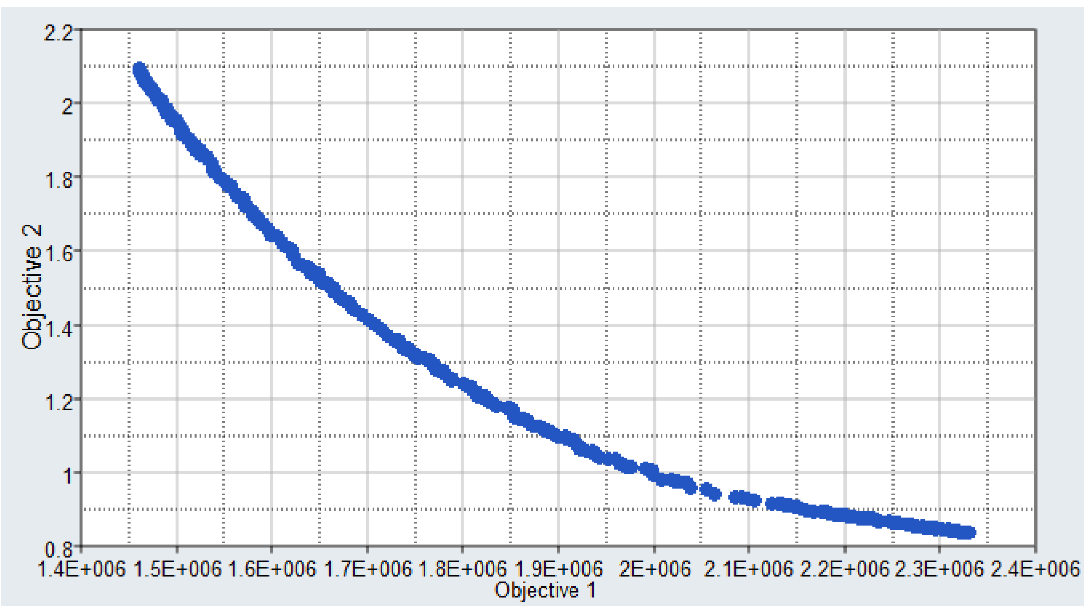


Figure 179:

Optima Post Processing

Visualize Pareto Frontiers.

Visualize Pareto Frontiers

Visualize Pareto Frontiers in the Optima post processing tab. Visually emphasize data in the plot by appending additional dimensions in the form of bubbles.

The Optima tab displays a plot of two different quantities. This is frequently done to plot multiple objectives against each other, which is known as a Pareto Front.

1. From the Post Processing step, click the **Optima** tab.

2. Select data to plot.

- Use the Channel selector to select data to plot along the X and Y axis.

> **Tip:** For the Y Axis, multiple channels can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

3. Visualize the Pareto Frontiers.

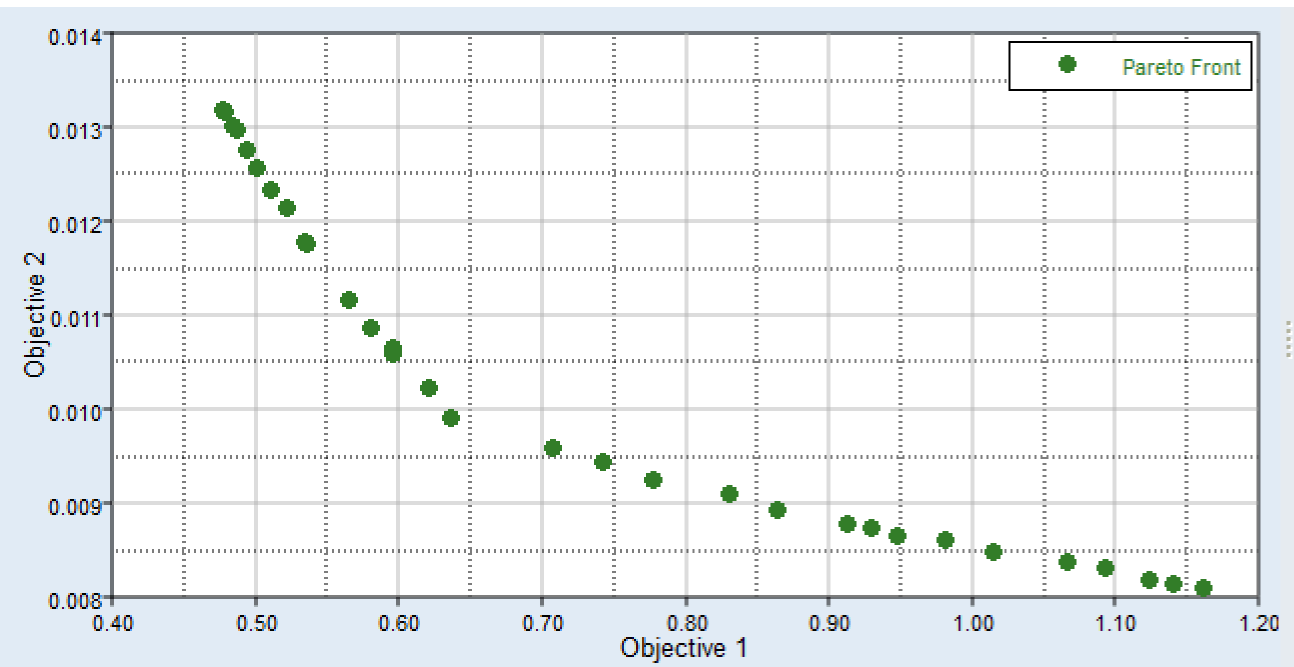


Figure 180:

> **Tip:** Query data on the Pareto Font by clicking **Show Iterations** from the Channel selector. In the **Optimal points** dialog, the row of data selected in the list of optimal points is marked in the original plot to enable you to easily identify the data and run pairs.

Configure the Optima tab's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Optima Tab Settings.

Optima Tab Settings

Settings to configure the plots displayed in the Optima post processing tab.

In the Scatter post tab tool, there are two methods for selecting data to display in the Optima plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size settings:

Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.

Color settings:

Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	Description
HyperStudy Data	Generates a data report (`*.data`).
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.

Report Type	Description
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel.

3. Click **Create Report**.

4.2.4 Setup Sampling Fit Studies

A Sampling Fit is a combination of space-filling DOE method and mathematical model trained by the data generated.

Add a Sampling Fit Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
   a) In the Label field, enter a name for the Sampling Fit.
   b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.

      By default, the Definition defined in the study Setup is selected.
   c) Under Select Type, select Sampling Fit.
   d) Click **OK**.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.
3. Test Models.
4. Define Output Responses.

5. Review definitions in the following ways:

To: **Do this:**

Review status

Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

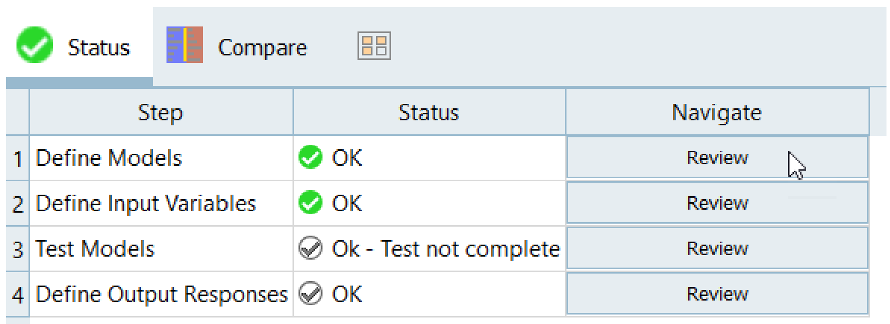


Figure 181:

Compare definitions

Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

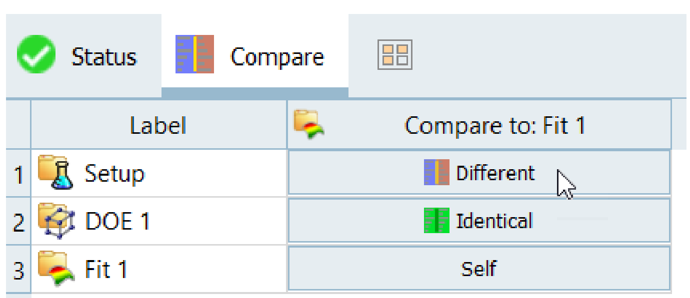
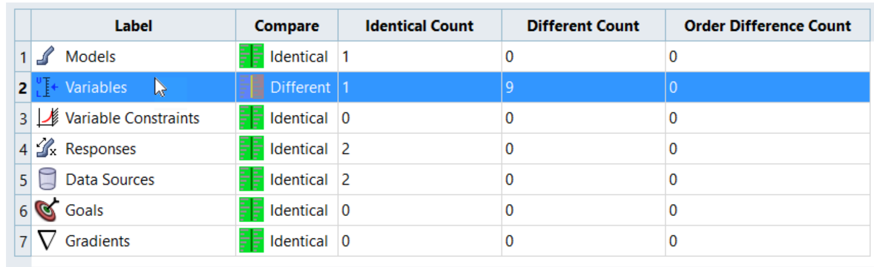


Figure 182:

   The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.
4. Click a channel to display a detailed comparison.

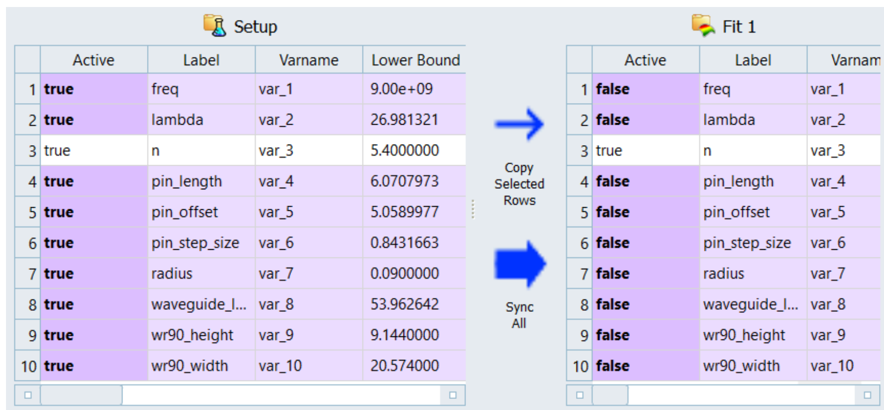
To:	**Do this:**



	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 183:

5. Sync data.
- Click **Copy Selected Rows** to sync the single row.
- Click **Sync All** to sync all rows.



	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 184:

Select a Numerical Method

Select a numerical method to use when evaluating the Sampling Fit.

1. In the Specifications step, go to the Specifications tab.
2. In the work area, Fit Type column, select a numerical method for each output response.
   By default, FAST is selected.
3. Optional: In the Settings tab, change settings as needed.
4. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog.

Fit Methods

Numerical methods available for a Fit approach.

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
Fit Automatically Selected by Training	General	N/A	N/A	Choose methods for Fit Automatically Selected by Training to consider.	Selects the most appropriate method and settings. It it recommended that you use this method unless you desire a specific method and settings.
HyperKriging	Interpolated data	###	##		The time to build the Fit and use the Fit (Evaluate From) increases with both the number of runs and the number of design variables in the input matrix. The number of design variables has more influence than the number of runs if order is larger than 1.
Least Squares Regression	Data trend lines	#	###		Noises can be screened out with this method.

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
					Closed form equations are available.
Moving Least Squares Method (MLSM)	General	##	##		The time to build the Fit and use the Fit (Evaluate From) increases with both the number of runs and the number of design variables in the input matrix. The number of design variables has more influence than the number of runs if order is larger than 1.
Radial Basis Function	Interpolate data	###	##		The time to build the Fit increases with both the number of runs and the number of design variables in the input matrix. The number of runs has more influence than the number of design variables. The run time for using the Fit in another

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
					approach (Evaluate From) is very small regardless of the size of the input matrix.

Fit Automatically Selected by Training

Selects the best available Fit from a list of available methods you have chosen. In addition to selecting the best method, Fit Automatically Selected by Training also automatically adjusts the individual settings (often called hyperparameters) to find the optimizing, predictive performance while avoiding overfitting.

Usability Characteristics

- Fits both noisy and non-noisy data.
- Reduces the methods on which Fit Automatically Selected by Training iterates in order to reduce the run time used to build the Fit.
- Can run in multi-execute, while simultaneously iterating over multiple responses.
- The Stepwise Regression Terms option for Least Squares Regression reduces the number of coefficients in the regression model to contain only the set that is statistically significant.
- The behavior and characteristics of the underlying methods are the same as when the methods are directly applied. See their respective documentation pages for details.
- Gradient information can be used to boost performance for the methods that support gradients.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Least Square Regression	On	On or Off	**On** Use Stepwise Regression Terms to reduce the number of terms in the regression to the statistically significant set. **Off** Do not consider Least Squares Regression in

Parameter	Default	Range	Description
			the ensemble list of methods.
Stepwise Regression Terms	Full Quadratic	Linear Squared Cubic Interaction Full Quadratic Full Cubic	Controls the maximal set of terms considered in stepwise Least Squares Regression. **Linear** First order terms only. y=A+Bx+Cy **Squared** Second order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2 **Cubic** Third order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2+Fx^3+Gy^3 **Interaction** Linear and the cross terms. y=A+Bx+Cy+Dxy **Full Quadratic** Complete second order polynomial. **Full Cubic** Complete third order polynomial.
Moving Least Squares	On	On or Off	**On** Consider Moving Least Squares Method in the ensemble list of methods. **Off** Do not consider Moving Least Squares Method in the ensemble list of methods.
Radial Basis Function	On	On or Off	**On** Consider Radial Basis Function in the ensemble list of methods.

Parameter	Default	Range	Description	
			Off	Do not consider Radial Basis Function in the ensemble list of methods.
Use Gradient Data	On	On or Off	**On**	Allow methods to be enhanced by gradient information when it is available.
			Off	Do not allow methods to be enhanced by gradient information.

HyperKriging

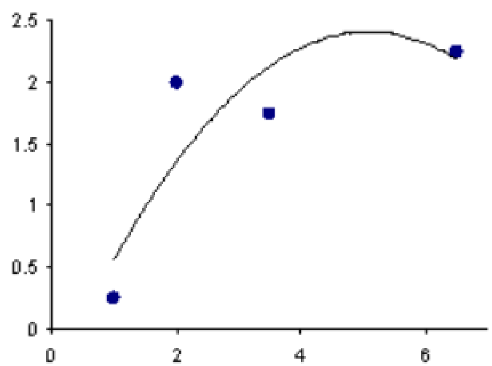
Creates predictive models with data sets coming from deterministic computer simulations, an area of application commonly known as the Design and Analysis of Computer Experiments (DACE).

These experiments are unique because they do not require some concepts such as replication. This approximation method is designed to tightly pass through and smoothly interpolate between the known points.

Usability Characteristics

- Attempts to go through the exact sampling points, and in general, the residuals are small, if not zero. As a result, diagnostic measures using only the complete input matrix do not produce meaningful values. Cross-validation results provide some diagnostics using a special scheme using only the input points. To get detailed diagnostics on the quality of a HyperKriging Fit, it is suggested that you use a testing matrix.
- Suitable for modeling highly nonlinear output response data that does not contain numerical noise.
- Applicability of HyperKriging and Radial Basis Function methods are similar in terms of physics (they both are suggested for highly nonlinear output responses with no noise). It is suggested that you use HyperKriging for large studies that contain a large number of sampling points, whereas, Radial Basis Function is suggested for studies with a large number of variables.

  > **Note:** As a result, Radial Basis Function Fits are evaluated faster than HyperKriging Fits when used in approaches.



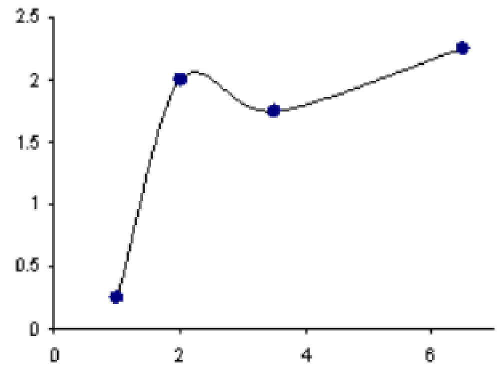


Figure 185: Comparison of a Quadratic Polynomial Model and aHyperKriging Model for a Function of a Single Variable

The plot on the left is a Least Squares Quadratic Regression, and the plot on the right is a HyperKriging model.

Settings

No settings available at this time.

Least Squares Regression

Creates a regression polynomial of the chosen order such that the sum of the squares of the differences (residuals) between output response values predicted by the regression model and the corresponding simulation model is minimized.

For example,

$$\min E = \sum_{i=1}^{n} \left(f_i^{predicted} - f_i \right)^2$$

where n is the number of designs, $f^{predicted}$ is the output response value predicted by the regression model for the ith design, and f is the output response value from the simulation of the ith design. This is achieved by finding the regression model coefficient values that sets the derivative of E, with respect to each unknown coefficient, to zero.

Least Squares Regression Model

The least squares regression model in HyperStudy is the polynomial expression that relates the output response of interest to the factors that were varied.

Selection of the proper model is required to create an accurate approximation. However this requires a prior knowledge of the behavior of the output responses (linear, non linear, noisy, and so on) and enough runs to feed the selected model.

Types of regression models include:

Linear Regression Model

$$F(x) = a_0 + a_1 x_1 + a_2 x_2 + (error)$$

Interaction Regression Model

$$F(x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + (error)$$

Quadratic Regression Model (2nd Order)

$$F(x) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1 2 + a_5 x_2 2 + (error)$$

An approximation is only as good as the uniformity of the design sampling and, for example, a two-level parameter only has a linear relationship in the regression. Higher order polynomials can be introduced by using more levels for the factors, but then, using more levels results in more runs.

If n is the number of input variables:

- A linear regression model requires $n+1$ runs.
- An interaction regression model requires $\frac{(n+1)(n+2)}{2} - n\ runs$.
- A quadratic regression model requires $\frac{(n+1)(n+2)}{2\ runs}$.

Usability Characteristics

- HyperStudy will create the least squares regression of any order, however, in most cases polynomials of the 4th order or higher do not increase accuracy.

  > **Note:** A custom order can be defined from the Regression Terms tab.

- Suppress regression terms that are known to be insignificant.
- Residuals and diagnostics should be used to gain an understanding of the quality of the Fit.
- Quality of a Least Squares Regression Fit is a function of the number of runs, order of the polynomial, and the behavior of the application.
- If the residuals and diagnostics are not good for a Least Squares Regression Fit, than you can increase the order of the Fit provided you have enough runs to fit that specific order.

  > **Note:** If n is the number of input variables:
  >
  > - A linear model requires $n+1$ runs.
  > - An interaction model requires $\frac{(n+1)(n+2)}{2} - n$ runs.
  > - A quadratic model requires $\frac{(n+1)(n+2)}{2}$ runs.

- If increasing the order does not improve the Fit quality, then you may want to inspect the input matrix collinearity and optionally add more runs. You should try the other available Fit methods as your application may have more non-linearity than polynomials can handle.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Regression Model	Linear	Linear Squared Cubic Interaction Full Quadratic Full Cubic Custom	**Linear** First order terms only. y=A+Bx+Cy **Squared** Second order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2 **Cubic** Third order without cross terms. y=A+Bx+Cy +Dx^2+Ey^2+Fx^3+Gy^3 **Interaction** Linear and the cross terms. y=A+Bx+Cy+Dxy **Full Quadratic** Complete second order polynomial. **Full Cubic** Complete third order polynomial. **Custom** User defined order and terms.

Moving Least Squares Method (MLSM)

Builds a weighted least squares model where the weights associated with the sampling points do not remain constant.

Weights are functions of the normalized distance from a sampling point to a point x, where the surrogate model is evaluated. The weight, associated to a sampling point, decays as the evaluation point moves away from it. The decay is defined through a decay function. For each point x it reconstructs a continuous function biased towards the region around that point.

Usability Characteristics

- Suggested to be used for nonlinear and noisy output responses.
- Residuals and diagnostics should be used to gain an understanding of the quality of the Fit.
- Use a Testing matrix in addition to an Input matrix for better diagnostics.

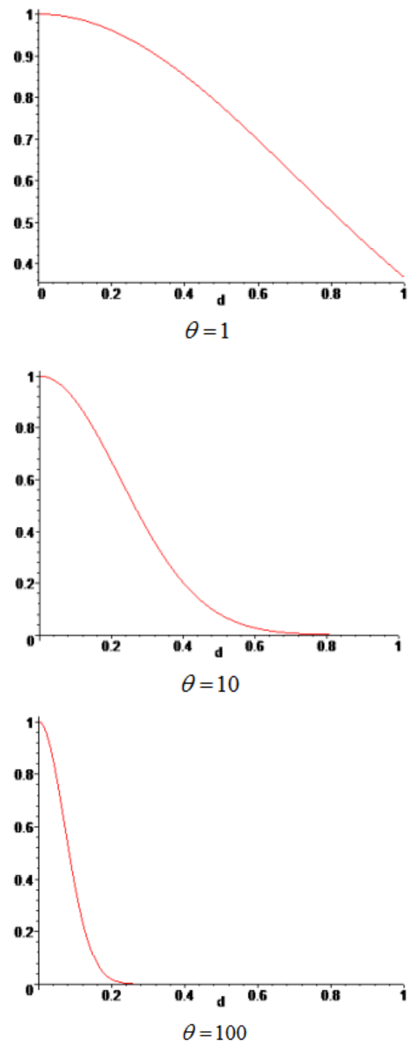
- Quality of a Moving Least Squares Method Fit is a function of the number of runs, order of the polynomial and the behavior of the application.
- If the residuals and diagnostics are not good for a Moving Least Squares Method Fit, than you can increase the order of the Fit provided you have enough runs to fit that specific order.
- Because the weights are not constant in Moving Least Squares Method, there is no analytical form and an equation can not be provided.

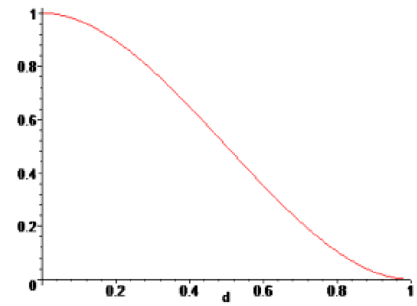
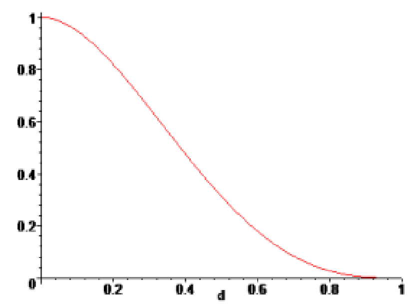
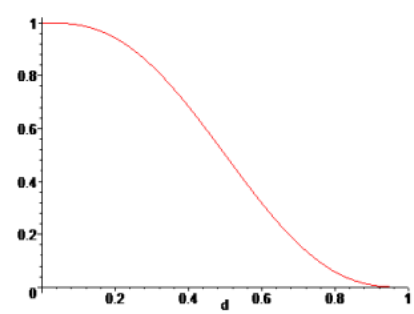
Settings

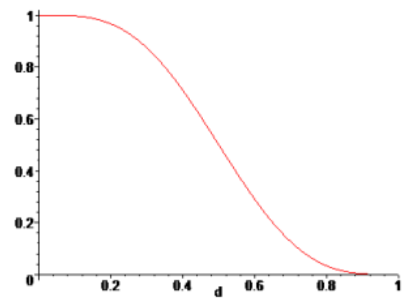
In the Specifications step, Settings tab, change method settings.

> **Note:** For most applications the default settings work optimally, and you may only need to change the Order to improve the Fit quality.

Parameter	Default	Range	Description
Fit Parameter	5.0	>= 0.0 <= 10.0	Controls the effect of screening out noise; the larger value, the less effect.
Minimum Weight	0.001	> 0.0	Minimum weight.
Number of Excess Points	3	>=0	Number of excessive points to build Moving Least Squares Method.
Regression Model	Linear	Linear Squared Cubic Interaction Full Quadratic Full Cubic Custom	Order of polynomial function.
Weighting Function	Gaussian	Gaussian (Recommended Cubic Fourth Order Fifth Order Seventh Order	Type of weighting function. *Gaussian* $W_i = \exp(-\theta r_i 2)$ where r_i is the normalized distance from the i-th sampling point to a current point. The parameter θ defines the "closeness of fit", the case θ=0 is equivalent to the traditional Least Squares Regression. When the parameter θ is large, it is possible to obtain a very close fit

Parameter	Default	Range	Description
			through the sampling points, if desired. The images in Figure 186 illustrate the change of the weight over the interval [0,1] where the sampling point is at r = 0.  *Figure 186:* *Cubic* $w_i = 1 - 3p_i^2 + 2p_i^3$ where $p_i = r_i / R_{\max}$, $R_{\max}$ is the normalized radius of the sphere of influence.

Parameter	**Default**	**Range**	**Description**
			 *Figure 187:* The normalized radius of the sphere of influence $R_{\max}$ inversely relates to the closeness of fit parameter, for example the smaller the value of $R_{\max}$, the closer fit is obtained. *Fourth Order* $w_i = 1 - 6p_i^2 + 8p_i^3 + 3p_i^4$  *Figure 188:* *Fifth Order* $w_i = 1 - 10p_i^3 + 15p_i^4 + 6p_i^5$  *Figure 189:*

Parameter	Default	Range	Description
			Seventh Order $w_i = 1 - 35p_i^4 + 84p_i^5 + 70p_i^6 + 20p_i^7$  *Figure 190:*

Radial Basis Function

Uses linear combinations of basis functions, such as linear, cubic, thin-plate spline, Gaussian, multiquadric, and inverse-multiquadric. These basis functions are observed to be accurate for highly nonlinear output responses but not for linear output responses.

To remedy this deficiency, in HyperStudy, a Radial Basis Function model is augmented with a polynomial function.

$$f(x) = \sum_{i=1}^{n} \lambda_i \phi(\| x - x_i \|) + \sum_{j=1}^{n} c_j p_j(x)$$

where n is the number of sampling points, x is a vector of input variables, x_i is the ith sampling point, $\| x - x_i \|$ is the Euclidean norm, ϕ is a basis function, and λ_i is the coefficient for the ith basis function. $p_j(x)$ is a low-order (constant or linear) polynomial function; k is the total number of terms in the polynomial, and $cj(j = 1, 2..k)$ are the unknown coefficients.

Usability Characteristics

- Attempts to go through the exact sampling points, and in general, the residuals are small, if not zero. As a result, diagnostic measures using only the complete input matrix do not produce meaningful values. Cross-validation results provide some diagnostics using a special scheme using only the input points. To get detailed diagnostics on the quality of a Radial Basis Function Fit, it is suggested that you use a testing matrix.
- Suitable for modeling highly nonlinear output response data that does not contain numerical noise.
- Applicability of HyperKriging and Radial Basis Function methods are similar in terms of physics (they both are suggested for highly nonlinear output responses with no noise). It is suggested that you use HyperKriging for large studies that contain a large number of sampling points, whereas, Radial Basis Function is suggested for studies with a large number of variables.

> **Note:** As a result, Radial Basis Function Fit are evaluated faster than HyperKriging Fits when used in approaches.

Settings

In the Specifications step, Settings tab, change method settings.

> **Note:** For most applications the default settings work optimally.

Parameter	Default	Range	Description
Augmented Function	Constant	Constant Linear Custom	Type of augmented function.
Maximum Points	2000	>=100	Maximum number of points for building Radial Basis Function; if number of building points is larger than maxnpt, then the point reduction algorithm is activated and a warning message is shown; the purpose of introducing maxnpt is to reduce computational effort for large scale problems.
RBF Type	CS21	Multiquadric CS21 (formally knows as Wu's Compactly Supported (2,1)) Gaussian	Type of Radial Basis Function.
Relaxation Parameters	1.0	>=0.0	Relaxation parameter d used in Radial Basis Function; if Radial Basis Function is CS21 or Gaussian, and d is set to 0.0 by users, then Radial Basis Function will automatically set d = 1.0e-6.

Select Matrices

Import and modify the design matrices and associated results for the creation of the approximation model.

The matrix and results should be imported from an existing DOE or Stochastic approach and can be further edited on the fly.

Active matrices are automatically imported.

1. In the Explorer, for the Fit, go to the Select Matrices step.
   A new matrix is created and added to the list of matrices.
2. Define the matrix by modifying its corresponding cells in the work area.
   a) Enter a label.
   b) Select a matrix type.

Input Matrix	Data will be used to create the fit and tune its parameters.
Testing Matrix	Data will be used to assess the quality of the fit.
Input + Testing Matrix	Data can be partitioned into input and testing by specifying the number of runs or percentage.

   c) Select a matrix origin.
   The origin settings names the approach from which the matrix is derived (in the current study).

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

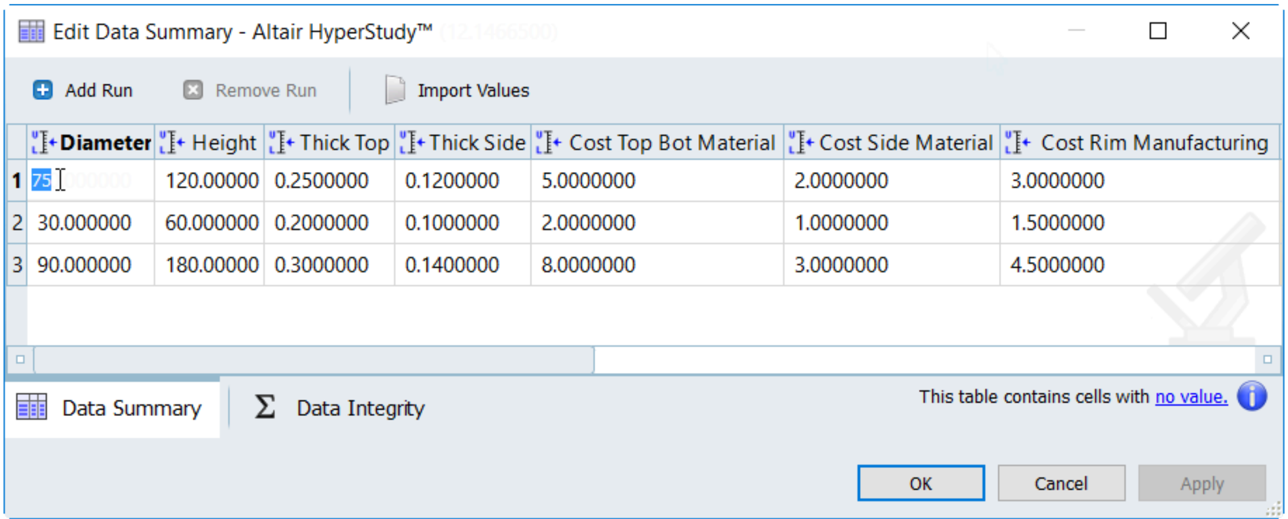


Figure 191:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.
   - Manually enter run data.
   - Copy and paste run data into the run matrix.

     Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

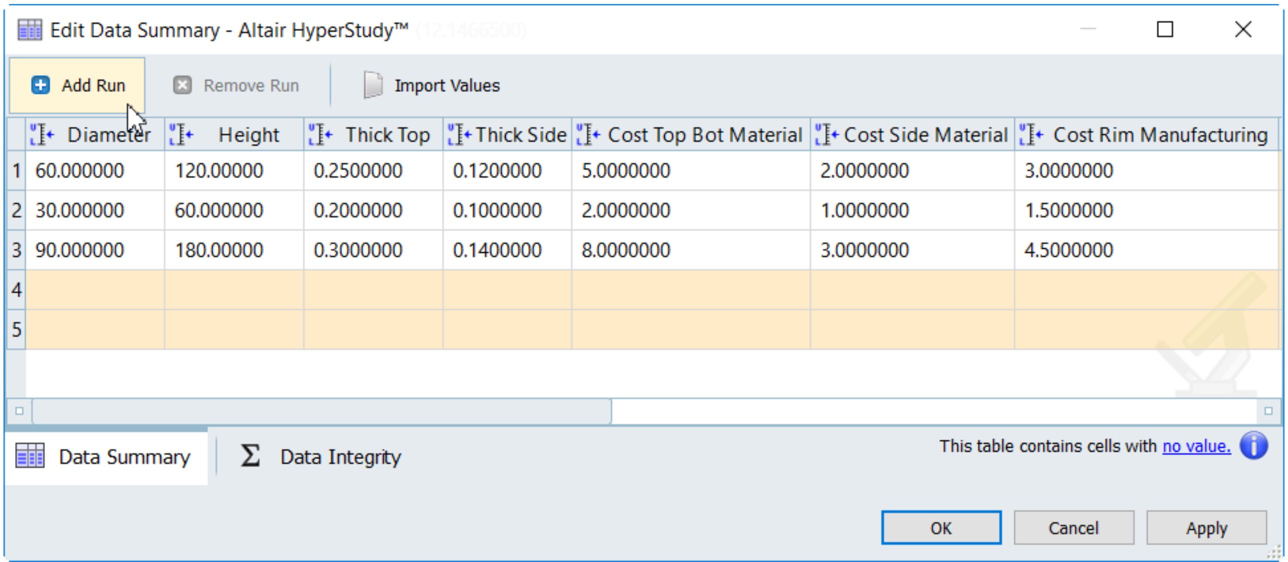


Figure 192:

Tip: Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.

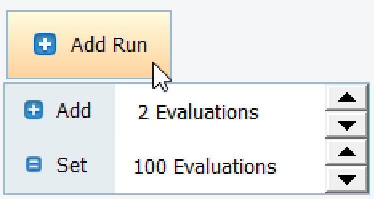


Figure 193:

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.
5. Select the source that contains run data.
   - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
   - For Approach evaluation data, select the approach that contains run data.

- For HyperStudy post processing file, select the source file.

6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.

Filter Run Data

Use the Filter tab to filter the run data included in a Fit approach.

1. Go to the **Fit** > **Specifications** step and select the **Filter** tab.
2. Click the corresponding checkbox to activate a filter.

Filter Name	Filter Description
Filter Outliers	Removes outliers. For more information about outliers, refer to About Box Plots.
Filter Duplicates	Removes duplicate information. **Note:** Filtering duplicates can cause unexpected results when using interpolating fits.
Filter Bad Numbers	Removes runs that contain undefined or unrepresentable data (NaN, inf, and so on).
Filter Excluded	Removes runs that were marked as excluded from post-processing in the source matricies. For more information about editing run matricies, refer to Edit the Run Matrix.

3. Click **Apply**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.
2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.
3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

   By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

	Active	Task	Batch
1	☐	Create Design	☐
2	☑	Write Input Files	☐
3	☑	Execute Analysis	☐
4	☑	Extract Output Responses	☐
5	☐	Purge ...	☐
6	☐	Create Reports	☐

Figure 194:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to stop the submission of new jobs. Click ≡ to select an execution mode.

Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

For an Optimization, multi-execution is affected by your choice in method. To learn more, refer to each method listed in Optimization Methods.

Each evaluation is independent so multi-execute can be used to run in parallel.

A Fit can run in multi-execute while simultaneously iterating over multiple responses.

Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

- Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.

- Horizontal execution mode sequences the write, execute, and extract task for each run independently.
- Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

Sampling Fit Output Files

Output files generated from the a Sampling Fit.

<samplingfit_variable_name>_anova.dat

File Creation

This file is created upon saving the study if Least Squares approximations have been created.

File Location

`<study_directory>/approaches/<samplingfit_variable_name>/<samplingfit_variable_name>_anova.dat`

File Contents

Result	Format	Description
ANOVA	ASCII	The analysis of variance (ANOVA) results are given in table form for each response approximation.

<samplingfit_variable_name>_approximations.slk

File Creation

This file is created upon saving the study if Least Squares approximations have been created.

File Location

`<study_directory>/approaches/<samplingfit_variable_name>/<samplingfit_variable_name>_approximations.slk`

File Contents

Result	Format	Description
Regression	Excel	Spreadsheet that holds the response surface created in the approximation. Use this file to make trade-off studies by modifying the Current Value of each input variable.

Result	Format	Description
		It also contains sensitivity information.

<samplingfit_variable_name>.hstds

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<samplingfit_variable_name>/<samplingfit_variable_name>.hstds`

File Contents

Result	Format	Description
Run Matrix Data	hstds, binary	Hstds files stores the retained data sources; direct access data using the `.hstds` file is not suggested.

<fit_variable_name>.hstdf

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<samplingfit_variable_name>/<samplingfit_variable_name>.hstdf`

File Contents

Result	Format	Description
Run Matrix Data	hstdf, binary	Hstdf files store the run data; however, direct access to the data using the hstdf files are not suggested.

Review Evaluation Results

Review the input variable and output response values for each run, as well as review the run files.

View Cross-Validation R^2 History Summary

View a detailed cross-validation R^2 history summary of all output responses in a tabular format using the R^2 History tool.

1. From the Evaluate step, click the R^2 Plot tab.
2. From the Channel selector, select the channels to display in the table.
3. Analyze the iteration history summary.

	Area Top	Area Top Bot	Area Side	Volume	Mater... Cost	Manu...Cost	Total Cost	Styling	Iteration Index	Evaluation Index
1	-Infinity	-Infinity	-Infinity	-Infinity	-Infinity	-Infinity	-Infinity	-Infinity	1	2
2	0.9478098	0.9478098	-0.7777778	-0.7777778	-0.7777778	0.9753114	-0.7777778	-0.7777778	2	4
3	1.0000000	1.0000000	1.0000000	0.7270779	0.6873946	1.0000000	0.6826707	0.5225883	3	6
4	1.0000000	1.0000000	1.0000000	0.8105061	0.8435154	1.0000000	0.8468060	0.7262046	4	8
5	1.0000000	1.0000000	1.0000000	0.9993626	0.9490636	1.0000000	0.9498065	0.9928450	5	10

Figure 195:

Analyze R^2 Plot

Plot the iteration history of cross-validation R^2 value for each output response in a 2D chart using the R^2 Plot tab.

1. From the Evaluate step, click the R^2 Plot tab.
2. From the Channel selector, select the output response to plot R^2 along the y-axis.

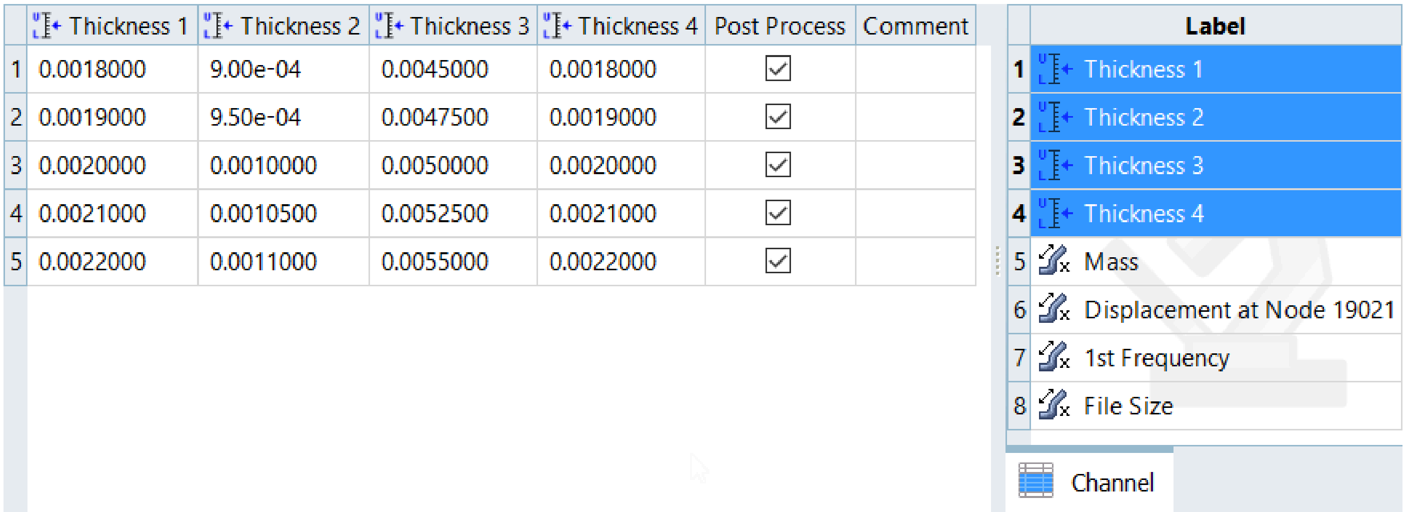
   The x-axis respresents the iteration history.
3. Analyze the plot.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processingEvaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Post Processing step, click the **Summary** tab.
3. From the Channel selector, select the channels to display in the summary table.
4. Analyze the run data summary.
5. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Channel

Figure 196:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

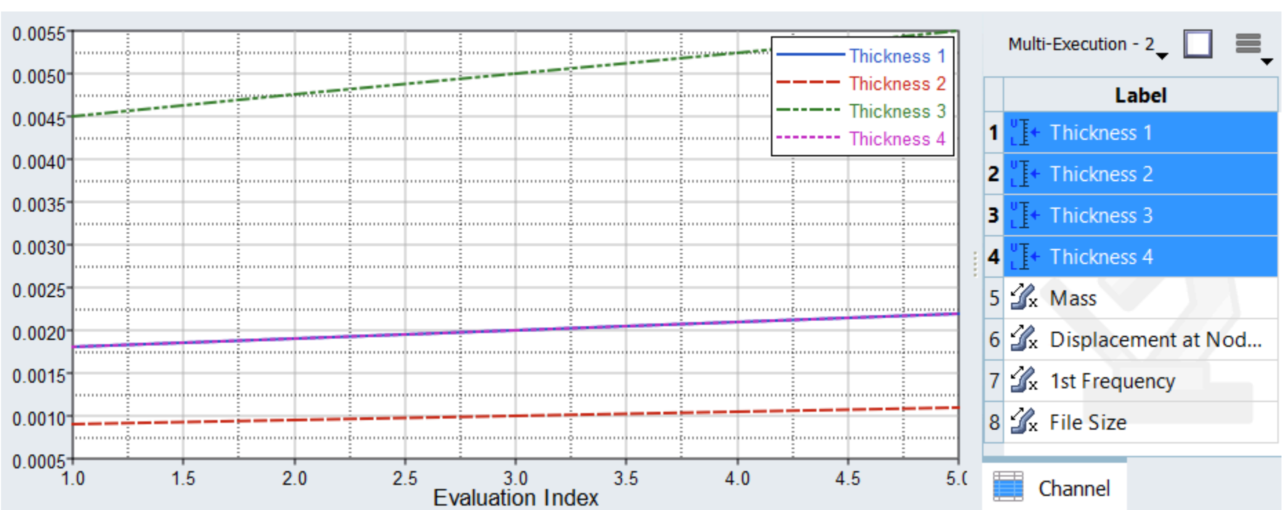


Figure 197:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processingEvaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. From the Post Processing step, click the **Scatter** tab.
3. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

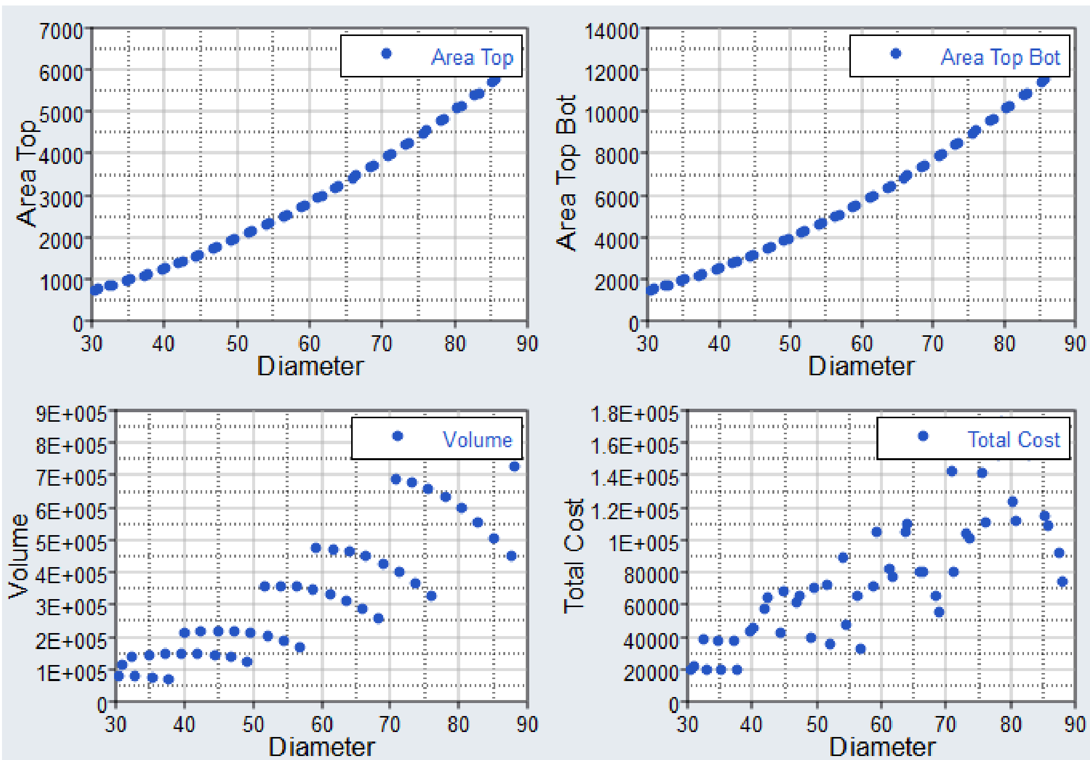


Figure 198:

    - Use the Correlation selector to select one or more values from the correlation map to plot.

      Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

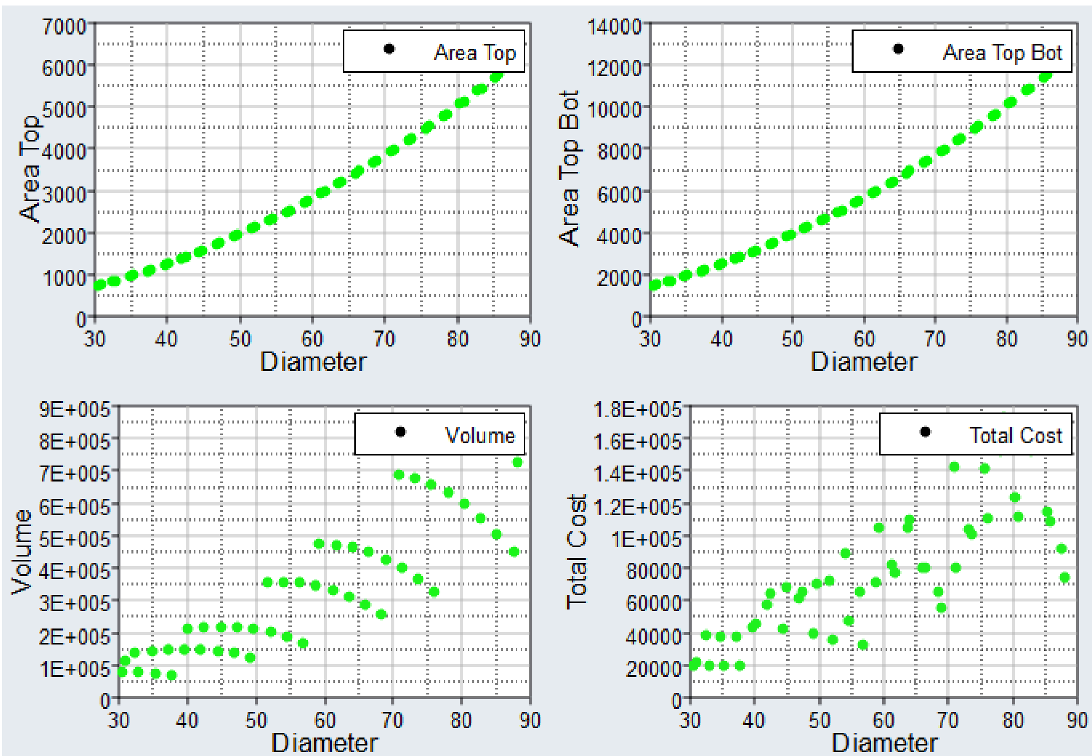


Figure 199:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

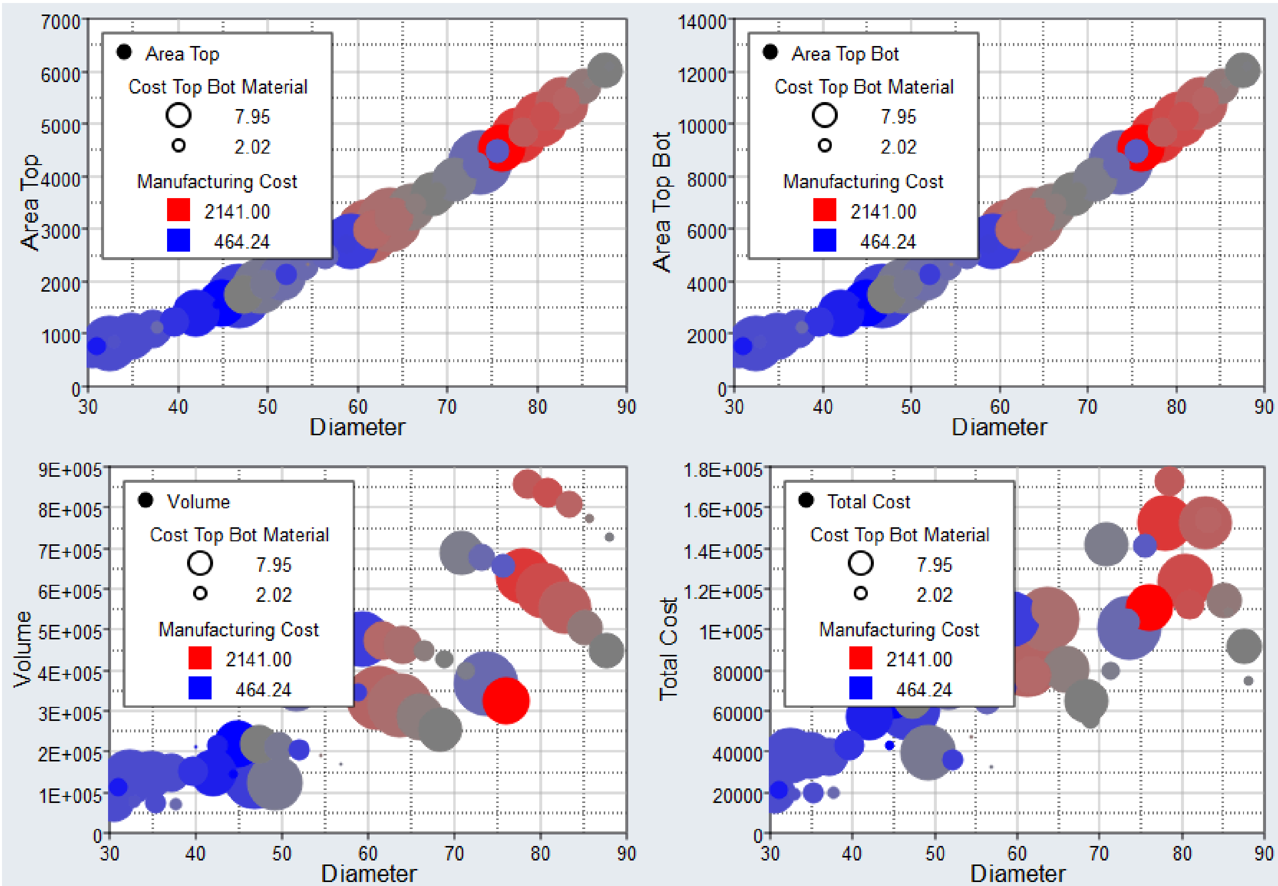


Figure 200:

4. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation ScatterScatter Tab Settings

Settings to configure the plots displayed in the Evaluation ScatterScatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds
Display the X bounds in the plot.

Y-Bounds
Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 201: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 202: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥
Show only the column/rows with cells over the specified threshold.

Show Variables and Responses
Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients	
X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Post Processing

View the computational results from the . approach

Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
    - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
    - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.

- **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
- **Quality** Values typically used in Quality Engineering.

	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.072752
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.019807
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.255904
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

	Category
1	Health
2	Summary
3	Distribution
4	Quality

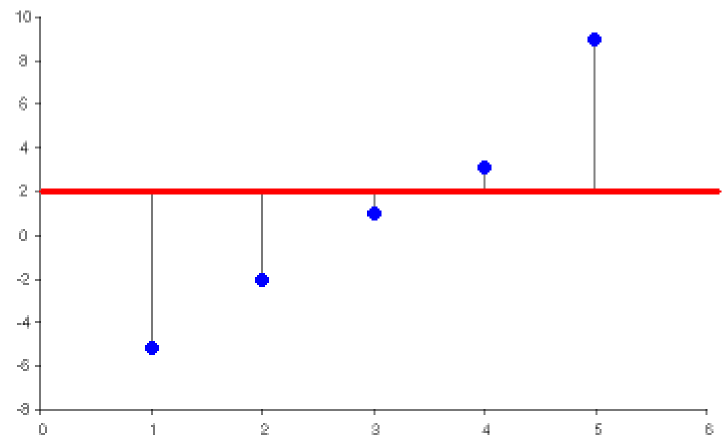
Channel

Figure 203:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	Description
Avg Dev (Average Deviation)	Average deviation is evaluated using: $\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$ In Figure 204, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 204:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $$CoV = \frac{Standard\ Deviation}{Mean}$$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $\sqrt{\frac{\sum x_i^2}{N}}$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$

Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processingEvaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Post Processing step, click the **Summary** tab.
3. From the Channel selector, select the channels to display in the summary table.
4. Analyze the run data summary.
5. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.

	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Channel

Figure 205:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or

Option	Description
	Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot. **Note:** Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 206 run data was selected for a set of lines. In Figure 207, you can see that when Styling is low, Height is high.

Option	**Description**

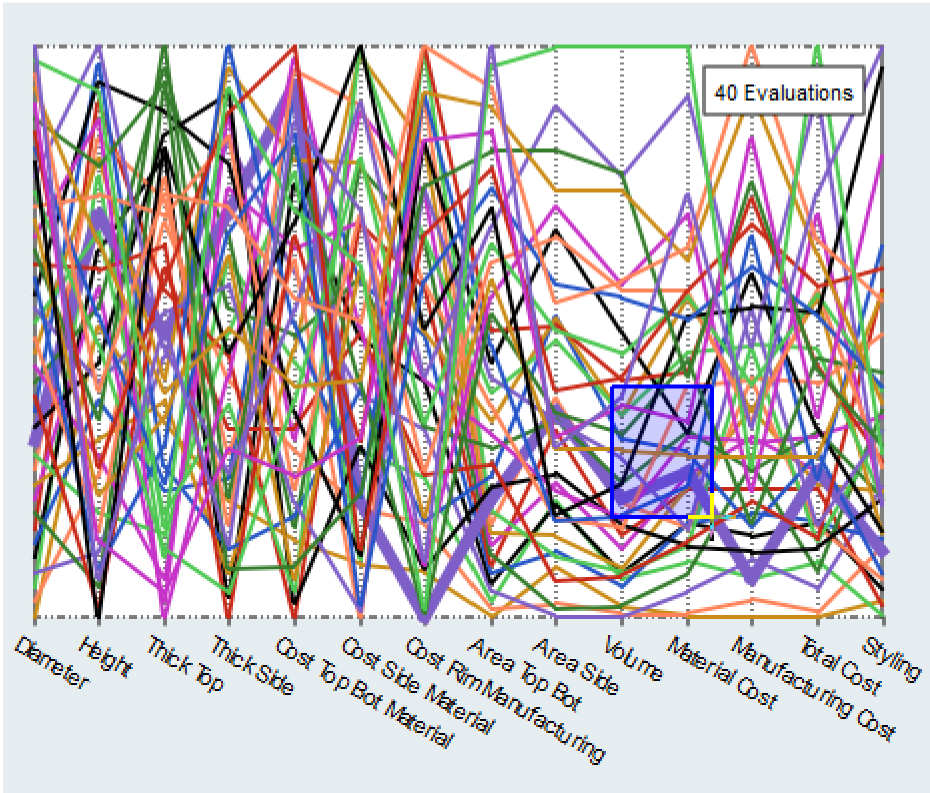
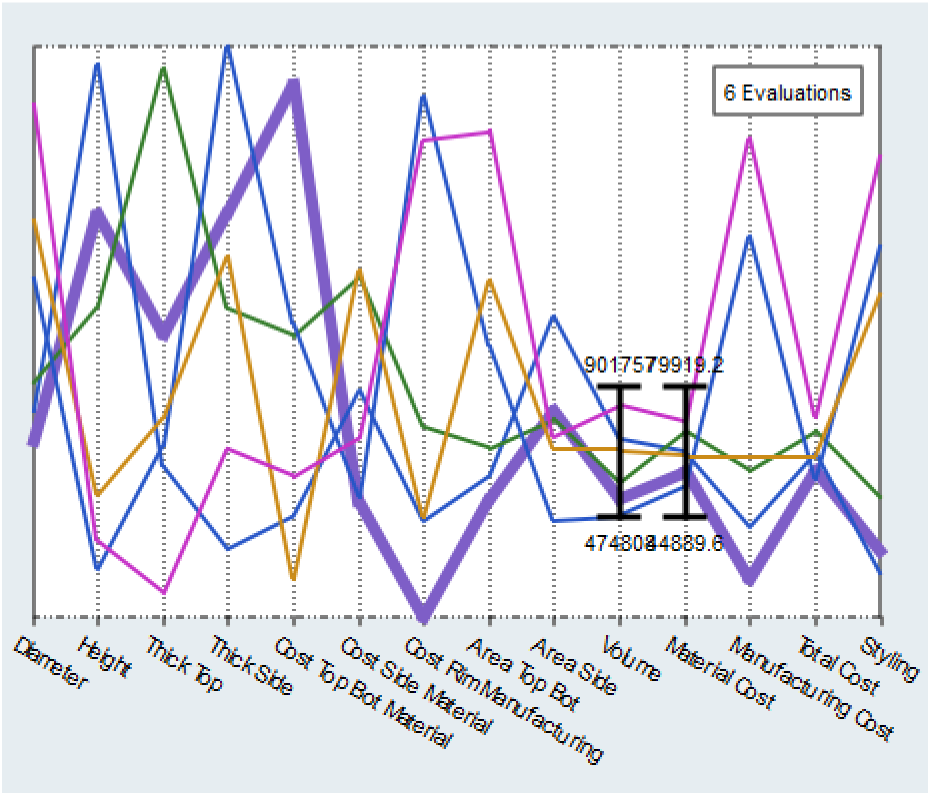


Figure 206:

Option	Description
	 *Figure 207:*
Filter run data	Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog. • From the Filter column, select the input variables and output responses to plot. • From the Filter Min and Filter Max columns, enter values to filter. The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

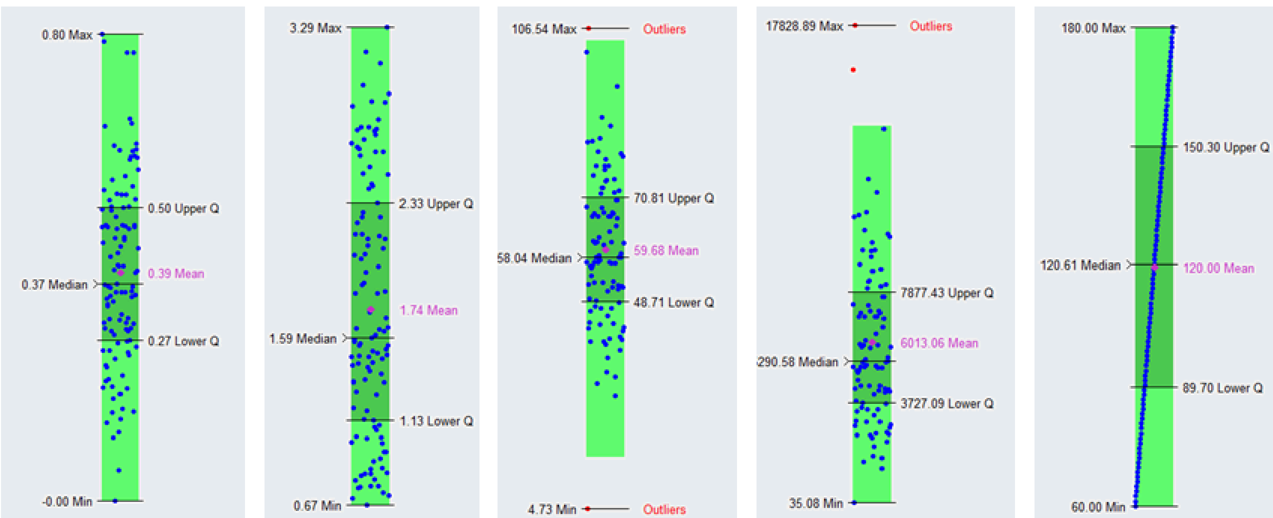


Figure 208:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 209 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

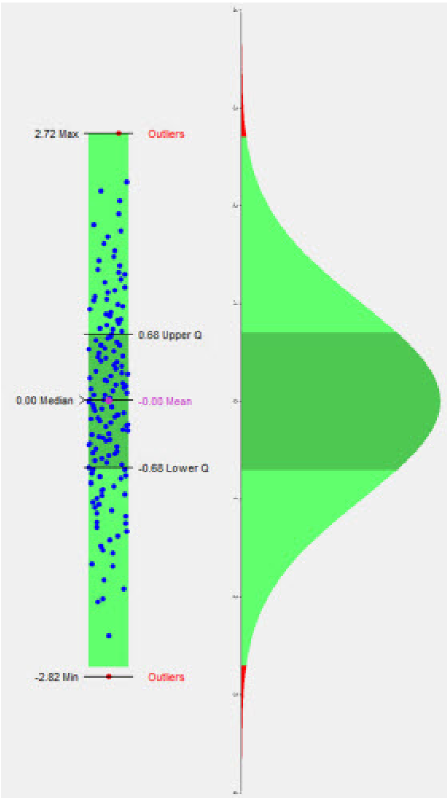


Figure 209:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

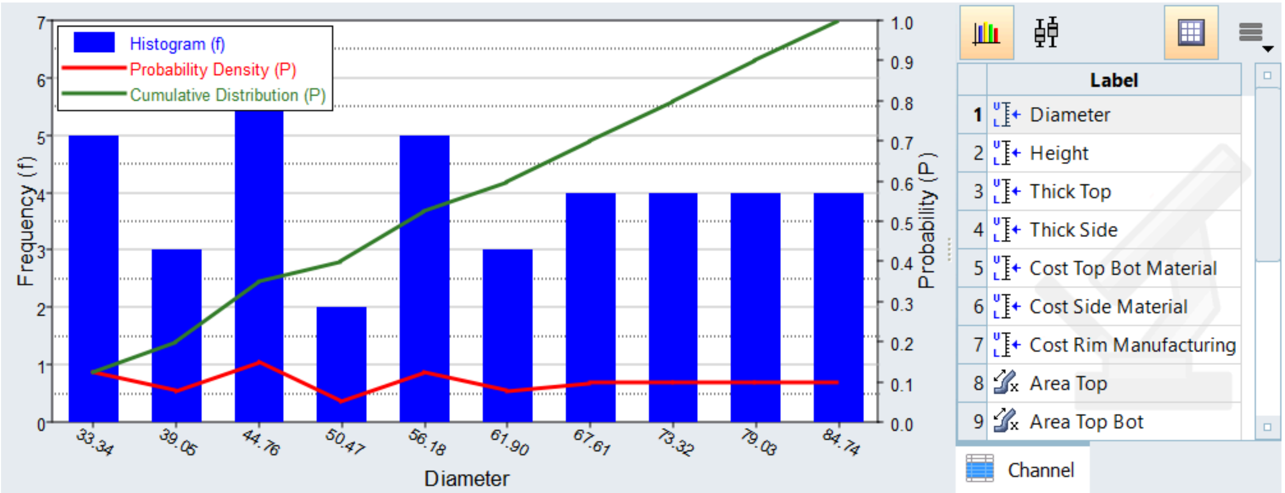


Figure 210:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processingEvaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. From the Post Processing step, click the **Scatter** tab.
3. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

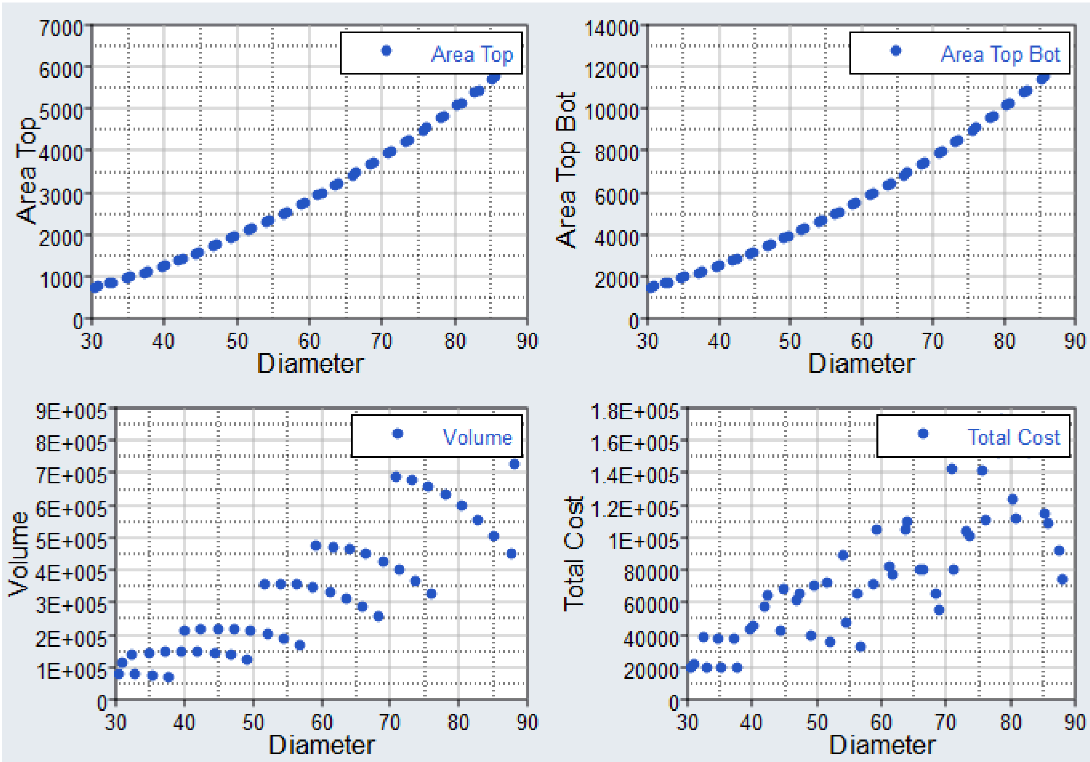


Figure 211:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

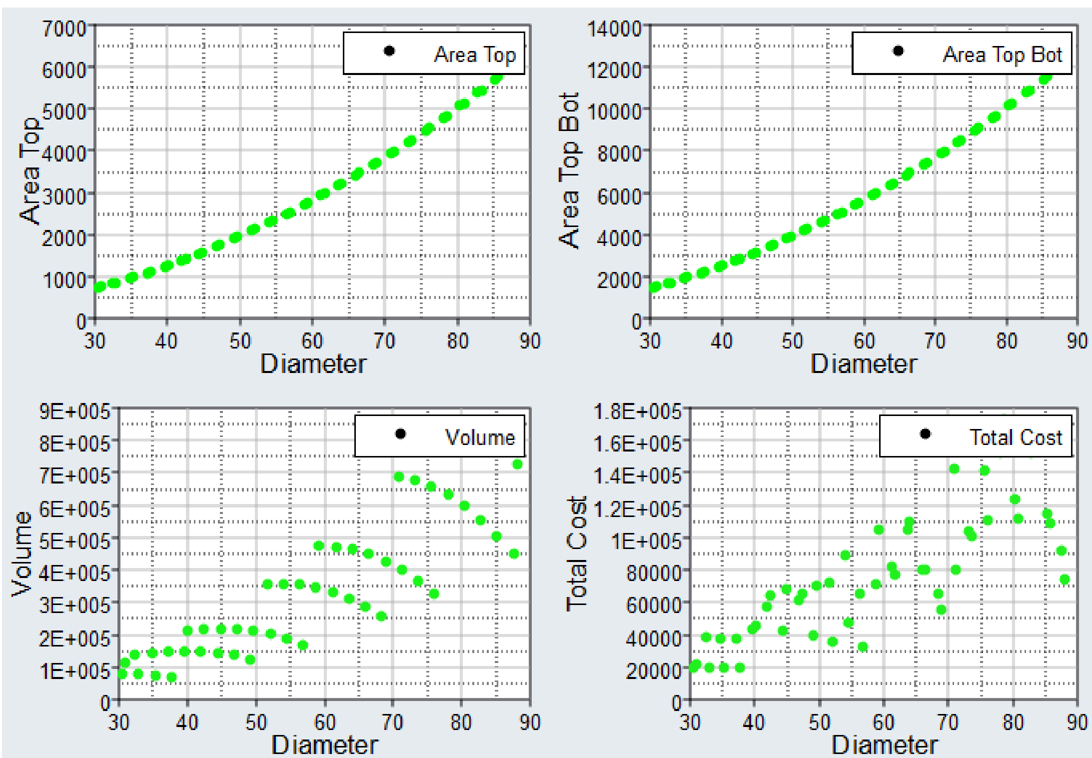


Figure 212:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

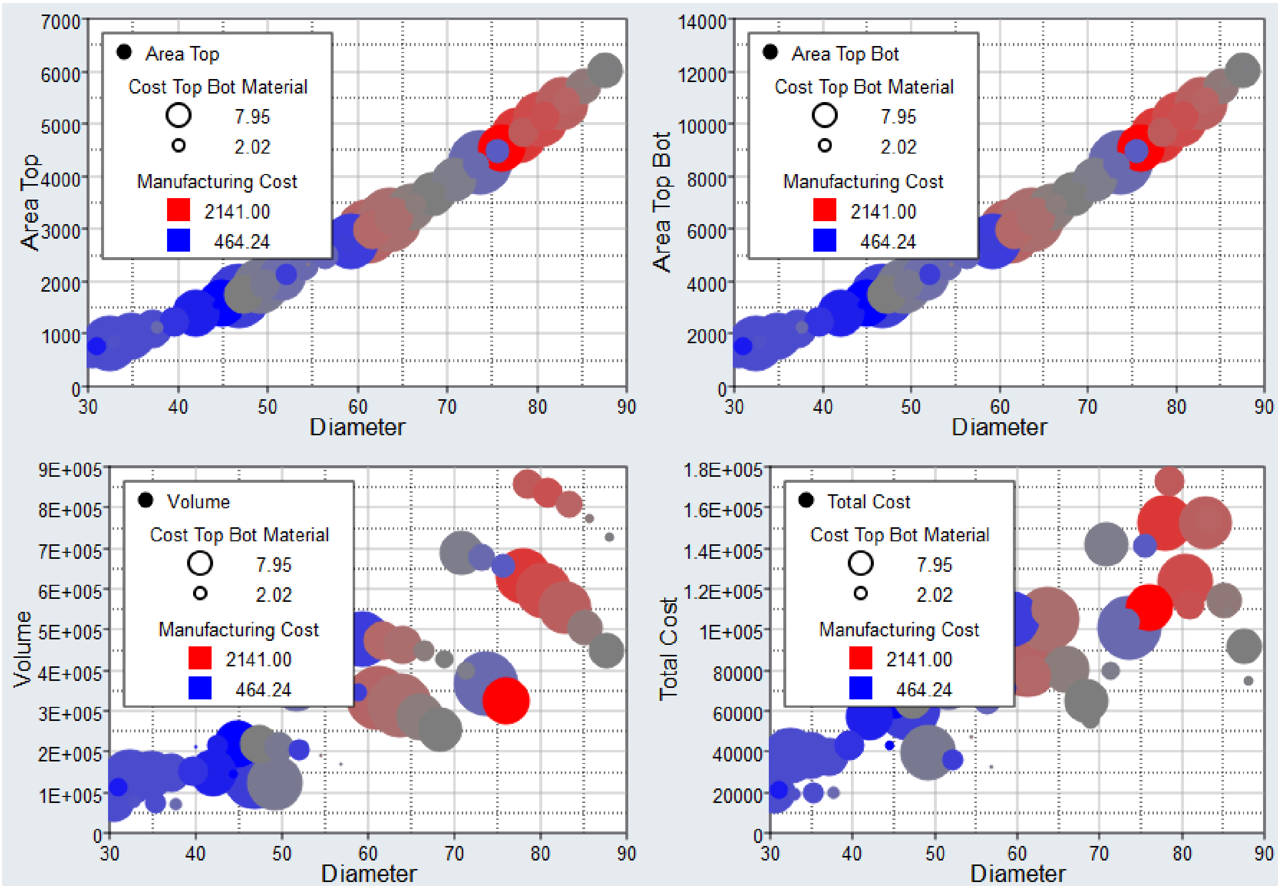


Figure 213:

4. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation ScatterScatter Tab Settings

Settings to configure the plots displayed in the Evaluation ScatterScatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds
Display the X bounds in the plot.

Y-Bounds
Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 214: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 215: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥
Show only the column/rows with cells over the specified threshold.

Show Variables and Responses
Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients	
X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

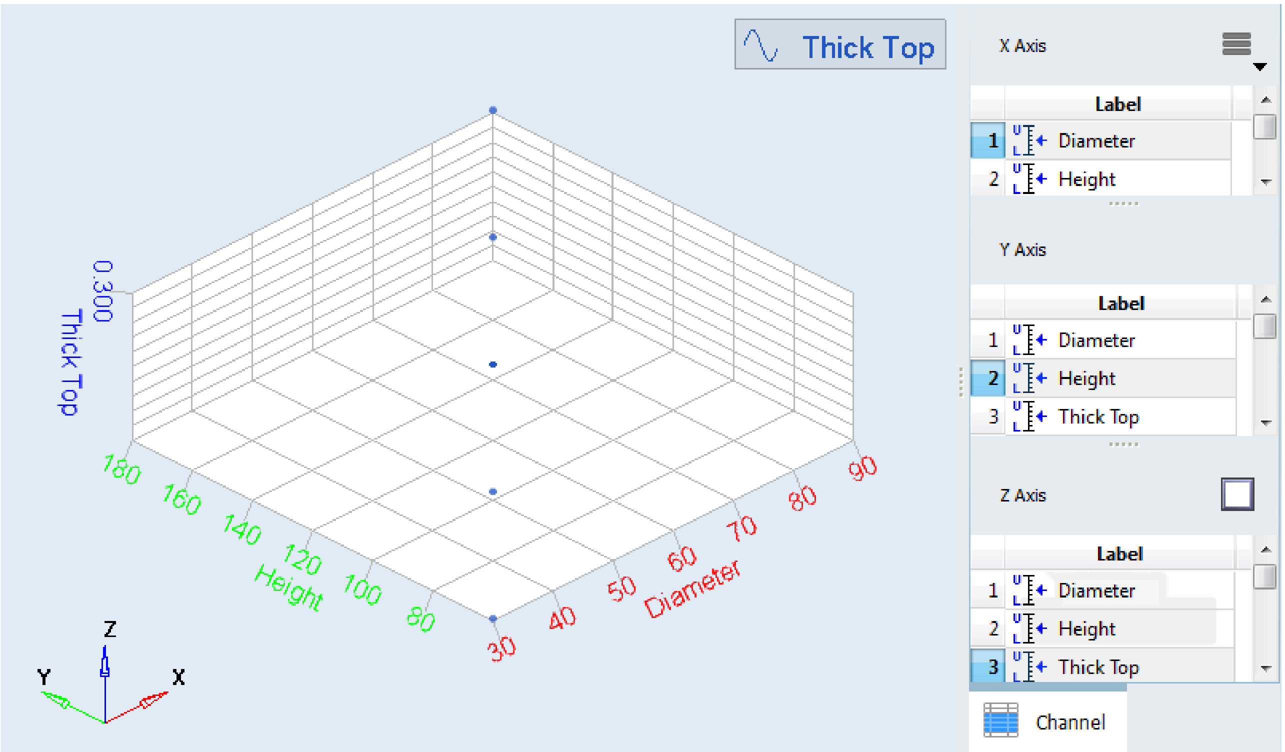


Figure 216:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

Tip: For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the biplot.

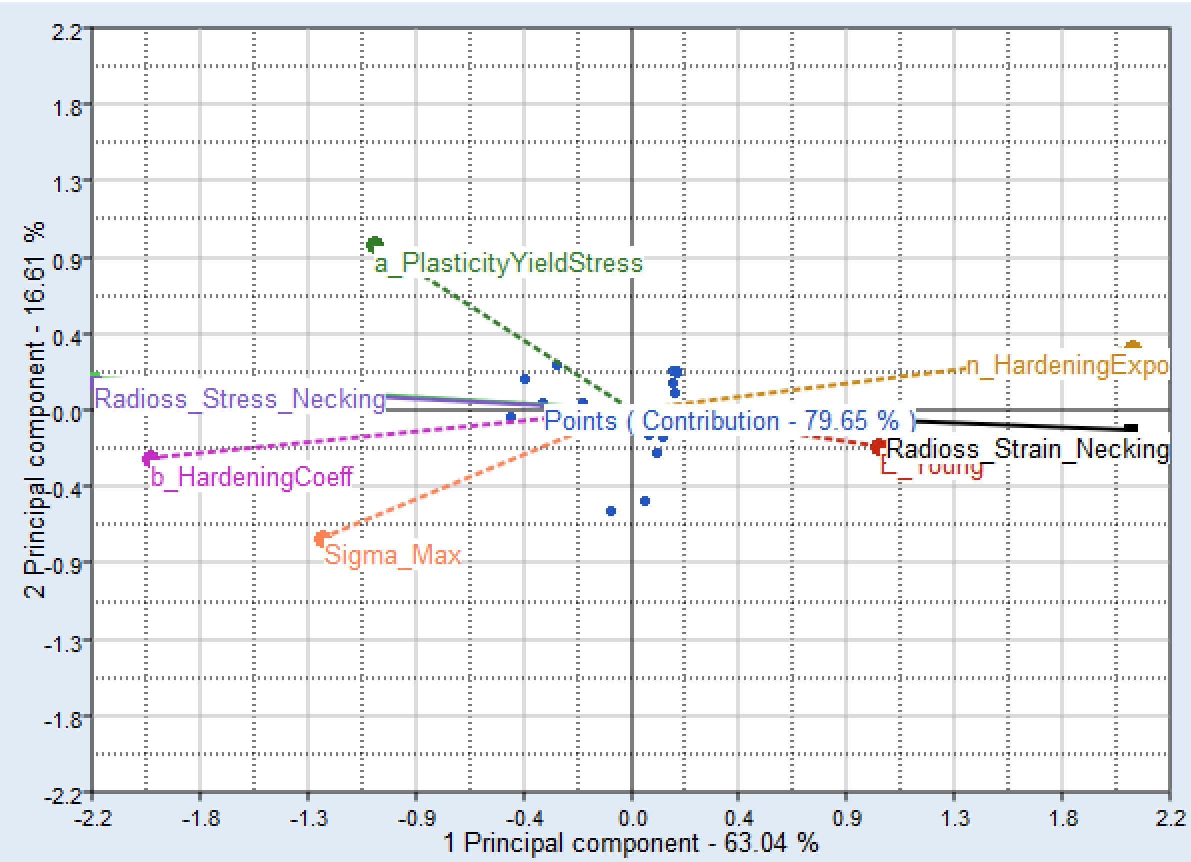


Figure 217:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
    a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

    > **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

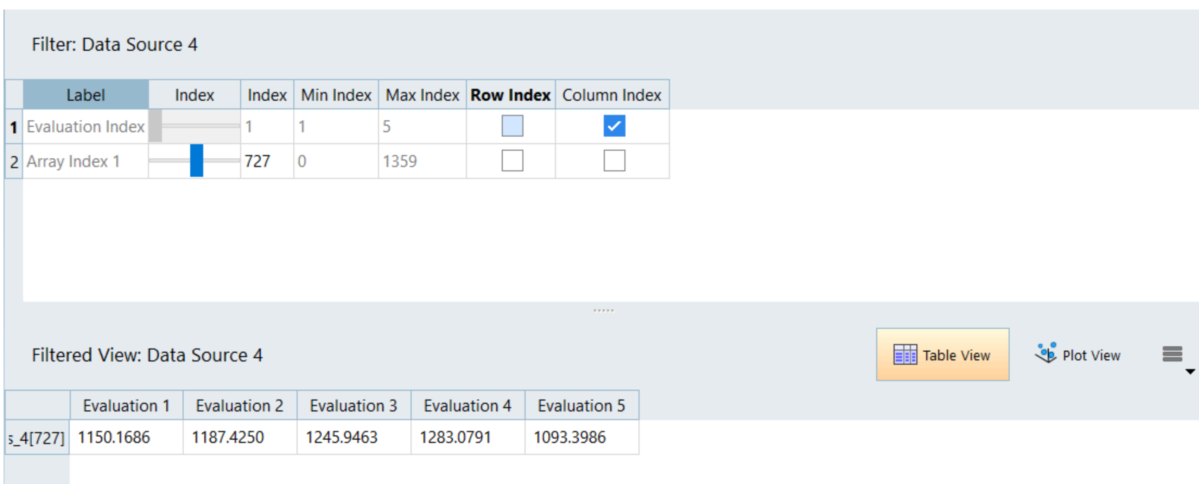


Figure 218:

5. Analyze the table.

Gradient Post Processing

Visualize gradients using vectors.

Analyze Vector

Analyze the vector in a gradient plot from the Gradient tab. Representing gradients as a vector field is an effective way to see gradients in space.

1. From the Post-Processing step, click the **Gradient** tab.

2. Use the Inputs and Output tabs of the Channel selector to select three dimensions of data to plot.
3. Analyze the direction and intensity of the vector created using the selected data sets.

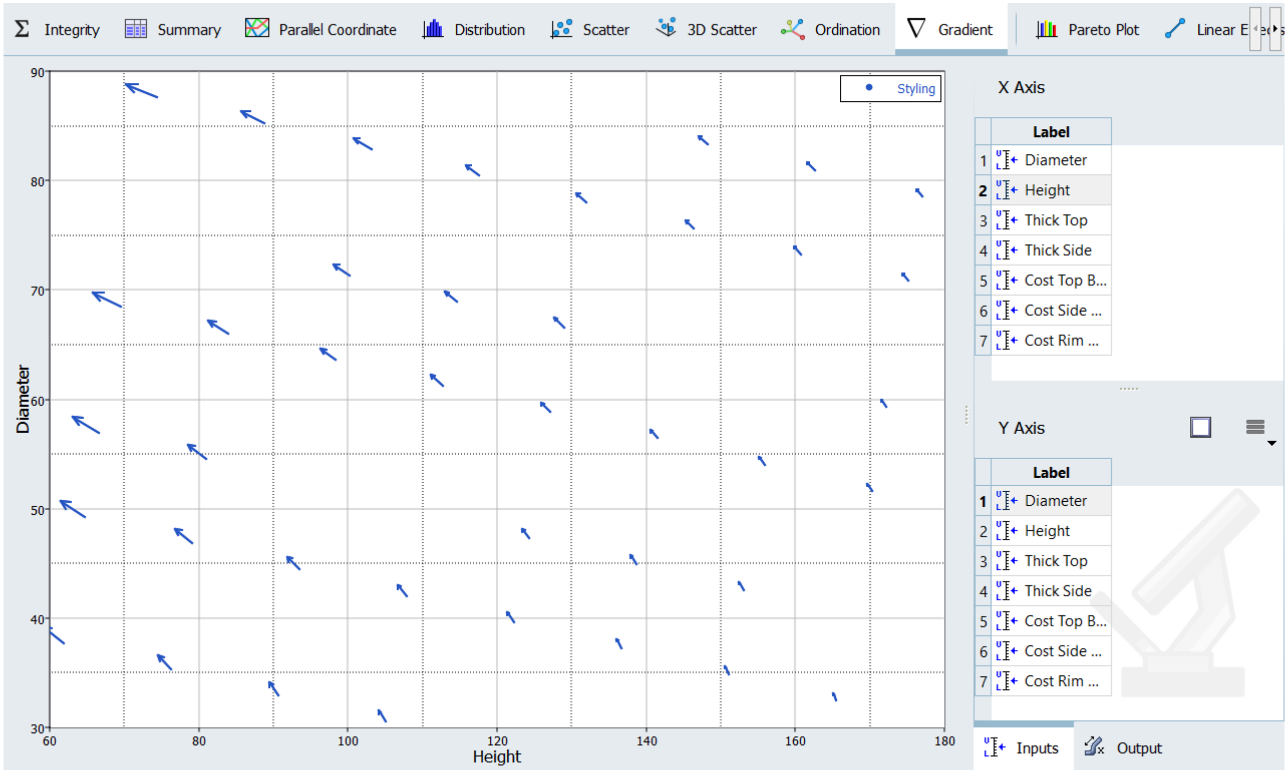


Figure 219:

Gradient Tab Settings

Settings to configure the plots displayed in the Gradient tab.

In the Gradient tab, use the Channel selector to select data to display in the gradient plot.

Channel Settings

Inputs	X-Bounds: display the X bounds in the plot.
	Y-Bounds: display the Y bounds in the plot.
Output	Gradient: display the vector in the plot.

Diagnostics Post Processing

Analyze the response surface quality.

Analyze the Predictive Model Quality

Analyze the Fit quality.

1. From the Post Processing step, click the **Diagnostics** tab.
2. In the work area, select the output response to analyze.
3. Click the tabs, below the output responses, to change the diagnostics used to analyze the selected output response.
   - **Detailed Diagnostics** displays diagnostic information for the Input matrix, Cross-Validation matrix, and Testing Matrix.
   - **Regression Terms** displays the confidence intervals which consist of an upper and lower bound on the coefficients of the regression equation.

     Bounds represent the confidence that the true value of the coefficient lies within the bounds, based on the given sample.

     Change the confidence value from the % Confidence settings. A higher confidence value will result in wider bounds; a 95% confidence interval is typically used.

     > **Note:** Only available for Least Squares Regression.

   - **Regression Equation** displays the complete formula for the predictive model as a function of the input variables.

     > **Note:** Only available for Least Squares Regression.

   - **ANOVA** estimates the error variance and determines the relative importance of various factors.

     Often used to identify which variables are explaining the variance in the data. This is done by examining the resulting increase in the unexplained error when variables are removed.

     > **Note:** Only available for Least Squares Regression.

   - **Confusion Matrix** summarizes the performance of a classifier. Correctly identified data is listed on the diagonal, and misclassifications are presented on the off-diagonals.

     > **Tip:** Click ≡ to toggle the confusion display from absolute count to percentages. Also, click ≡ to control the display of the confusion matrix between the input, cross-validation, and testing data set.

Configure the Diagnostics tab display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Diagnostic Tab Settings.

Diagnostic Tab Settings

Settings to configure the diagnostics displayed in the Diagnostic post processing tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the pane that displays the different types of diagnostics).

% Confidence	Change the confidence value. **Note:** Only available for Regression Terms diagnostics.

Diagnostic Definitions

Definitions used to describe diagnostic concepts.

For a given set of n input values, denoted as y_i, the Fit predictions at the same points are denoted as $\bar{y}_i$. The mean of the input values is expressed $\bar{y}$. For a Least Squares Regression, p is the number of unknown coefficients in the regression.

The following values are defined as follows:

Total Sum of Squares

$$SS_{tot} = \sum_{i=1}^{n} \left[y_i - \bar{y} \right]^2$$

Explained Sum of Squares

$$SS_{\mathrm{exp}} = \sum_{i=1}^{n} \left[\bar{y}_i - \bar{y} \right]^2$$

Residual Sum of Squares

$$SS_{\mathrm{eer}} = \sum_{i=1}^{n} \left[y_i - \bar{y}_i \right]^2$$

Average Absolute Error

$$\frac{1}{n} \sum_{i=1}^{n} \left[abs(y_i - \bar{y}_i) \right]$$

Standard Deviation

$$\sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} \left[y_i - \bar{y}_i \right]^2 \right)}$$

Detailed Diagnostic

Data displayed in the Detailed Diagnostic tab of the Diagnostics post process tool.

Input Matrix

The Input Matrix column shows the diagnostic information using only the input matrix. For methods which go through the data points, such as HyperKriging or Radial Basis Functions, input matrix diagnostics are not useful.

Cross-Validation Matrix

The Cross-Validation Matrix column shows the diagnostic information using a k-fold scheme, which means input data is broken into k groups. For each group, the group's data is used as a validation set for a new approximate model using only the other k-1 group's data. This allows for diagnostic information without the need of a testing matrix.

Testing Matrix

The Testing Matrix column compares the approximate model, which was built using the input matrix, against a separate set of user supplied points. Using a Testing matrix is the best method to get accurate diagnostic information.

Criterion

R-Sqaure

Commonly called the coefficient of determination, is a measure of how well the Fit can reproduce known data points. Graphically, this can be visualized by scatter plotting the known values versus the predicted values. If the model perfectly predicts the known values, R-Square will have its maximum possible value of 1.0, and the scatter points will lie on a perfect diagonal line, as shown in the Figure 220.

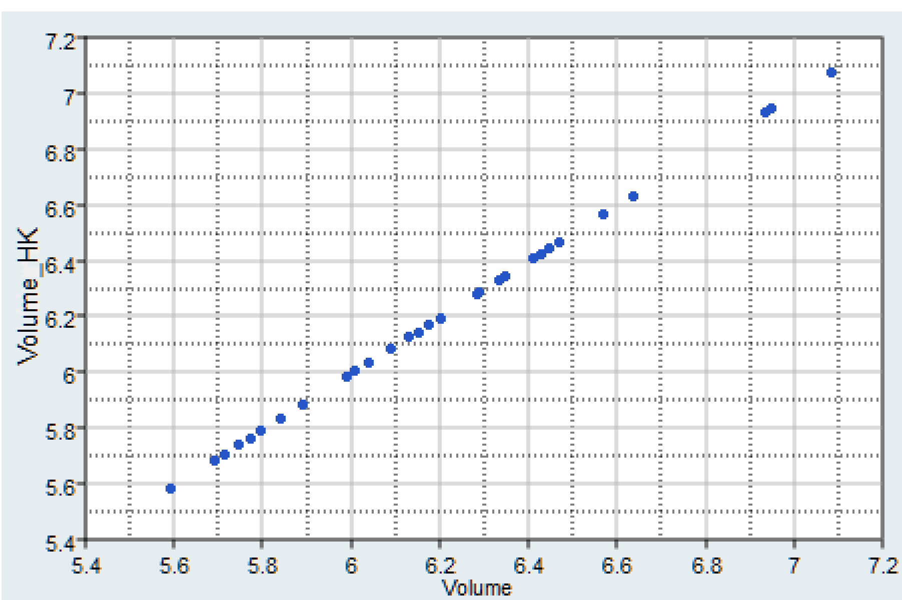


Figure 220:

More typically, the Fit introduces modeling error, and the scatter points will deviate from the straight diagonal line, as shown in the Figure 221.

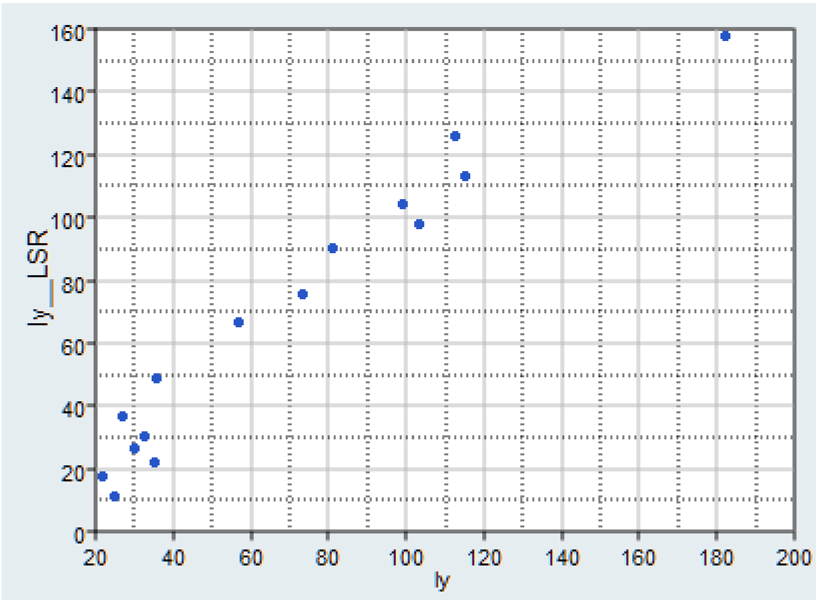


Figure 221:

The value of R-Square decreases as errors increase and the scatter plot deviates more from a straight line. The main interpretation of R-Square is that it represents the proportion of variance within the data which is explained by the Fit. For example, if R-Square = 0.84, then 84% of the variance in the data is predictable by the Fit. The higher the value of R-Square, the better the quality of the Fit. In practice, a value above 0.92 is often very good and a value lower than 0.7 necessitates investigation using other metrics. If R-Square is 1.0, you should be skeptical of this result unless the data was expected to be perfectly predicted by the Fit. There are some cases in which R-Square can be negative. A negative R-Square value indicates that using the raw mean would be a better predictor than the Fit itself; the Fit is very poor quality.
In the work area, these numbers are presented with a spark line to indicate the relative value of the number (values typically vary between 0 and 1). Values are color coded based on the following:

Red	When R^2 is less than 0.65 ($R^2 < 0.65$) it is displayed red, which indicates the value is not good.
Green	When R^2 is between 0.8 and 0.995 ($0.8 < R^2 < 0.995$) it is displayed green, which indicates the value is good.
Black	Indicates that you should apply judgment when determining whether the value is or is not good.

R-Sqaure is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left[y_i - \bar{y}_i\right]^2}{\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2}$$

R-Sqaure Adjusted

Due to its formulation, adding a variable to the model will always increase R-Square. R-Square Adjusted is a modification of R-Square that adjusts for the explanatory terms in the model. Unlike R-Square, R-Square Adjusted increases only if the new term improves the model more than would be expected by chance. The adjusted R-Square can be negative, and will always be less than or equal to R-Square. If R-Square and R-Square Adjusted differ dramatically, it indicates that non-significant terms may have been included in the model.

R-Sqaure Adjusted is defined as:

$$R^2\,adjusted = 1 - \frac{n-1}{n-p-1}\left(1 - R^2\right)$$

In the work area, these numbers are presented with a spark line to indicate the relative value of the number (values typically vary between 0 and 1). Values are color coded based on the following:

Red	When R^2 adjusted is less than 0.65 (R^2 adjusted < 0.65) it is displayed red, which indicates the value is not good.
Green	When R^2 adjusted is between 0.8 and 0.995 (0.8 < R^2 adjusted < 0.995) it is displayed green, which indicates the value is good.
Black	Indicates that you should apply judgment when determining whether the value is or is not good.

Multiple R

The multiple correlation coefficient between actual and predicted values, and in most cases it is the square root of R-Square. It is an indication of the relationship between two variables.

> **Note:** Only available for Least Squares Regression.

Relative Average Absolute Error

The ratio of the average absolute error to the standard deviation. A low ratio is more desirable as it indicates that the variance in the Fit's predicted value are dominated by the actual variance in the data and not by modeling error.

Relative Average Absolute Error is defined as:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left[abs(y_i - \bar{y}_i)\right]}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \bar{y}\right]^2\right)}}$$

Maximum Absolute Error

The maximum difference, in absolute value, between the observed and predicted values. For the input and validation matrices, this value can also be observed in the Residuals tab.

Maximum Absolute Error is defined as:

$$\max\left(abs\left(y_i - \bar{y}_i\right)\right)$$

Root Mean Square Error

A measure of weighted average error. A higher quality Fit will have a lower value.

Root Mean Square Error is defined as:

$$\sqrt{\frac{\sum_{i=1}^{n}\left[y_i - \bar{y}_i\right]^2}{n}}$$

Number of Samples

The number of data points used in the diagnostic computations.

Regression Terms

Data displayed in the Regression Terms tab of the Diagnostics post process tool.

t-value is defined as:

$$t_j = \frac{\beta_j}{\sqrt{\sigma^2 c_{jj}}}$$

where β_j is the corresponding regression coefficient (the Values column) and SE is the standard error. The standard error is defined as:

$$SE = \sqrt{\sigma^2 c_{jj}}$$

and

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left[y_i - \bar{y}_i\right]^2}{n - p}$$

where c_{jj} is the diagonal coefficient of the information matrix used during the regression calculation.

p-values are computed using the standard error and t-value to perform a student’s t-test. The p-value indicates the statistical probability that the quantity in the Value column could have resulted from a random sample and that the real value of the coefficient is actually zero (the null hypothesis). A low value, typically less than 0.05, leads to a rejection of the null-hypothesis, meaning the term is statistically significant.

ANOVA

Data displayed in the ANOVA (Analysis of Variance) tab of the Diagnostics post process tool.

Degrees of Freedom

Number of terms in the regression associated with the variable. All degrees of freedom not associated with a variable are retained in the Error assessment. More degrees of freedom associated with the error increases the statistical certainty of the results: the p-values. Higher order terms have more degrees of freedom; for example a second order polynomial will have two degrees of freedom for a variable: one for both the linear and quadratic terms.

Sum of Squares

For each variable, the quantity shown is the increase in unexplained variance if the variable were to be removed from the regression. A variable which has a small value is less critical in explaining the data variance than a variable which has a larger value.

The row Error, represents the variance not explained by the model, which is SS_{err}.

The row Total, which is SS_{tot}, will generally not equal to the sum of the others rows.

Mean Squares

The ratio between unexplained error increase and degrees of freedom, computed as the Sum of Squares divided by the associated degrees of freedom.

Mean Squares Percent

Interpreted as the relative contribution of the variables to the Fit quality, computed as the ratio of the Mean Square to the summed total of the Mean Squares. A variable with a higher percentage is more critical to explaining the variance in the given data than a variable with a lower percentage.

F-value

The quotient of the mean squares from the variable to the mean squares from the error. This is a relative measure of the variable's explanatory variance to overall unexplained variance.

p-value

The result of an F-test on the corresponding F-value. The p-value indicates the statistical probability that the same pattern of relative variable importance could have resulted from a random sample and that the variable actually has no effect at all (the null hypothesis). A low value, typically less than 0.05, leads to a rejection of the null-hypothesis, meaning the variable is statistically significant.

Residuals Post Processing

Identify design errors.

Identify Design Errors

Identify design errors in the Residuals post processing tab.

1. From the Post Processing step, click the **Residuals** tab.
2. From the Channel selector, select one output and one or more inputs to investigate.

3. Optional: Switch the view from residuals table to residuals plot by clicking , located next to the Channel selector.

> **Note:** The plot shows the residuals versus the original output response. This plot is useful to visually assess the performance of a Fit. It is desirable to not have any visual pattern to the residuals; unbiased and homoscedatic residuals appear similar to randomness.

4. Click ≡ (located above the Matrix Residuals table) to select the type of residuals displayed in the table.

   By default, Input Matrix residuals are displayed.

The error (and percentage) between the original output response and the approximation is listed for each run of the Input, Cross-Validation, or Testing matrices.

> **CAUTION:**
>
> View the Percent Error column with caution when the values of the output response approach zero. In this situation, the Percent Error can be very high and potentially misleading.

	Cost Side Material	Cost Rim Manufacturing	Area Top	Area Top	Error	Percent Error
1	2.7614467	2.2016189	1385.4424	1385.4424	-9.09e-13	-6.56e-14
2	2.5228935	2.9032379	2290.2210	2290.2210	4.55e-13	1.99e-14
3	2.1054261	3.4036476	3421.1944	3421.1944	0.0000000	0.0000000
4	2.5527131	3.9518238	4778.3624	4778.3624	0.0000000	0.0000000
5	1.5192770	3.8582212	824.47958	824.47958	1.36e-12	1.65e-13
6	2.5948835	1.7132828	1548.3025	1548.3025	-1.59e-12	-1.03e-13
7	1.6704901	2.5683444	2498.3201	2498.3201	9.09e-13	3.64e-14
8	1.2530227	3.0687541	3674.5324	3674.5324	4.55e-13	1.24e-14

Figure 222: Cross-Validation Matrix Residuals

> **Tip:** Search for specific cases using the Find and Sort options, which can be accessed from the context menu that opens when you right-click in the work area.

Configure the Residuals tab display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Residuals Tab Settings.

Residuals Tab Settings

Settings to configure the results displayed in the Residuals post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Matrix Residuals table).

Input Matrix Residuals	Display Input matrix residuals.
Cross-Validation Matrix Residuals	Display Cross-Validation matrix residuals.
Testing Matrix Residuals	Display Testing matrix residuals.

Trade Off Post Processing

Perform "What If" scenarios.

Perform "What If" Scenarios

Perform "What If" scenarios with interactive response surface tools in the Trade-Off post process tab.

1. From the Post Processing step, click the **Trade-Off** tab.
2. From the Channel selector, select the output response(s) to analyze in the Output Table.
3. Analyze the effect on inputs vs. outputs.

Option	Description
Modify the values of input variables to see their effect on output response approximations	In the Inputs pane, change each input variable by moving the slider in the first Value column, or by entering a value into the second Value column. Set input variables to their initial, minimum, or maximum values by moving the slider in the upper right-hand corner of the Inputs frame.

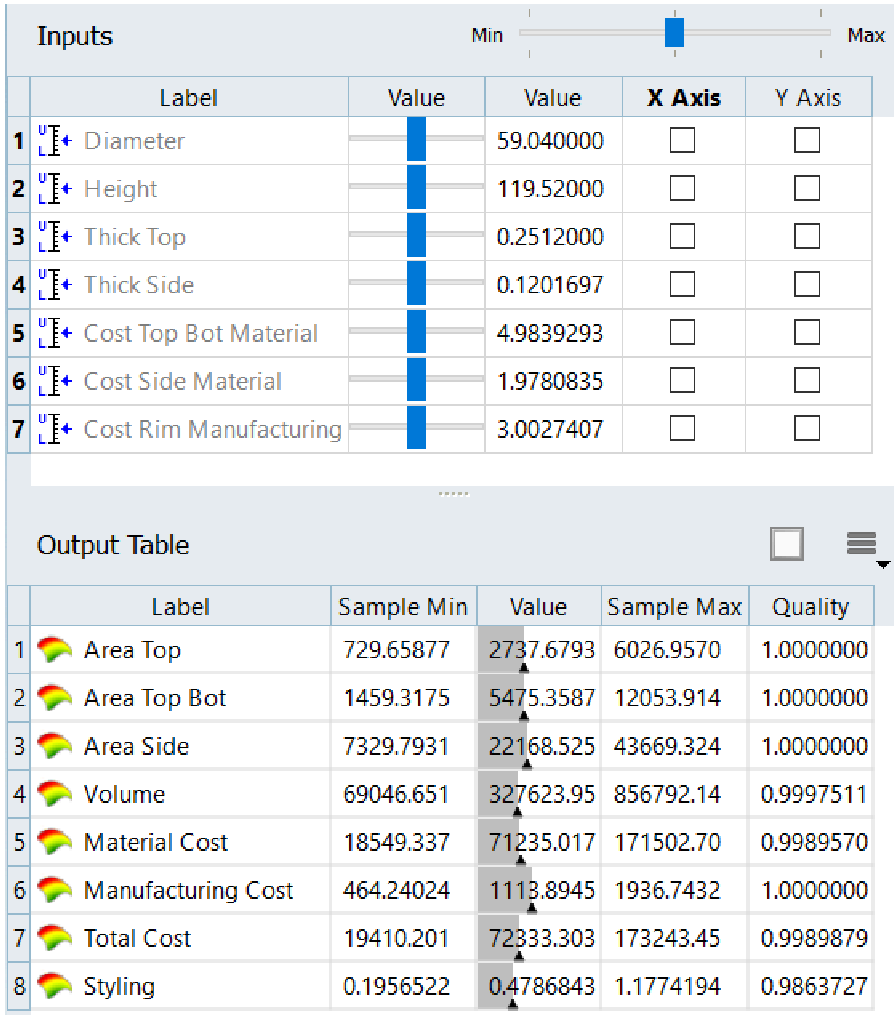
Option	**Description**
	

Figure 223:

Plot the effect of input variables on output response approximations	In the Inputs pane, select an input variable to plot by selecting its corresponding X Axis and/or Y Axis checkbox. • Create a 2D trade-off by enabling the X Axis checkbox.

Option	Description

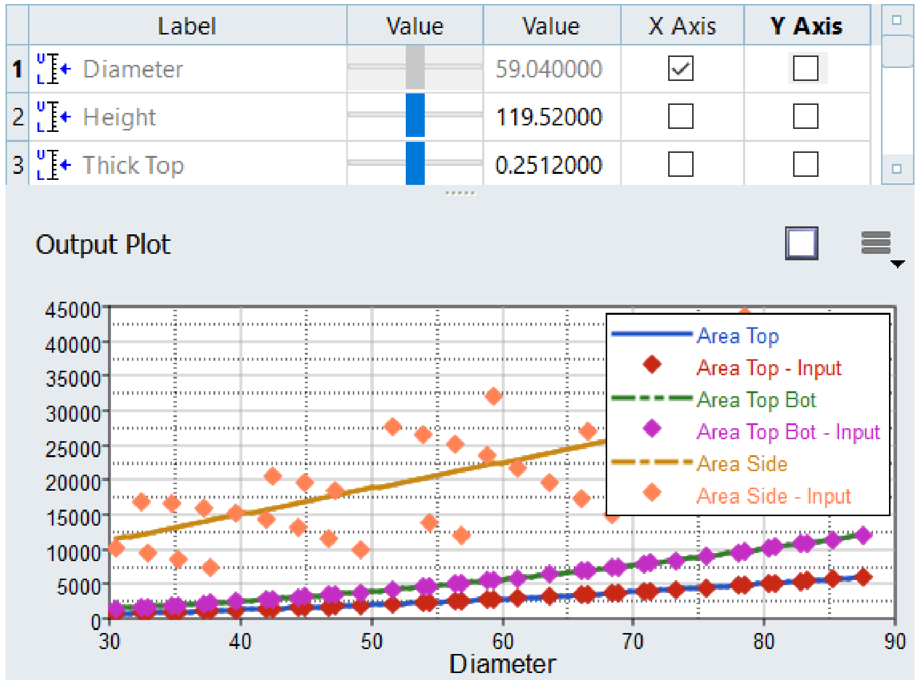


Figure 224:

- Create a 3D trade off by enabling X Axis and Y Axis checkboxes.

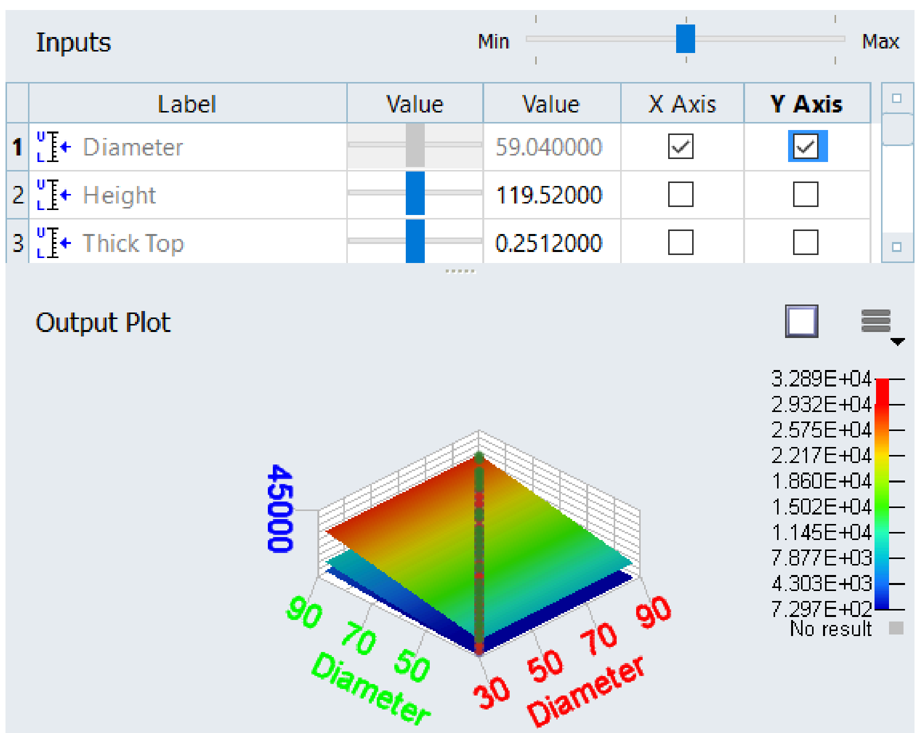


Figure 225:

Option	Description
	The output responses selected with the Channel selector are plotted. The values for the input variables which are not plotted can be modified by moving the sliders in the Value column to modify the other input variables, while studying the output response throughout the design space.

For the given values of the input variables, the output responses' predictions are calculated by the Fit, and displayed in the Output Plot pane. Table shading is used to indicate the output response's value between the minimum and maximum values contained in the input design matrix. When shading extends into either the Sample Min or Sample Max column, this indicates that the predicted value is beyond the bounds contained in the input matrix. If the shading extends significantly into these regions, it is suggested that you asses the validity of this value based on experience and knowledge of the modeled problem.

The Quality column is provided as a measure to assess both the accuracy and trust in the Fit at a specific point in the design space. Both global and local metrics are combined to create a metric that runs between 0 and an Upper Bound limited by the Fit's R^2 value. The quality will be highest at points inside the convex hull formed around the Fit's input points, where the predictive model has been trained to explain variance in the data. The quality metric decreases proportional to the distance outside the convex hull as the predictions at these points becomes less reliable, partially due to the values increasing based on an extrapolation of the data.

Tip: In a 2D trade-off the metrics shown in the Quality column can be plotted alongside the output response curve by selecting **Fit Quality** from the menu that displays when you click (located in the Output Plot pane).

Configure the Trade Off tab's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Trade-Off Tab Settings.

Trade-Off Tab Settings

Settings to configure the results displayed in the Trade-Off post processing tab.

Access settings from the menu that displays when you click ≡ (located in the Output Plot pane).

Fit	Display the predicated curve (2D plot) or surface (3D plot) of the Fit.
Fit Quality	Plot the estimated quality (value shown in the Quality column) alongside the output response curve.
Input Matrix	Display the scatter points of the Input matrix.
Testing Matrix	Display the scatter points of the Testing matrix.

# samples	Change the number of discretized points used when drawing the trade-off (2D plot) or surface (3D plot). **Note:** Increasing this number will result in a smoother representation, which could be at the cost of interface responsiveness.
Discrete Surface Contour	Display a discrete color profile of the surface. Disable this checkbox to display a blended color profile of the surface (3D plot).
Mesh lines	Display a visual projection of the samples' mesh grid lines onto the surface (3D plot).

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	Description
HyperStudy Data	Generates a data report (`*.data`).
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel. In the Excel spreadsheet, click the **Trade-Off** sheet to perform trade-off studies in Excel, independent of HyperStudy. In a Fit Excel report, the Trade-Off 1D tab is a reflection of the corresponding Trade-Off 1D tab within HyperStudy. From the Trade-Off 1D tab in Excel, you can adjust the input variable values on the right-hand side to change the predicted output responses values displayed on the left-hand side.

Report Type	Description
	Restriction: To use this feature, you must add the HstAddinFit add-in to Excel. To install the Excel Plug-in for HyperStudy Fit engines, go to the `<ALTAIR_HOME>\hst\plugins\externals\hstfitaddin\` directory and double-click `hstfitaddin_install.vbs`. The Excel Plug-in requires access to a valid HyperWorks license for all features to work properly.

3. Click **Create Report**.

4.2.5 Setup Stochastic Studies

A Stochastic approach is a method of probabilistic analysis where the input variables are defined by a probability distribution, and consequently the corresponding output responses are not a single deterministic value, but a distribution.

Add a Stochastic Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
    a) In the Label field, enter a name for the Stochastic.
    b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.

       By default, the Definition defined in the study Setup is selected.
    c) Under Select Type, select **Stochastic**.
    d) Click **OK**.

A new Stochastic is added to the Explorer.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.

3. Test Models.
4. Define Output Responses.
5. Review definitions in the following ways:

To: **Do this:**

Review status

Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

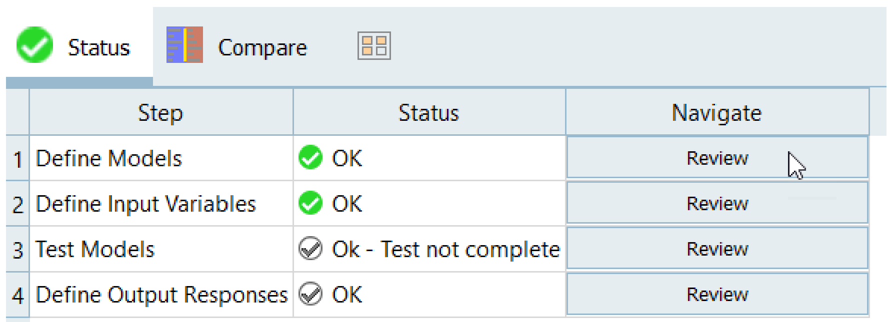


Figure 226:

Compare definitions

Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

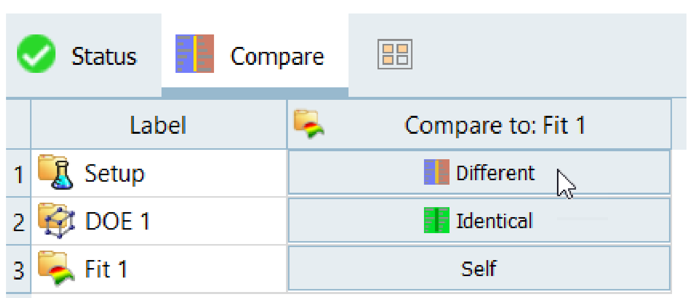
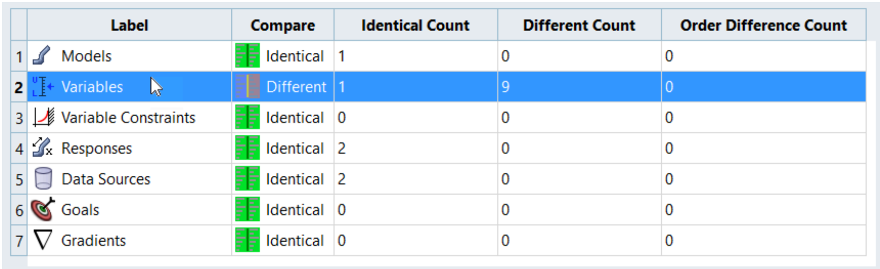


Figure 227:

To:	Do this:
	The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.

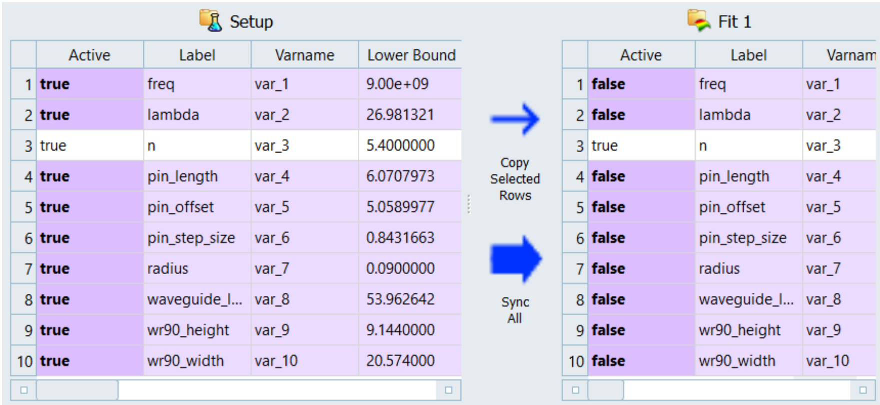
4. Click a channel to display a detailed comparison.



	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 228:

5. Sync data.
    - Click **Copy Selected Rows** to sync the single row.
    - Click **Sync All** to sync all rows.



Setup

	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

Fit 1

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 229:

Select a Numerical Method

Select a numerical method to use when evaluating the Stochastic approach.

1. In the Specifications step, Mode column, select a numerical method.
2. Optional: In the Settings tab, change settings as needed.
3. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog. For more information, see Edit the Run Matrix.

Stochastic Methods

Numerical methods available for an Stochastic approach.

Method	Efficiency	Basic Parameter	Comments
Hammersley	##	Number of runs	
Latin HyperCube	##	Number of runs	Maintain the value of the random seed to get repeatable designs.
Modified Extensible Lattice Sequence	##	Number of runs	Maintain the value of the random seed to get repeatable designs. **Note:** Modified Extensible Lattice Sequence can be extended upon itself to add points to a previously completed MELS DOE.
Simple Random	#	Number of runs	Maintain the value of the random seed to get repeatable designs.

Hammersley

Hammersley sampling belongs to the category of quasi-Monte Carlo methods. This technique uses a quasi-random number generator, based on the Hammersley points, to uniformly sample a unit hypercube.

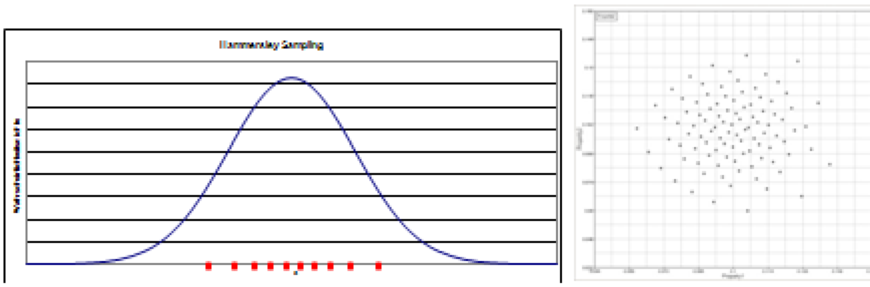


Figure 230: Hammersley Sampling

Usability Characteristics

- An efficient sampling method that provides reliable estimates of output statistics using fewer samples than random sampling.
- A correlation structure can be specified to reflect the correlation existing between random variables. Applying a correlation structure can be costly for a large number of input variables.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	100	> 0	Number of new designs to be evaluated.
Apply User Correlations	On	Off or On	Apply user specified correlations on the data.

Latin HyperCube

An approach which can yield precise estimates of output statistics with a lesser number of samples than simple random sampling.

The Latin HyperCube method uses a constrained or stratified sampling scheme.

Latin HyperCube sampling selects n different values from each of k variables $x1$, … xk in the following manner:

- The range of each random variable is divided into n non-overlapping intervals on the basis of equal probability.
- One value from each interval is selected at random with respect to the probability density in the interval.

- The n values thus obtained for $x1$ are paired in a random manner with the n values of $x2$. These n pairs are combined in a random manner with the n values of $x3$ to form n triplets and so on, until n k-tuplets are formed.

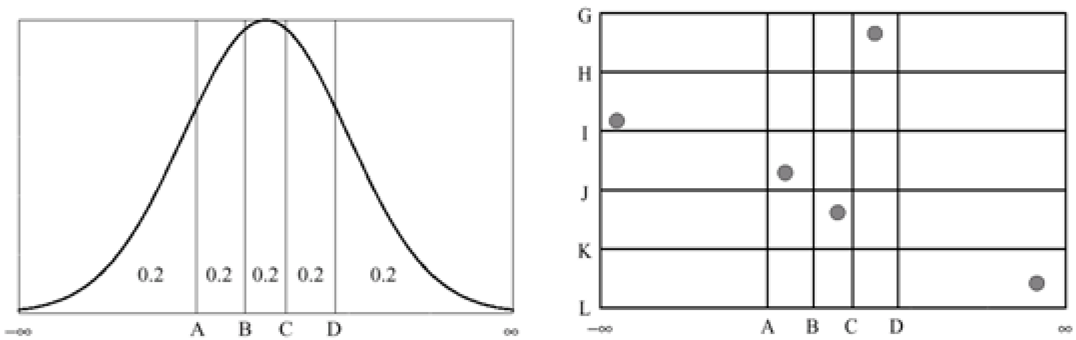


Figure 231: Latin HyperCube Sampling

Usability Characteristics

- A stratified sampling scheme like Latin HyperCube offers the advantage of selecting random variable values that are uniformly spread across the range of random variables while taking into account the probability density function of those random variables.
- A correlation structure can be specified to reflect the correlation existing between random variables. Applying a correlation structure can be costly for a large number of input variables.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	100	> 0	Number of new designs to be evaluated.
Random Seed	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Apply User Correlations	On	Off or On	Apply user specified correlations on the data.

Modified Extensible Lattice Sequence

A lattice sequence is a quasi-random sequence, or low discrepancy sequence, designed to equally spread out points in a space by minimizing clumps and empty spaces.

This property makes lattice sequences an excellent space filling Stochastic scheme.

Usability Characteristics

- An efficient sampling method that provides reliable estimates of output statistics using fewer samples than random sampling.
- A correlation structure can be specified to reflect the correlation existing between random variables. Applying a correlation structure can be costly for a large number of input variables.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	100	> 0	Number of new designs to be evaluated.
Sequence Offset	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Apply User Correlations	On	Off or On	Apply user specified correlations on the data.

Simple Random

The conventional approach of sampling is commonly called Simple Random or Monte Carlo. In Simple Random sampling, a pseudo-random number generator is used for generating random numbers from 0 to 1.

Design points are generated by using the Inverse Transform method. Clustering may occur in the design point distribution because the sequence of samples is random.

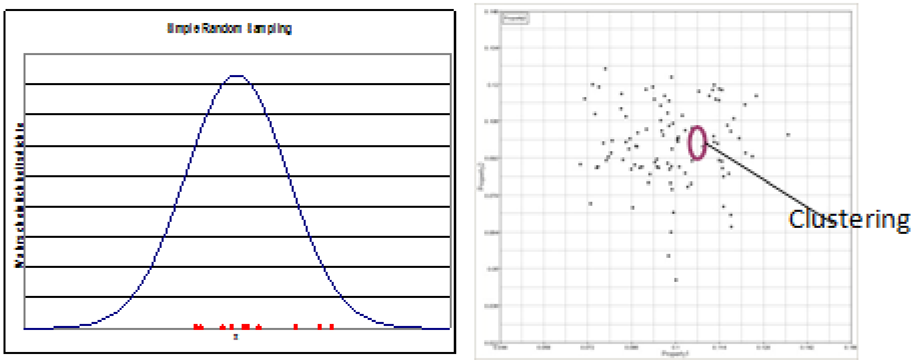


Figure 232:

Usability Characteristics

- The statistical measures (such as mean or standard deviation) of a random sample group requires large numbers of runs to converge the given probability distribution's statistical measures.
- A correlation structure can be specified to reflect the correlation existing between random variables. Applying a correlation structure can be costly for a large number of input variables.

Settings

In the Specifications step, Settings tab, change method settings.

Parameter	Default	Range	Description
Number of Runs	100	> 0	Number of new designs to be evaluated.
Random Seed	1	Integer 0 to 10000	Controlling repeatability of runs depending on the way the sequence of random numbers is generated. **0** Random (non-repeatable). **>0** Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.
Apply User Correlations	On	Off or On	Apply user specified correlations on the data.

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

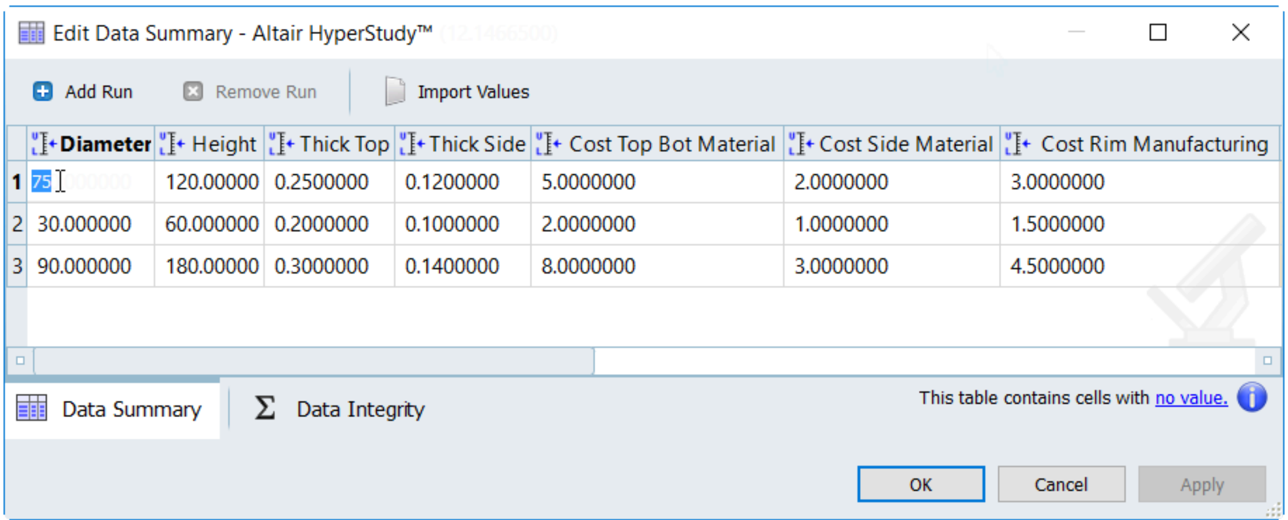


Figure 233:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.
   - Manually enter run data.
   - Copy and paste run data into the run matrix.

     Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

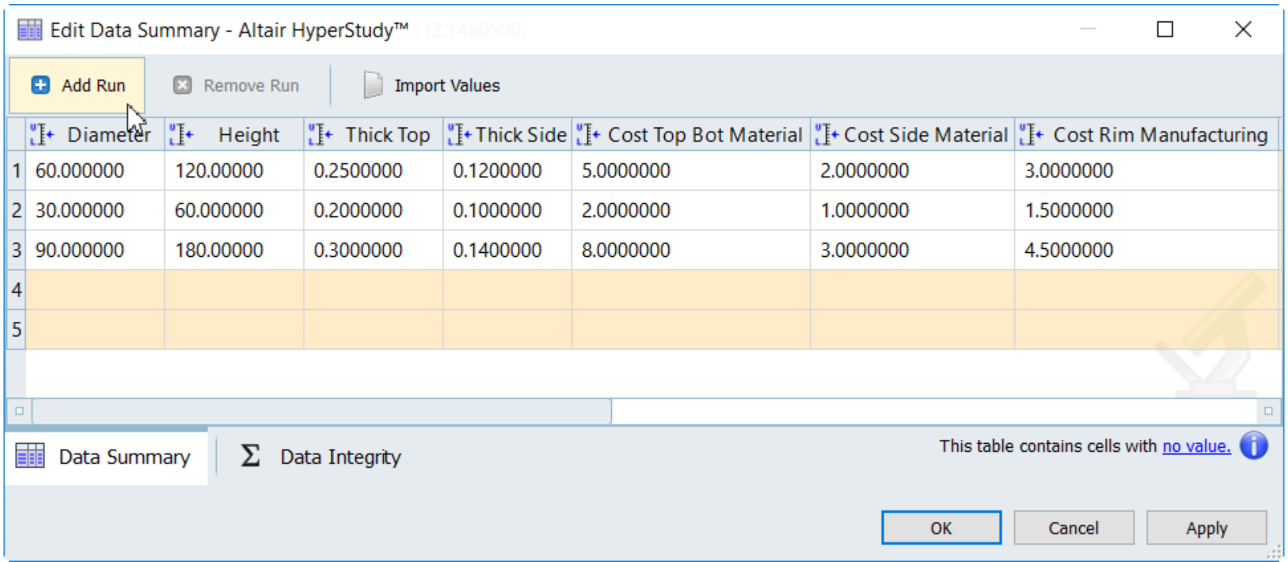


Figure 234:

Tip: Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.

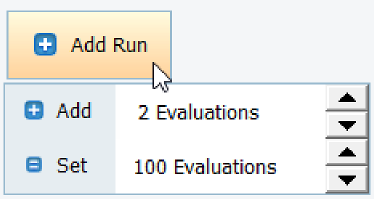


Figure 235:

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.
5. Select the source that contains run data.
   - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
   - For Approach evaluation data, select the approach that contains run data.

- For HyperStudy post processing file, select the source file.

6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.
2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.
3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

    By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

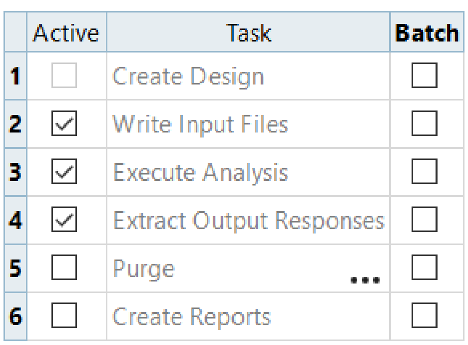


Figure 236:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to

Setting	Action
	stop the submission of new jobs. Click ≡ to select an execution mode.

Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

Each evaluation is independent so multi-execute can be used to run in parallel.

Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

- Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
- Horizontal execution mode sequences the write, execute, and extract task for each run independently.
- Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

Stochastic Output Files

Output files generated from the a Stochastic.

<sto_variable_name>.hstds

File Creation

This file is created when Apply is selected during the Specifications step.

File Location

`<study_directory>/approaches/<sto_variable_name>/<sto_variable_name>.hstds`

File Contents

Result	Format	Description
Run Matrix Data	hstds, binary	Hstds files stores the retained data sources; direct access data using the `.hstds` file is not suggested.

<sto_variable_name>.hstdf

File Creation

This file is created when **Apply** is selected during the Specifications step.

File Location

`<study_directory>/approaches/<sto_variable_name>/<sto_variable_name>.hstdf`

File Contents

Result	Format	Description
Run Matrix Data	hstdf, binary	Hstdf files store the run data; however, direct access to the data using the hstdf files are not suggested.

Evaluation Parameters

Modify the run environment settings for the Evaluation tasks.

1. From the Evaluation step, click the **Evaluation Parameters** tab.
2. In the Value column, modify settings accordingly.

> **Note:** Review the Effectuation column to determine the scope at which each setting takes effect.

Review Evaluation Results

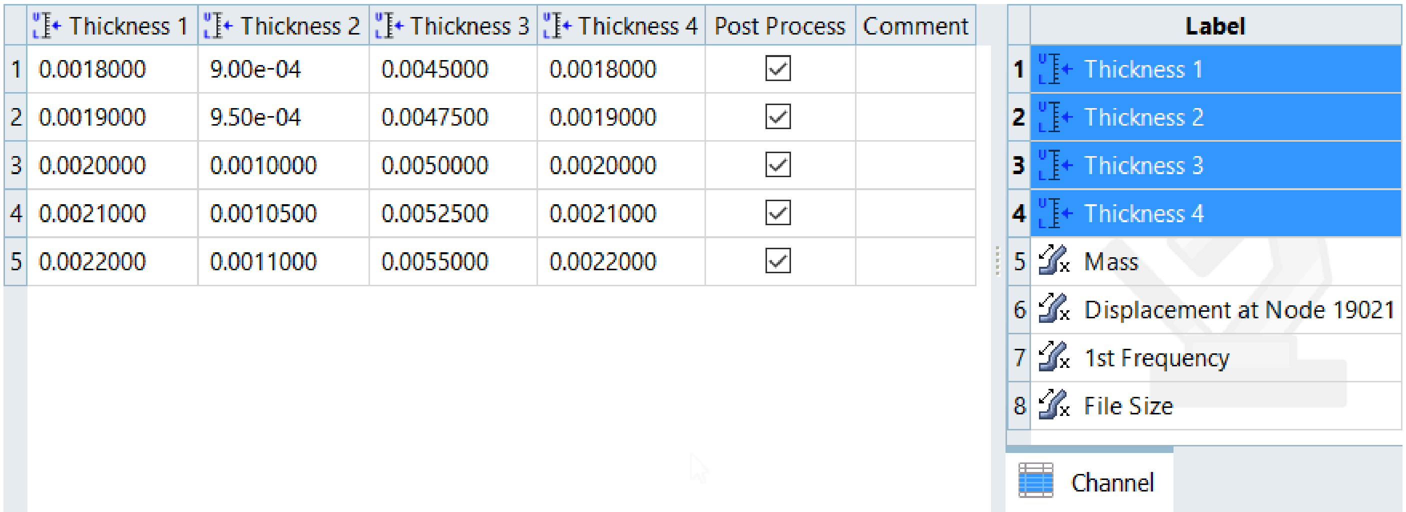
Review the input variable and output response values for each run, as well as review the run files.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Evaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.
4. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

Figure 237:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

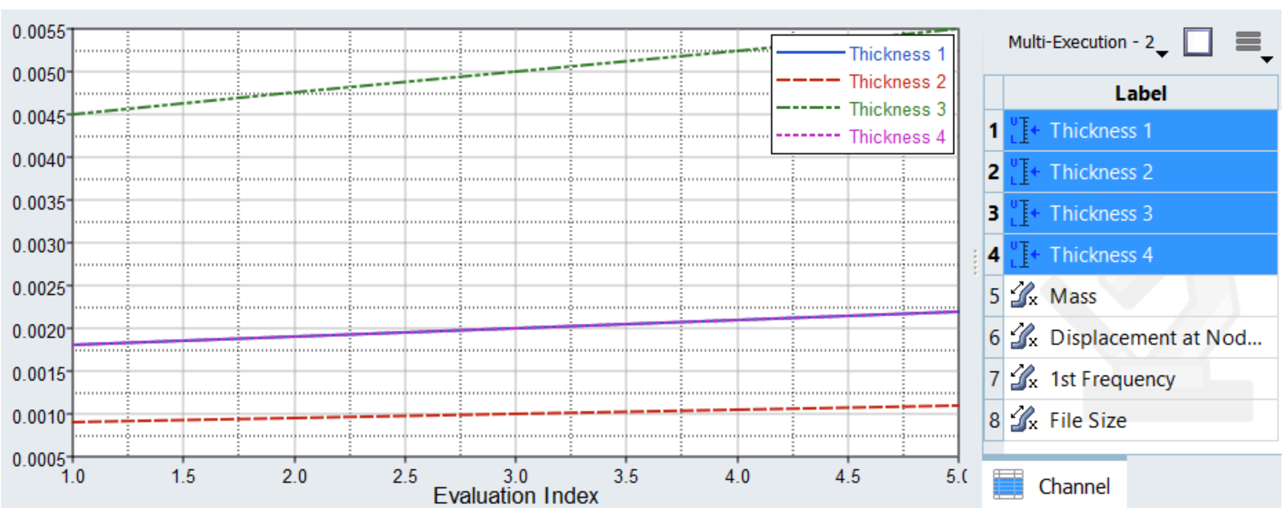


Figure 238:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Evaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. Select data to display in the scatter plot.
   - Use the Channel selector to select two dimensions of data to plot.

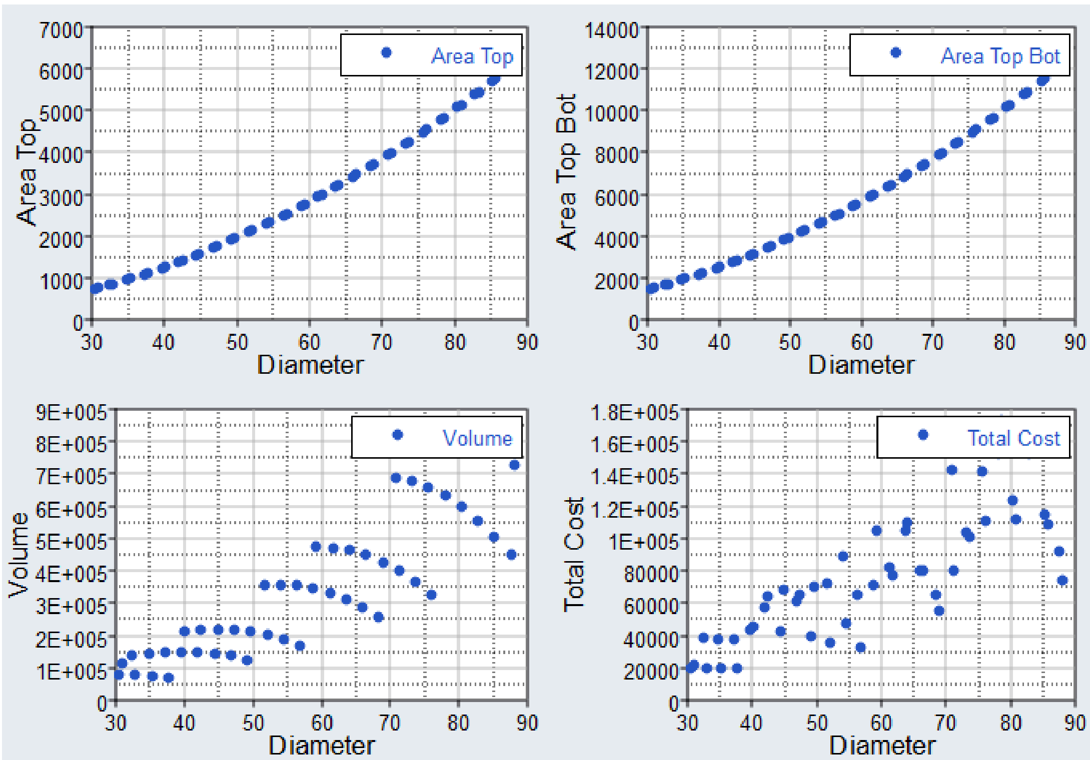


Figure 239:

   - Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

     The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

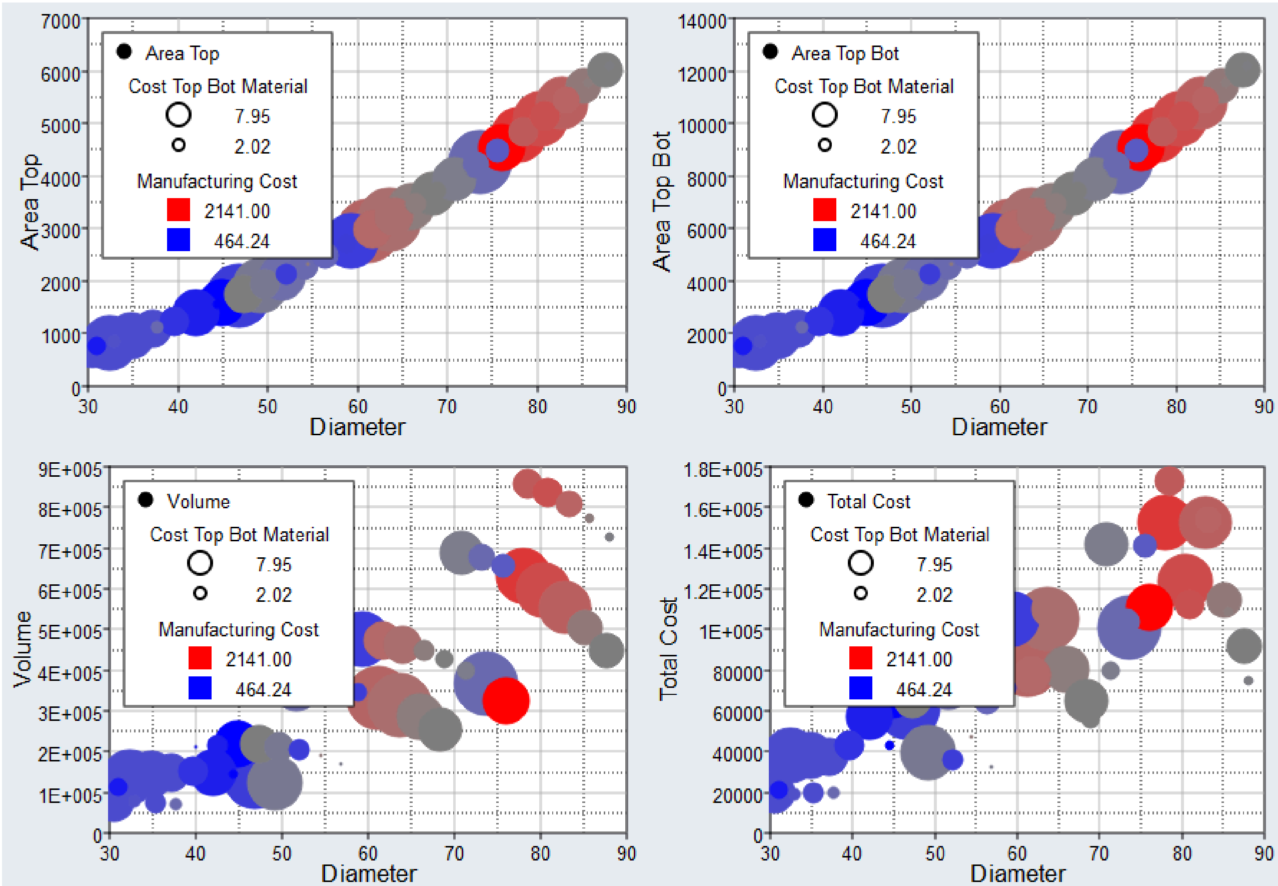


Figure 240:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation Scatter Tab Settings

Settings to configure the plots displayed in the Evaluation Scatter tab.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.

Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Review Evaluation Time

Inspect task wall-clock times.

Review the time spent in each task within the Evaluation Time tab. Identify bottlenecks in tabular or plot form.

1. From the Evaluate step, click the **Evaluation Time** tab.
2. Use the top channel selector to select the model(s) to review.
3. Use the bottom channel selector to identify the time categorises for review.

Option	Action
Write	Time spent in the write task.
Execute	Time spent in the execute task.
Extract	Time spent in the extract task.
Model Total	Total time of the write, execute, and extract tasks.
All Models Total	Summation of all Model Totals.

Option	Action
	Note: This category is independent of the selected models.

4. Switch the view between table and plot by clicking Table or Plot, located above the Channel selector.

Evaluation Time Settings

Settings to configure the plots and tables displayed in the Evaluation Time tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Cumulative Rows
: Each row entry is a summation of its own wall time and the wall times preceding it with a lower evaluation index.

Plot Time-Unit
: Controls the units of time when plotting the wall times.

Post Processing

View the computational results from the Stochastic. approach

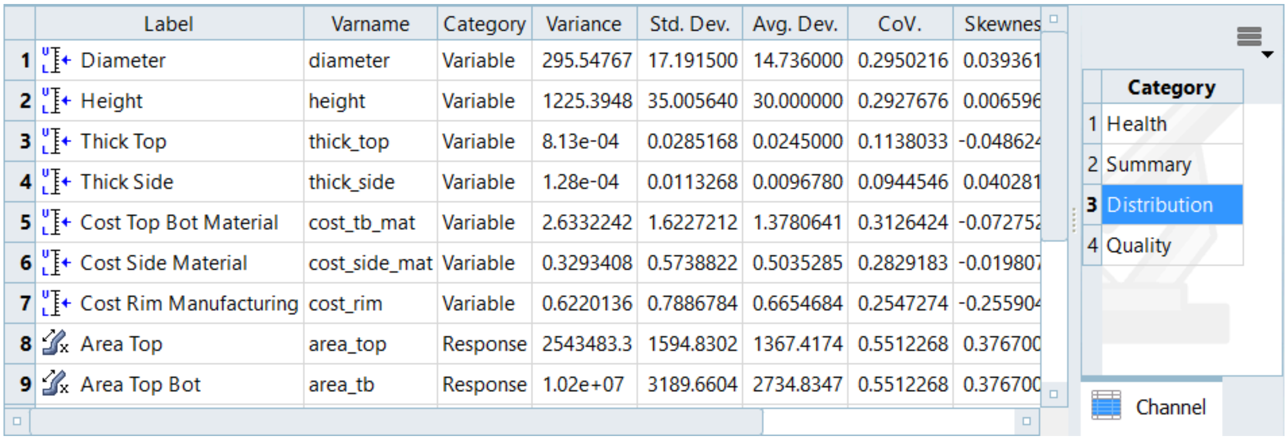
Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
   - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
   - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.
   - **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
   - **Quality** Values typically used in Quality Engineering.



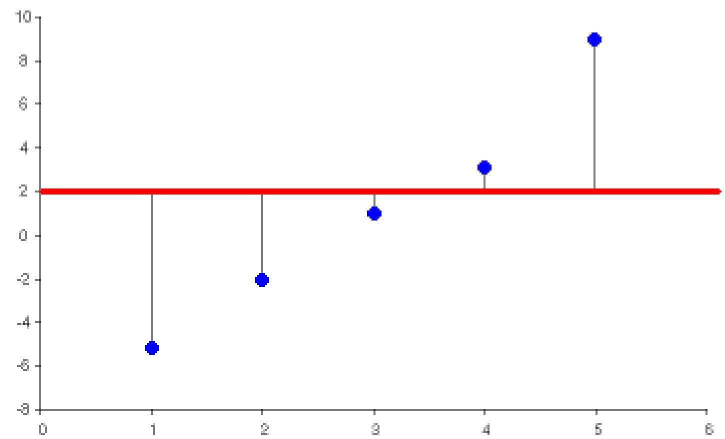
	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.072752
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.019807
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.255904
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

Figure 241:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	Description
Avg Dev (Average Deviation)	Average deviation is evaluated using: $$\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$$ In Figure 242, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 242:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $CoV = \frac{Standard\ Deviation}{Mean}$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $\sqrt{\frac{\sum x_i^2}{N}}$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $\frac{\sum_{i=1}^{N}(x_i-\overline{x})^2}{N-1}$

Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processing tab.

1. From the Post Processing step, click the **Summary** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.

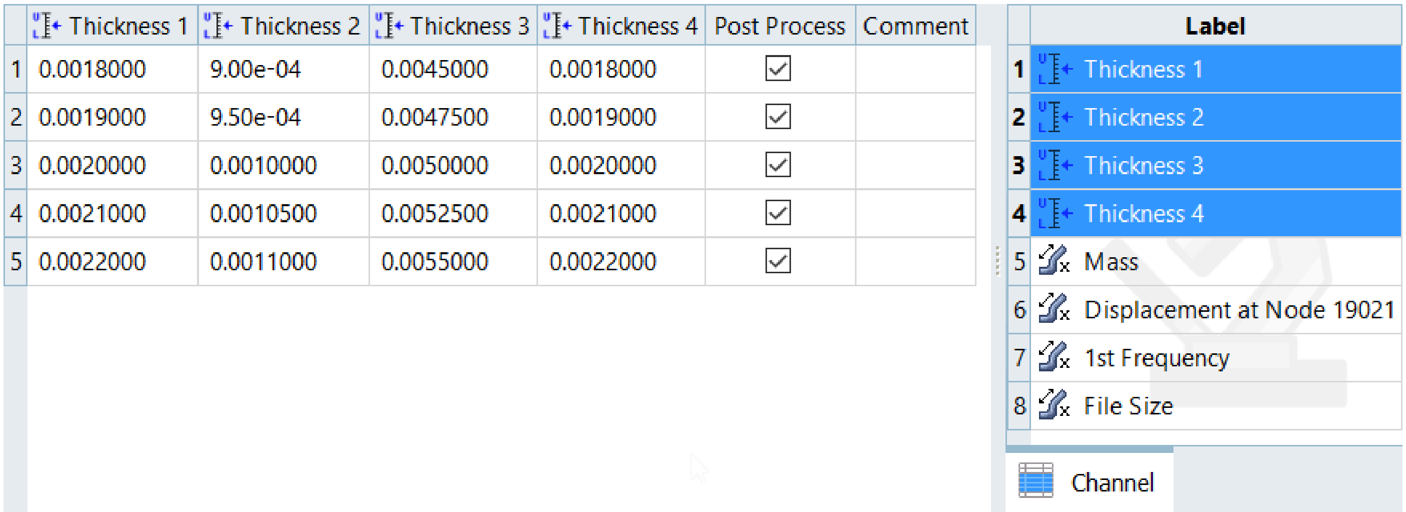


Figure 243:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

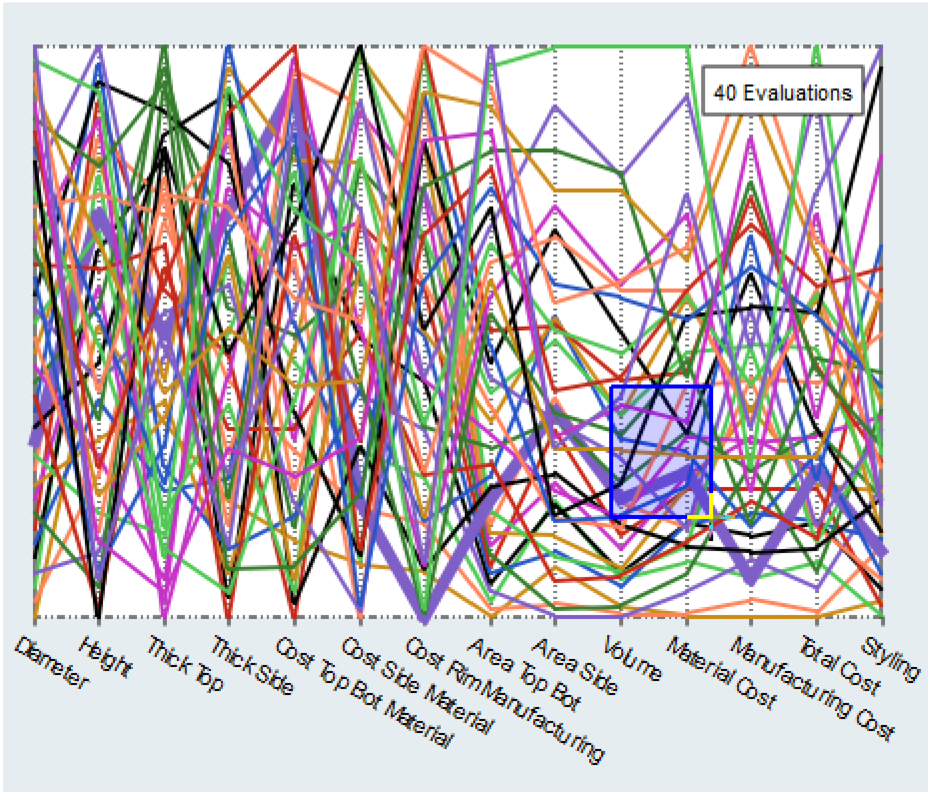
A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot.

Option	Description
	Note: Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 244 run data was selected for a set of lines. In Figure 245, you can see that when Styling is low, Height is high.

Option	Description
	 *Figure 244:*

Option	Description

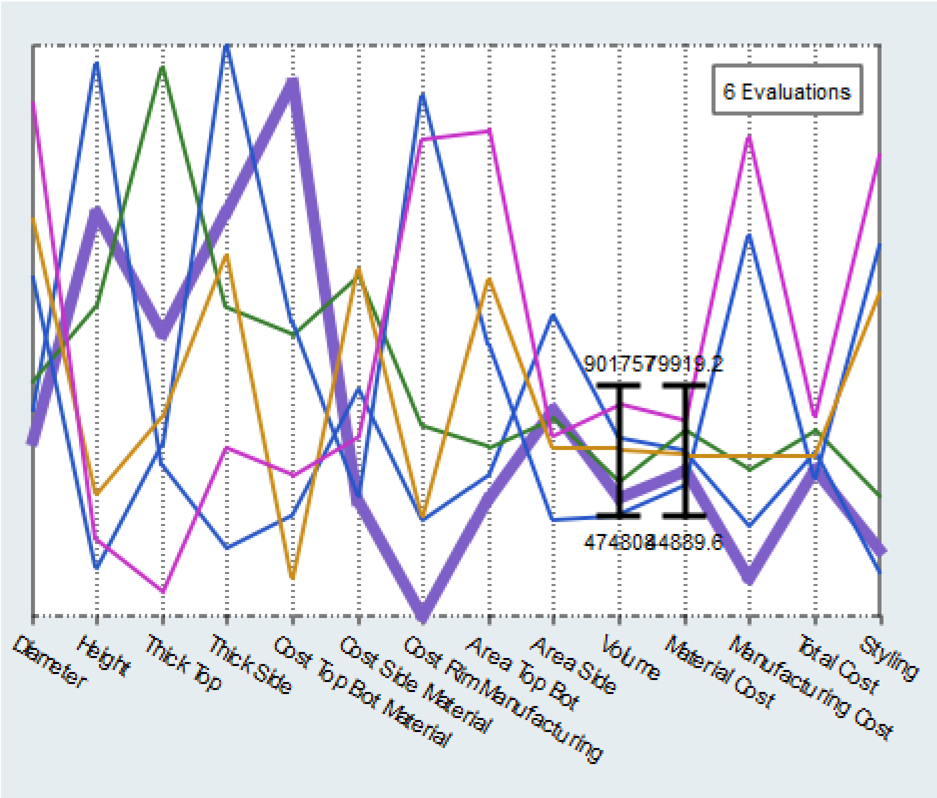


Figure 245:

Filter run data Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog.

- From the Filter column, select the input variables and output responses to plot.
- From the Filter Min and Filter Max columns, enter values to filter.

The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

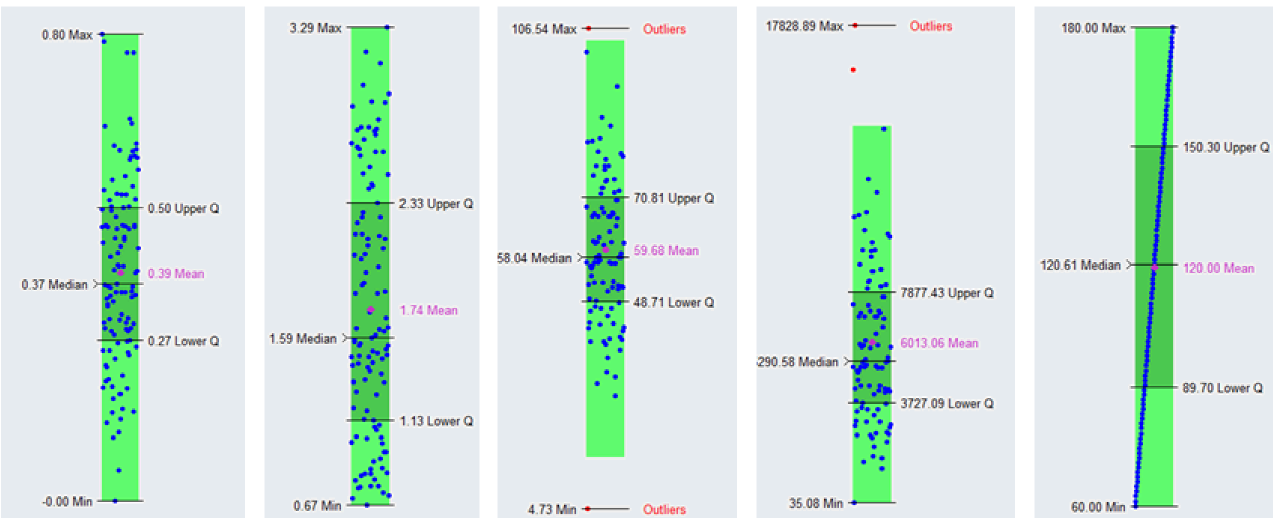


Figure 246:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 247 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

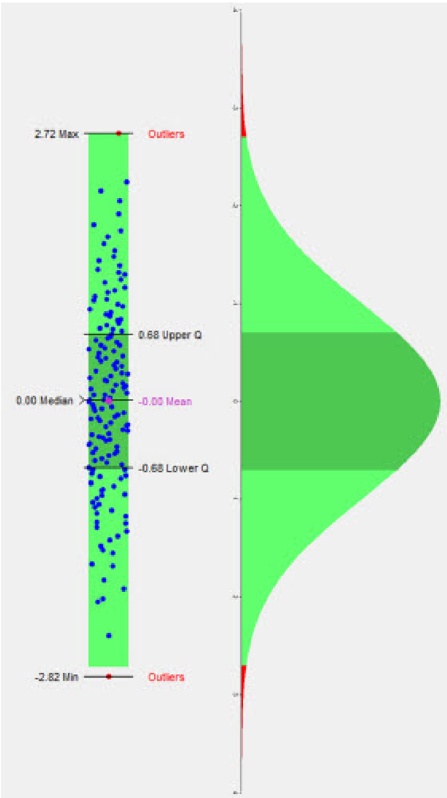


Figure 247:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

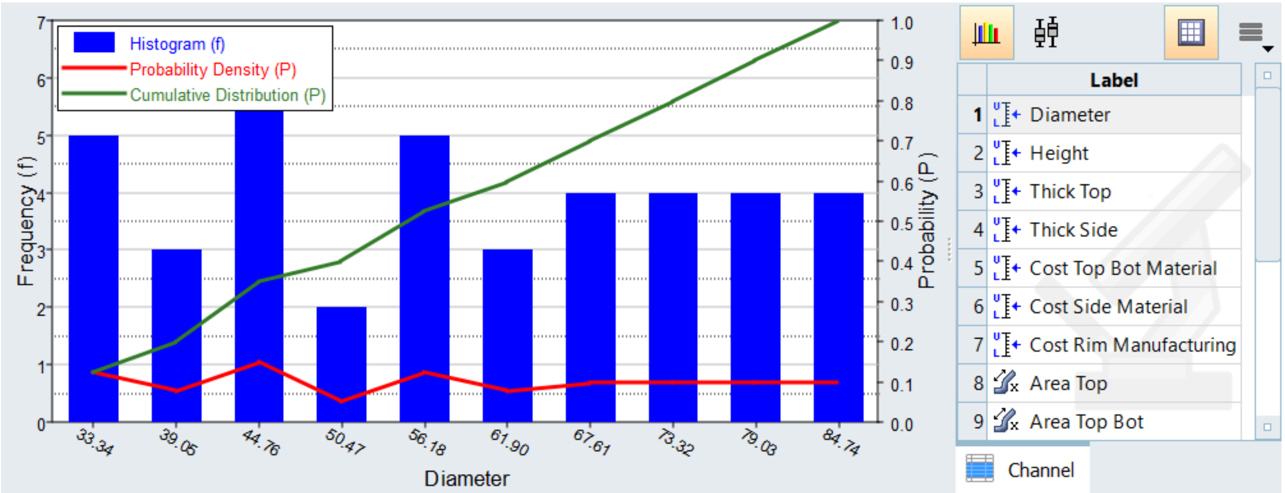


Figure 248:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processing tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Post Processing step, click the **Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

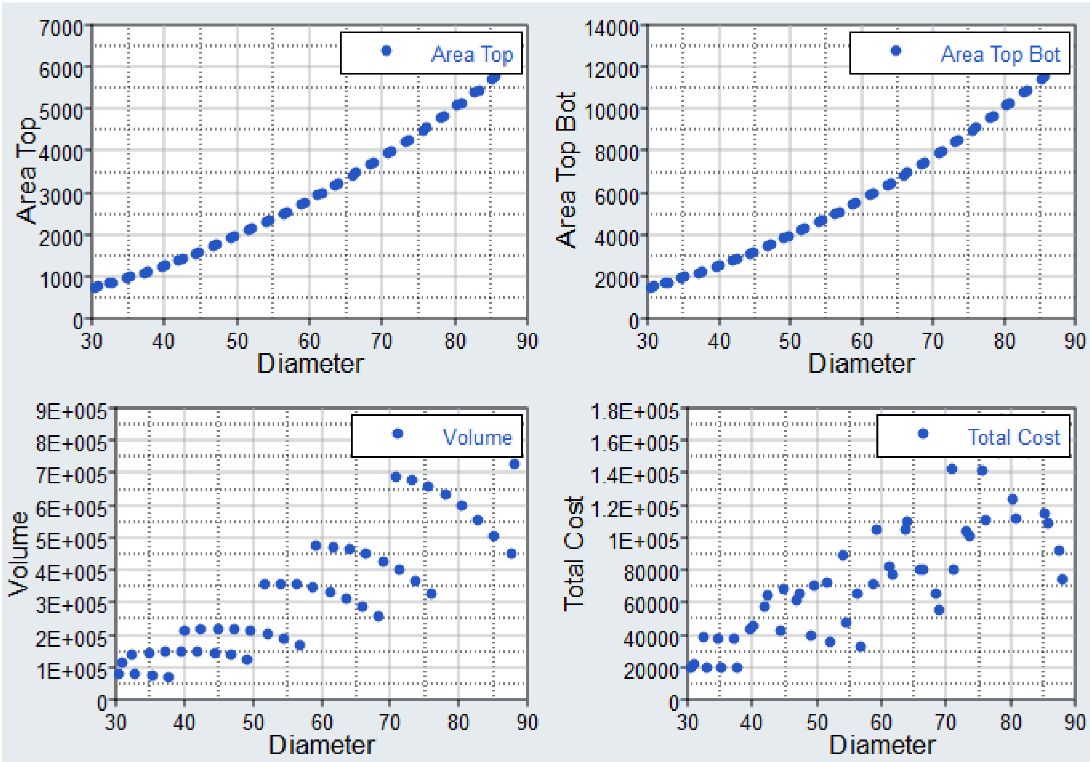


Figure 249:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

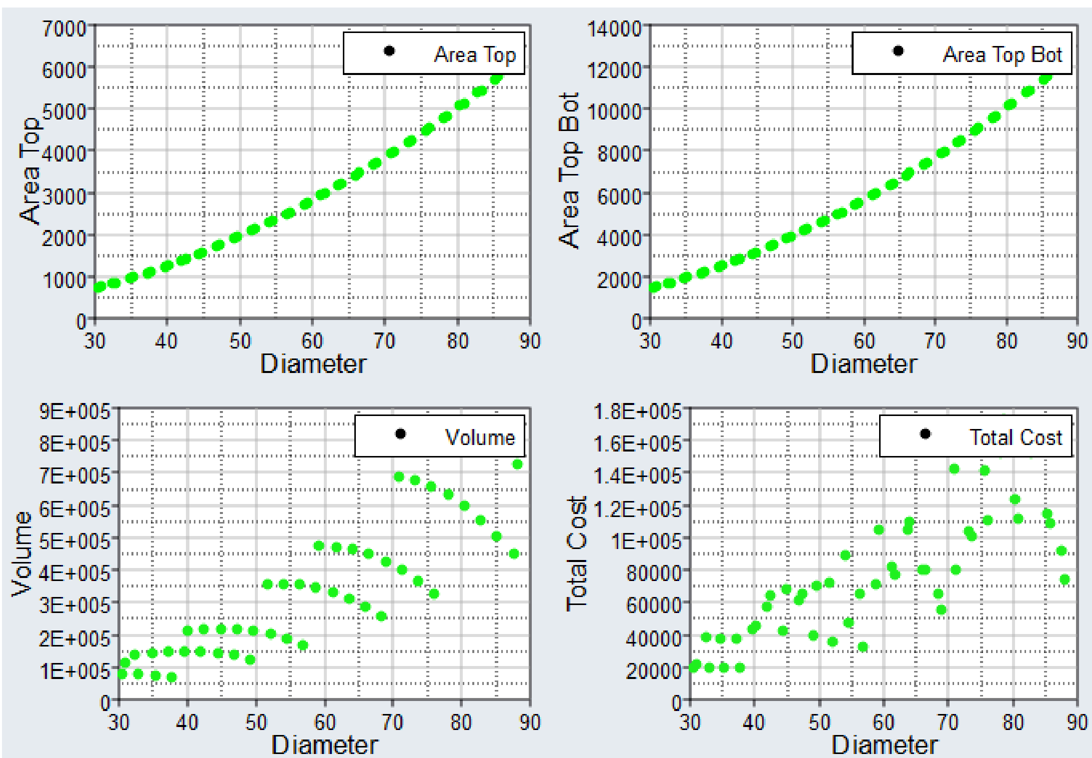


Figure 250:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

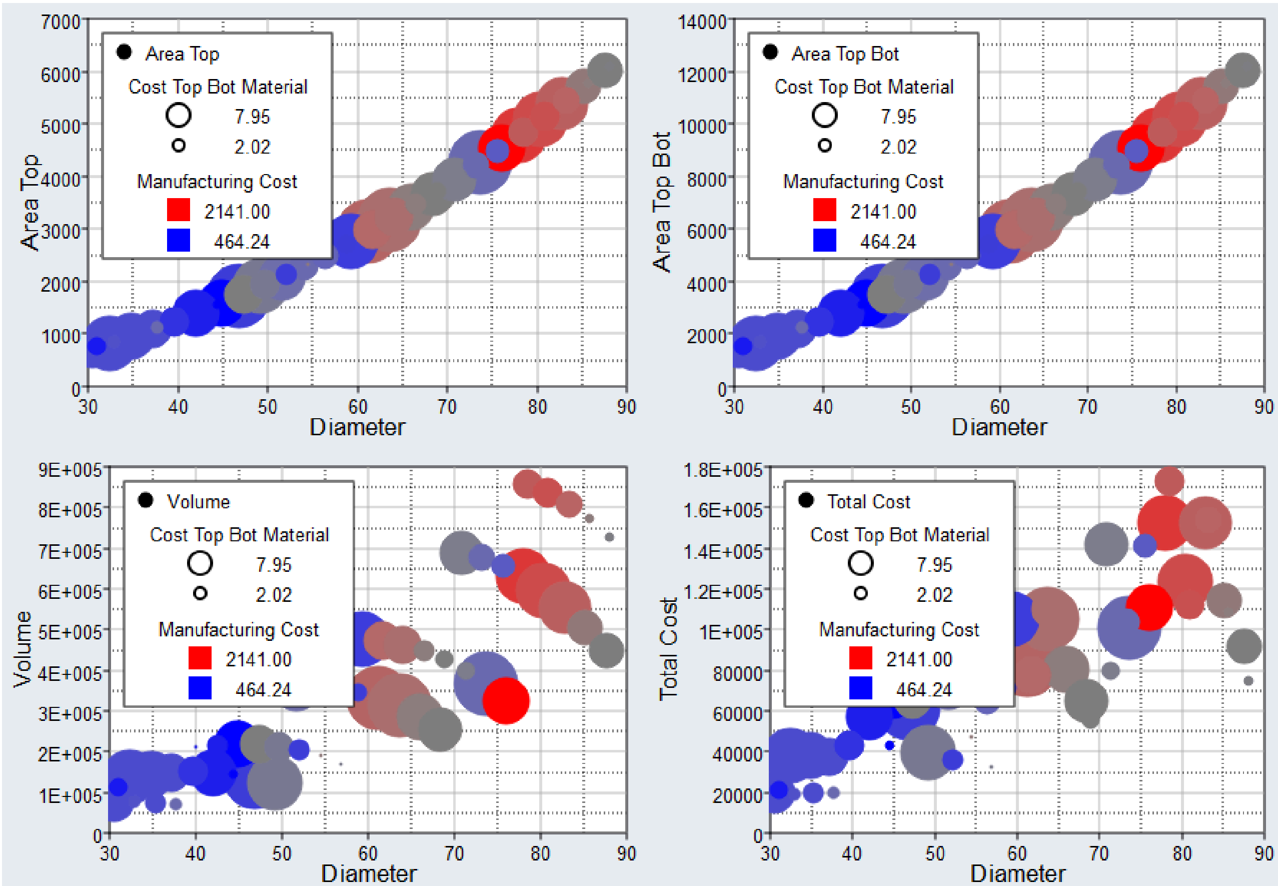


Figure 251:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Scatter Tab Settings

Settings to configure the plots displayed in the Scatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds
Display the X bounds in the plot.

Y-Bounds
Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 252: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 253: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥
Show only the column/rows with cells over the specified threshold.

Show Variables and Responses
Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

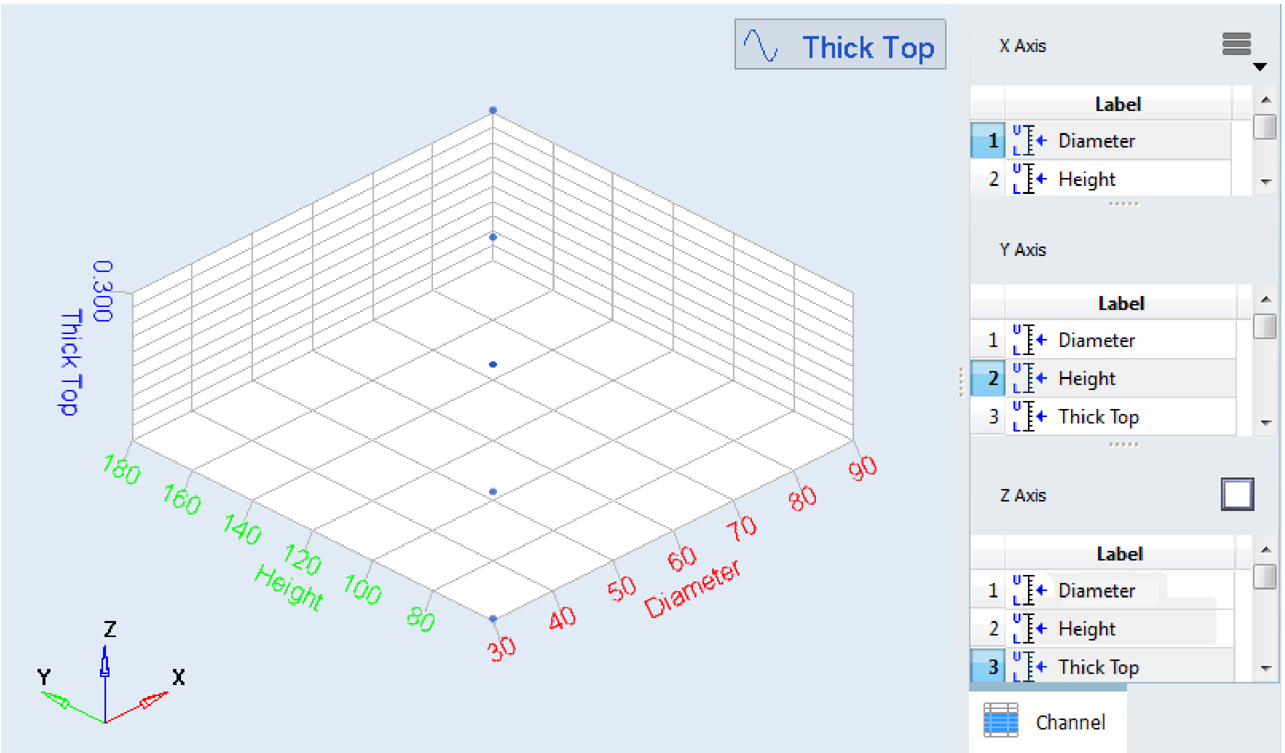


Figure 254:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

> **Tip:** For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the biplot.

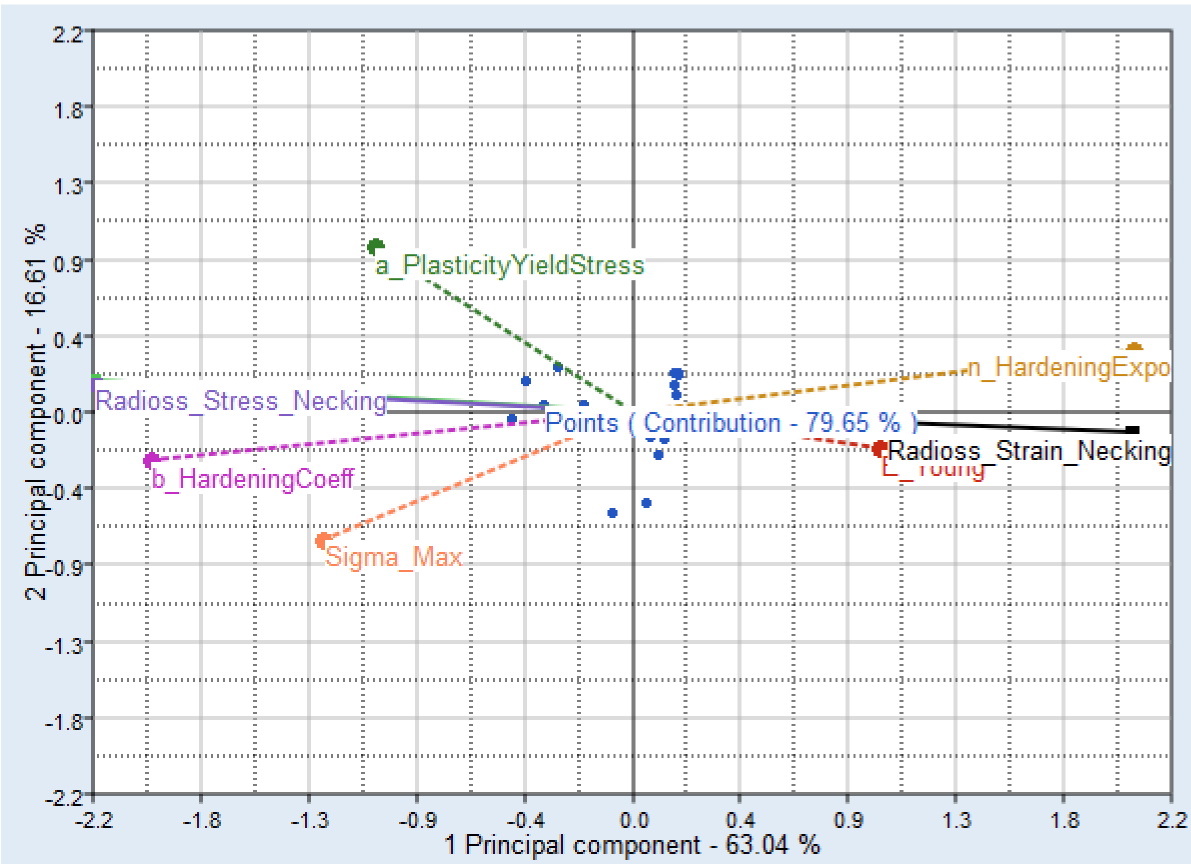


Figure 255:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
   a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

   > **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

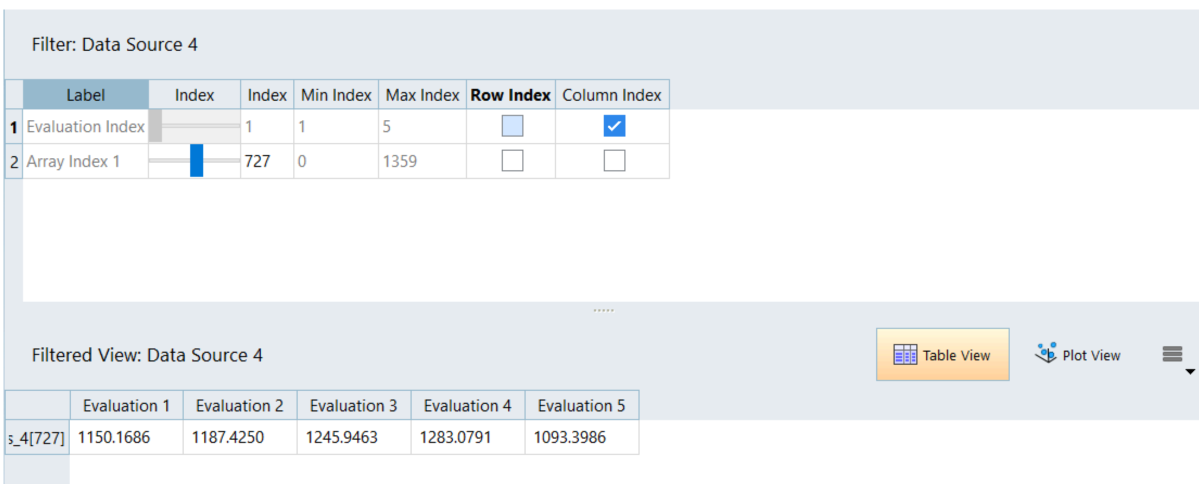


Figure 256:

5. Analyze the table.

Pareto Plot Post Processing

Plot the effects of input variables on output responses in hierarchical order (highest to lowest).

Plot the Effects of Variables on Responses in Hierarchical Order

Rank the effects of input variables on output responses in hierarchical order (highest to lowest) in the Pareto Plot post processing tab.

1. From the Post Processing step, click the **Pareto Plot** tab.

2. Using the Channel selector, select the response to plot.

   > **Tip:** Analyze multiple responses simultaneously by switching the Multiplot option to (multiple plots) and selecting the responses to plot using the Channel selector.

3. Analyze the pareto plot.

   The effect of input variables on output responses is indicated by bars. Hashed lines with a positive slope indicates a positive effect. If an input variable increases, the output response will also increase. Hashed lines with a negative slope indicates a negative effect. Increasing the input variable lowers the output response.

   A line represents the cumulative effect.

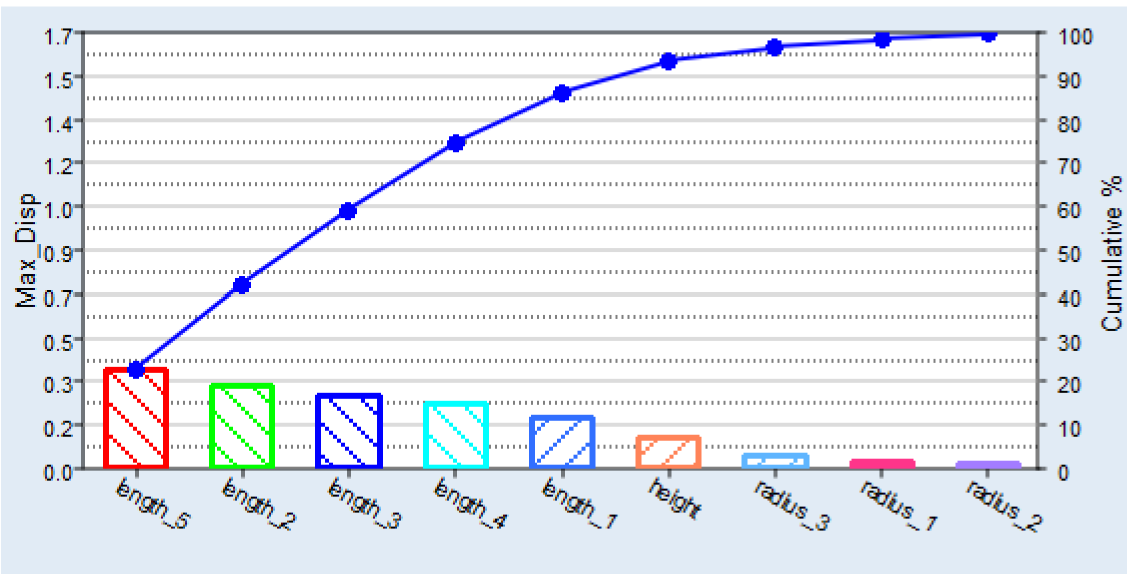


Figure 257:

Configure the pareto plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Pareto Plot Tab Settings.

Pareto Plot Tab Settings

Settings to configure the plots displayed in the Pareto Plot post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Effect curve	Show line to represent the cumulative effect.
# Top factors displayed	Specify the number of input variables (bars) displayed in the plot. **Note:** This settings does not change the calculated effects.
Multivariate Effects	Calculate the effect using all input variables simultaneously.

Linear Effects

Calculate the effect using each input variable independently.

For more information about linear effects, refer to Linear Effects Post Processing.

Include Interactions

Include first order, two way interactions along with first order effects, and calculate interactions consistently with the choice of linear or multi-variate effects.

For more information about interactions, refer to Interactions Post Processing.

Exclude dependent/linked inputs

Only show the independent input variables.

> **Tip:** Excluding dependent/link inputs reduces redundant information.

Reliability Post Processing

Lookup reliability values.

Lookup Reliability Values

Lookup specific reliability values in the Reliability post processing tab.

1. From the Post Processing step, click the **Reliability** tab.
2. Click **Add Reliability**.
3. In the Response column, select an output response.
4. In the Bound Value column, enter a threshold value for the selected output response.

The percentage of designs that would satisfy the constraint is populated in the Reliability column; this percentage is known as the reliability. The remaining percentage which indicates the probability of failure is populated in the Probability of Failure column.

	Active	Response	Bound Type	Bound Value	Reliability	Probability of Failure	Comment
1	☑	Max_Disp (r_1)	<=	3.0000000	1.0000000	0.0000000	...
2	☑	Volume (r_2)	<=	2013690.0	0.9500000	0.0500000	...
3	☑	Max_Stress (r_3)	<=	400.00000	0.9900000	0.0100000	...

Figure 258:

Reliability Plot Post Processing

Visualize representations of system reliability.

Visualize Representation of System Reliability

Visualize the representation of system reliability in the Reliability Plot post processing tab.

1. From the Post Processing step, click the **Reliability Plot** tab.
2. Using the Channel selector, select the channel(s) to plot.
3. Analyze the plot.

The relationship between the desired threshold and the reliability of the system is plotted.

Refer to the vertical axis to locate a desired reliability and find the corresponding x-axis value of the curve to identify the required threshold. For example, "To have 95% reliability I would have to design this value." You can also find a threshold from the x-axis and locate the corresponding reliability. For example, "91% of my designs are below X value". The latter use case is more common.

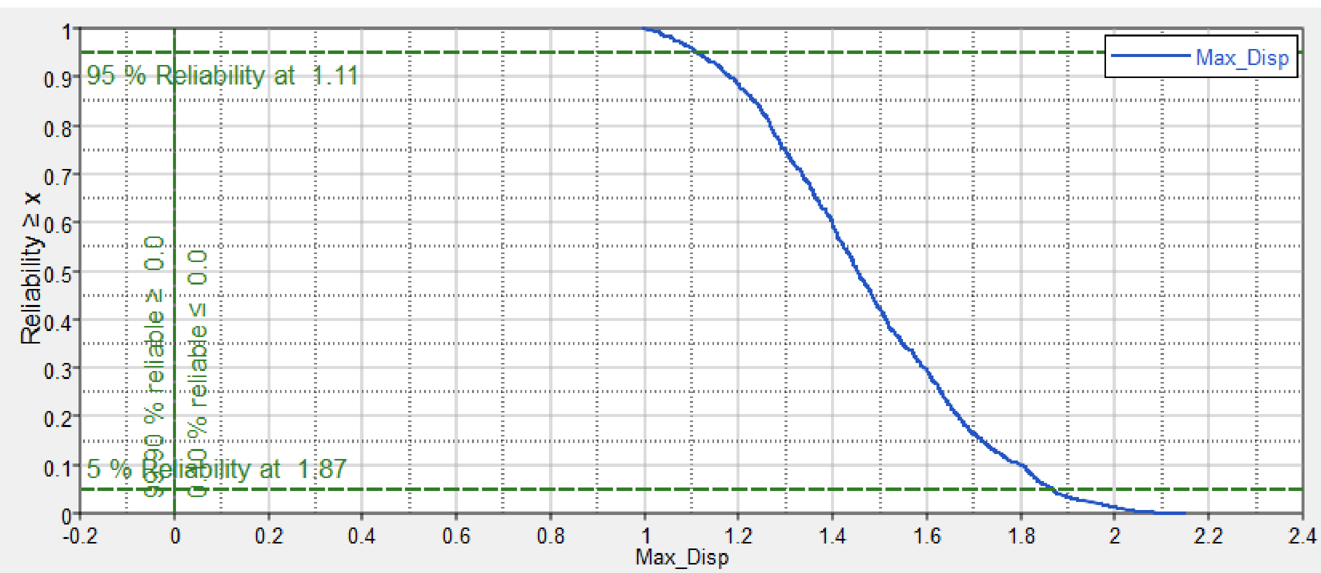


Figure 259:

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between ☐ (single plot) and ⊞ (multiple plots).

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	Description
HyperStudy Data	Generates a data report (`*.data`).
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel.

3. Click **Create Report**.

4.2.6 Setup Basic Studies

A Basic approach can be used to test nominal values and bounds by performing a nominal run, system bound check, or sweep.

Add a Basic Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
   a) In the Label field, enter a name for the Basic.
   b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.
      By default, the Definition defined in the study Setup is selected.
   c) Under Select Type, select **Basic**.
   d) Click **OK**.

A new Basic is added to the Explorer.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.
3. Test Models.
4. Define Output Responses.
5. Review definitions in the following ways:

To: | **Do this:**

Review status

Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

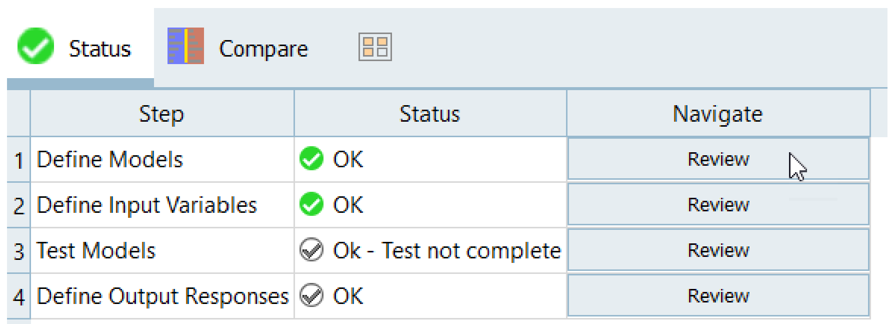


Figure 260:

Compare definitions

Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

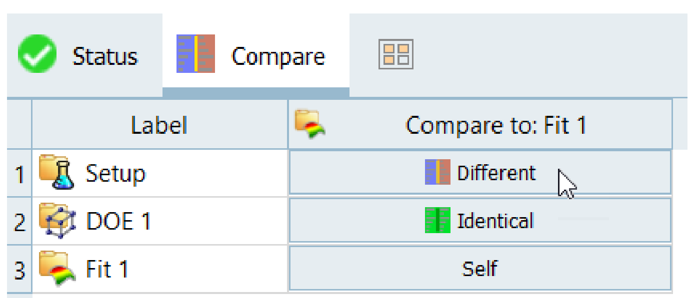
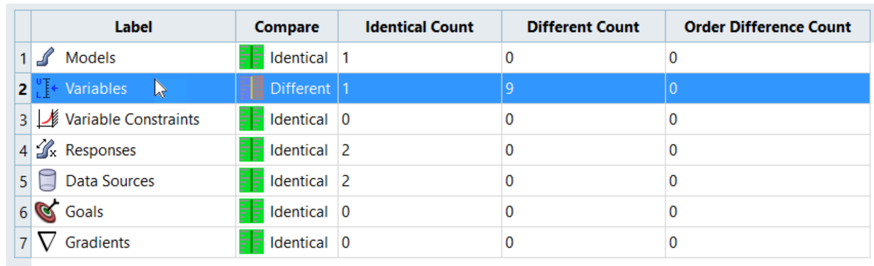


Figure 261:

To:	**Do this:**
	The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.

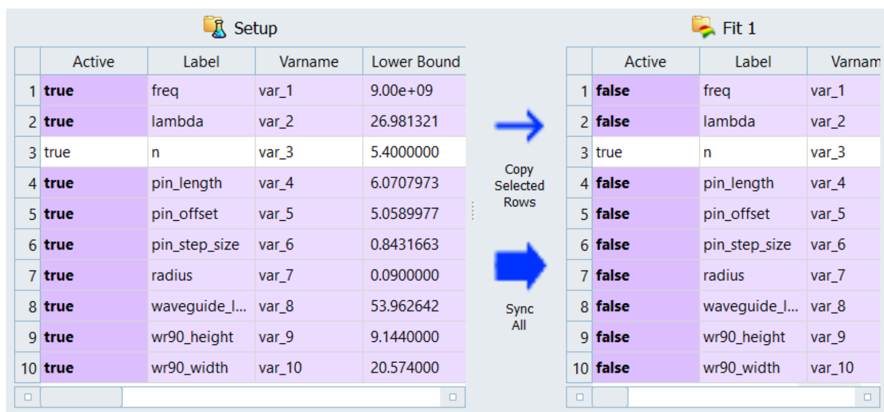
4. Click a channel to display a detailed comparison.



	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 262:

5. Sync data.
   - Click **Copy Selected Rows** to sync the single row.
   - Click **Sync All** to sync all rows.



	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 263:

Select a Numerical Method

Select a numerical method to use when evaluating the Basic approach.

1. In the Specifications step, Mode column, select a numerical method.
2. Optional: In the Settings tab, change settings as needed.
3. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog. For more information, see Edit the Run Matrix.

Basic Methods

Numerical methods available for a Basic approach.

Table 26: Basic Methods

Method	Description
Nominal Run	Runs one simulation, and sets the input variable's values to the initial values.
System Bound Check	Checks the study setup and the design space using three runs. The first run sets all of the input variables to their nominal values, the second run sets all of the values to their lower bounds, and the third run sets all of the values to their upper bounds.
Sweep	The values for a discrete input variable are iterated by index. If the number of runs exceeds the maximum index of the discrete list, the indexing scheme returns to the first index in a periodic sense.

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

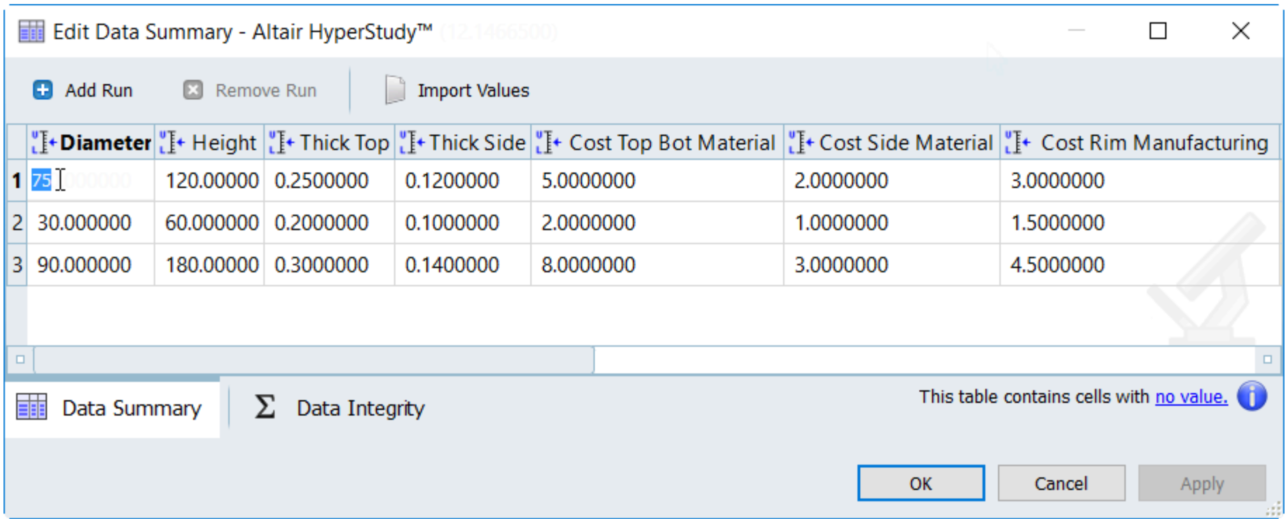


Figure 264:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.
   - Manually enter run data.
   - Copy and paste run data into the run matrix.

     Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

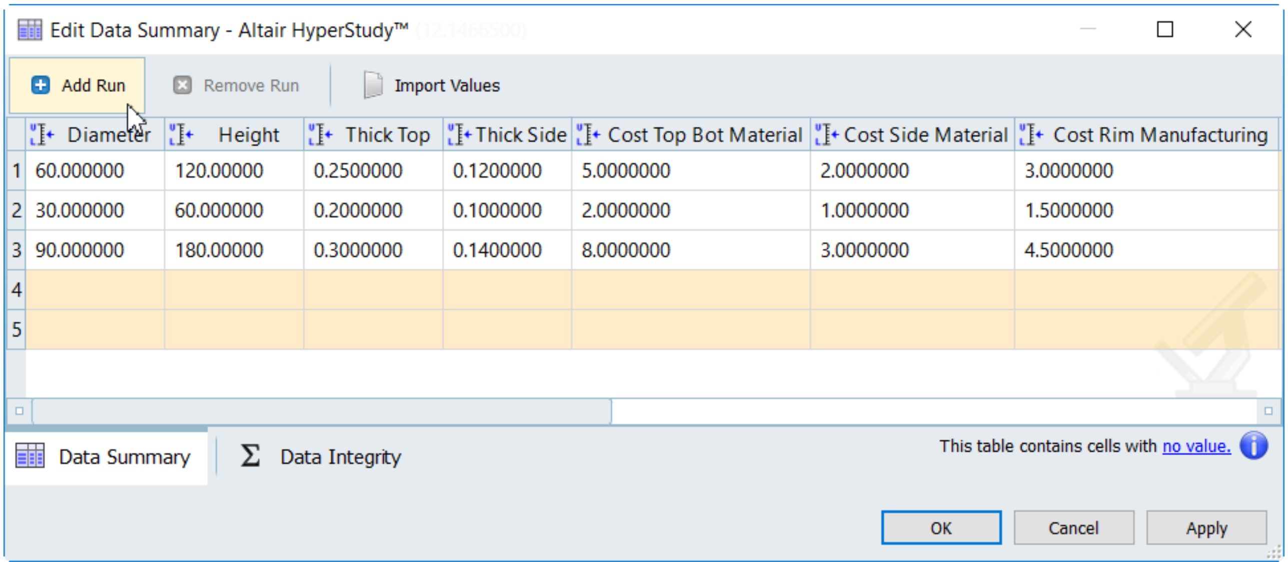


Figure 265:

Tip: Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.

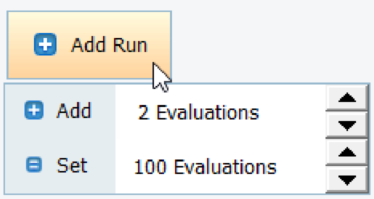


Figure 266:

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.
5. Select the source that contains run data.
   - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
   - For Approach evaluation data, select the approach that contains run data.

- For HyperStudy post processing file, select the source file.

6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.

Reuse Run Data

An Inclusion matrix contains existing data that will be appended into the newly created approach as known data points. This data typically comes from other approaches, such as DOEs or previously run Optimizations.

1. Go to the **Specifications** step for the Basic.
2. In the top, right of the work area, click **Edit Matrix** > **Inclusion Matrix**.
3. In the **Edit Inclusion Matrix** dialog, click **Import Values**.
4. In the **Import Values** dialog, select **Approach evaluation data** and click **Next**.
5. For Approach evaluation data, select the approach that contains run data.
6. Click **Next**.
7. Define the variable to column assignment(s).
    a) From the Variable to ColumnAssignment table, select a variable to which run data will be assigned.
    b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
    c) Click **Assign**.
8. Click **Finish**.
9. Review the imported run data.
10. Click **OK**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.
2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.

3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

   By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

	Active	Task	Batch
1	☐	Create Design	☐
2	☑	Write Input Files	☐
3	☑	Execute Analysis	☐
4	☑	Extract Output Responses	☐
5	☐	Purge ...	☐
6	☐	Create Reports	☐

Figure 267:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to stop the submission of new jobs. Click ≡ to select an execution mode.

Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

Each evaluation is independent so multi-execute can be used to run in parallel.

Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

- Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
- Horizontal execution mode sequences the write, execute, and extract task for each run independently.
- Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

Evaluation Parameters

Modify the run environment settings for the Evaluation tasks.

1. From the Evaluation step, click the **Evaluation Parameters** tab.
2. In the Value column, modify settings accordingly.

> **Note:** Review the Effectuation column to determine the scope at which each setting takes effect.

Review Evaluation Results

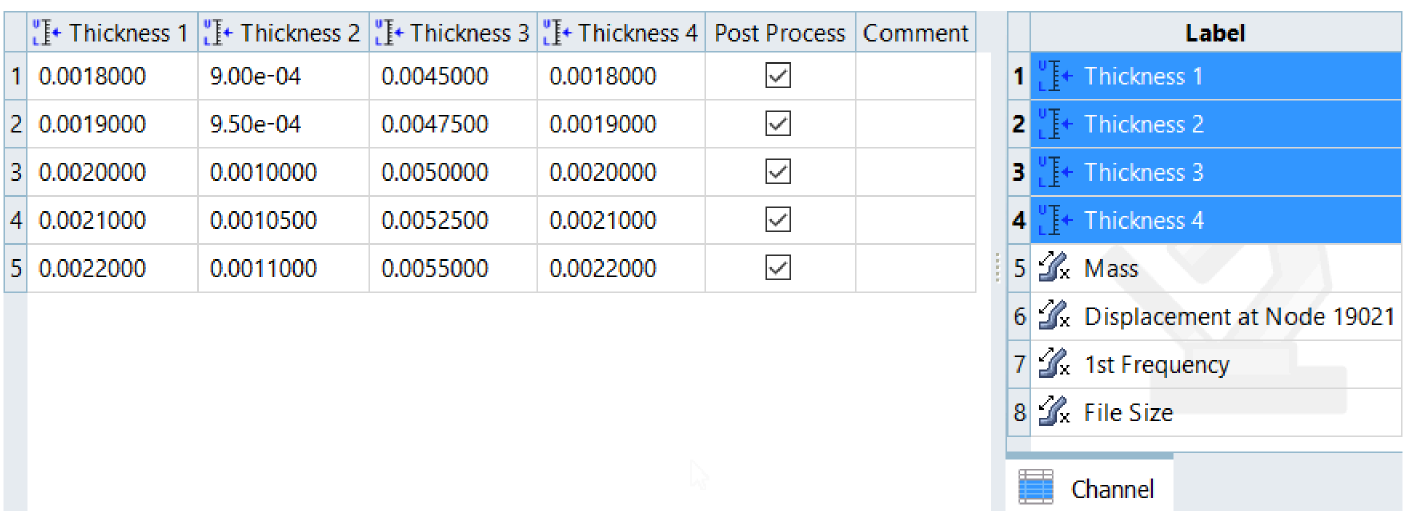
Review the input variable and output response values for each run, as well as review the run files.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Evaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.
4. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Figure 268:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

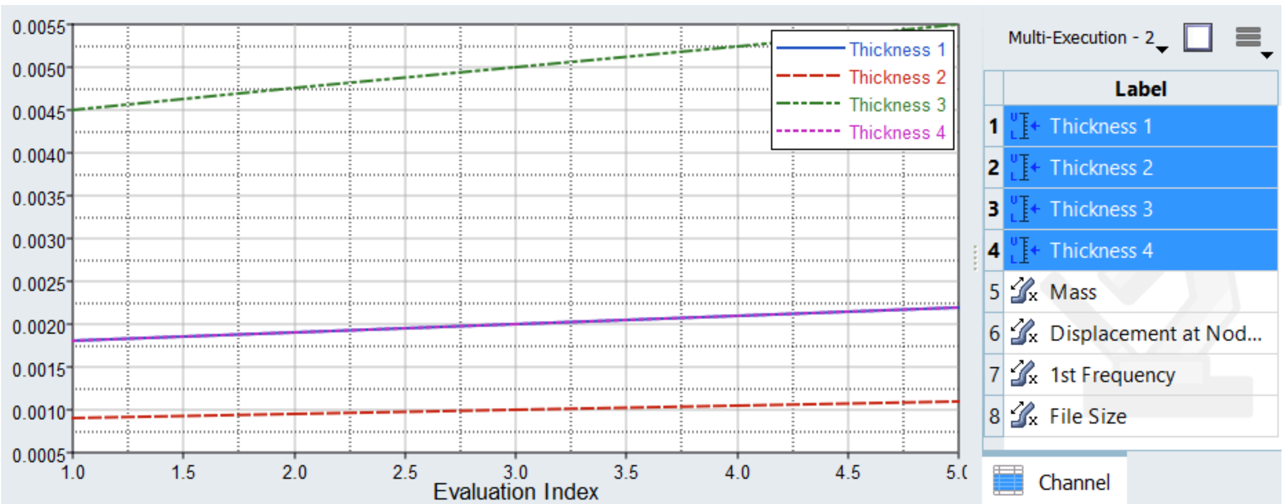


Figure 269:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Evaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. Select data to display in the scatter plot.
   - Use the Channel selector to select two dimensions of data to plot.

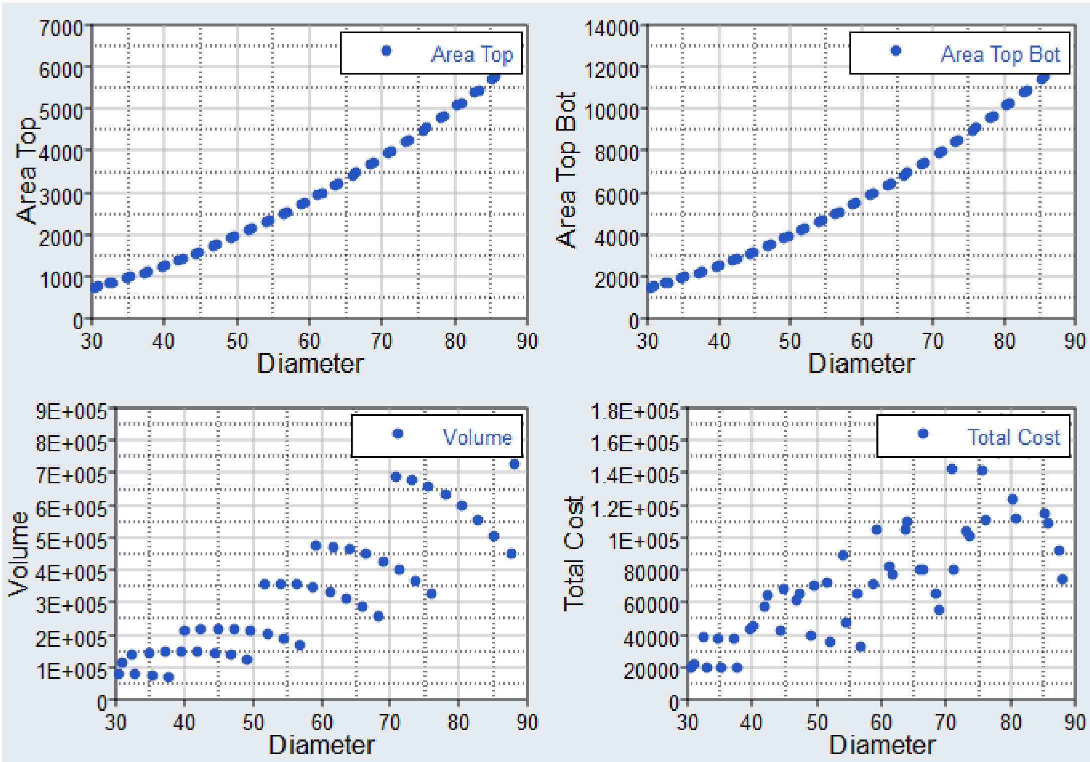


Figure 270:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

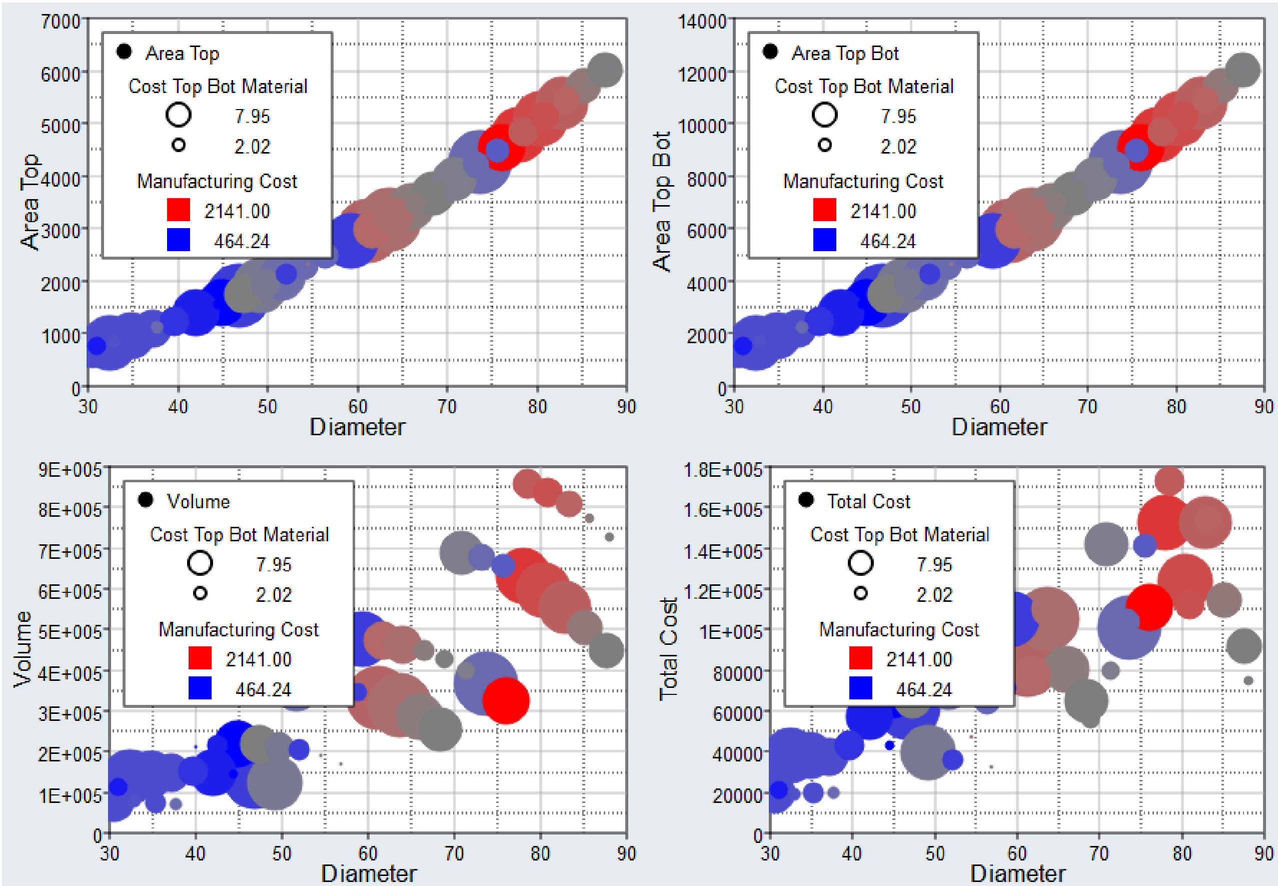


Figure 271:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation Scatter Tab Settings

Settings to configure the plots displayed in the Evaluation Scatter tab.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.

Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Review Evaluation Time

Inspect task wall-clock times.

Review the time spent in each task within the Evaluation Time tab. Identify bottlenecks in tabular or plot form.

1. From the Evaluate step, click the **Evaluation Time** tab.
2. Use the top channel selector to select the model(s) to review.
3. Use the bottom channel selector to identify the time categorises for review.

Option	Action
Write	Time spent in the write task.
Execute	Time spent in the execute task.
Extract	Time spent in the extract task.
Model Total	Total time of the write, execute, and extract tasks.
All Models Total	Summation of all Model Totals.

Option	Action
	Note: This category is independent of the selected models.

4. Switch the view between table and plot by clicking Table or Plot, located above the Channel selector.

Evaluation Time Settings

Settings to configure the plots and tables displayed in the Evaluation Time tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Cumulative Rows
: Each row entry is a summation of its own wall time and the wall times preceding it with a lower evaluation index.

Plot Time-Unit
: Controls the units of time when plotting the wall times.

Post Processing

View the computational results from the Basic.

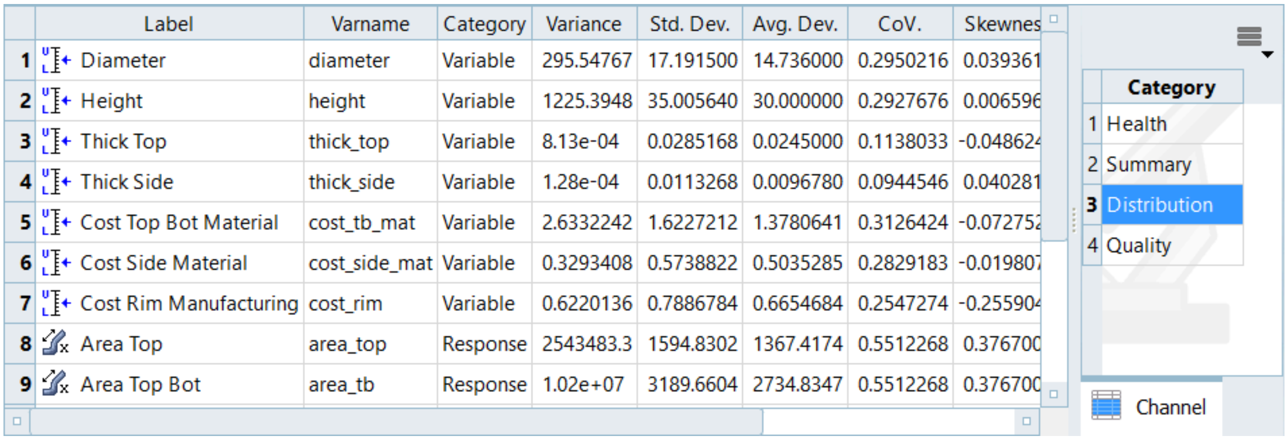
Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
   - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
   - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.
   - **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
   - **Quality** Values typically used in Quality Engineering.



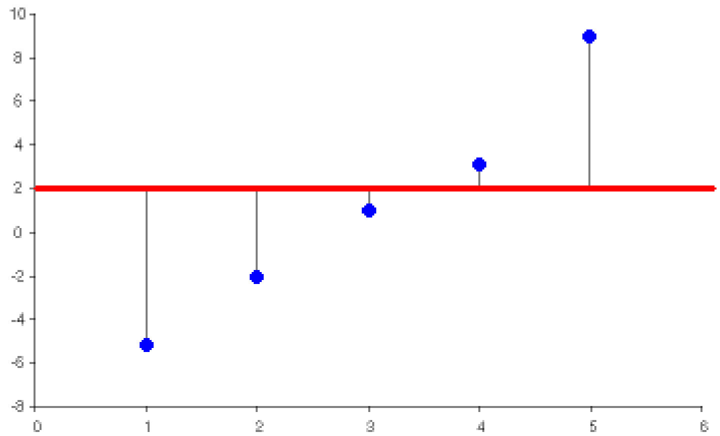
	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.072752
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.019807
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.255904
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

Figure 272:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	Description
Avg Dev (Average Deviation)	Average deviation is evaluated using: $$\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$$ In Figure 273, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 273:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $$CoV = \frac{Standard\ Deviation}{Mean}$$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $$\sqrt{\frac{\sum x_i^2}{N}}$$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $$\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}$$

Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processing tab.

1. From the Post Processing step, click the **Summary** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.

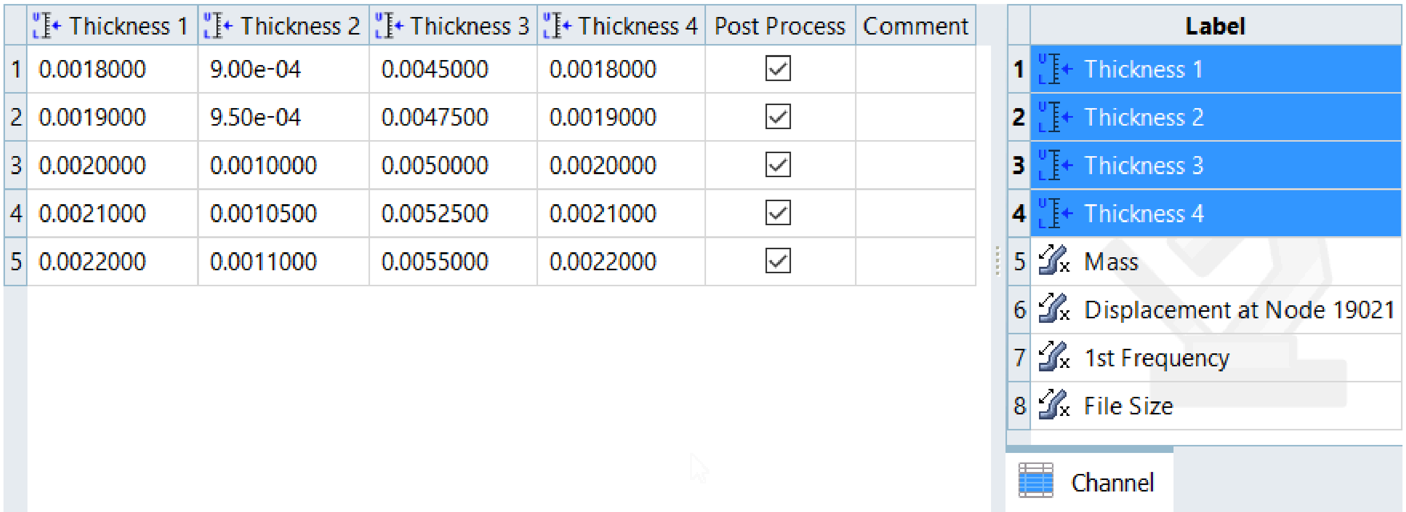


Figure 274:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot.

Option	Description
	Note: Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 275 run data was selected for a set of lines. In Figure 276, you can see that when Styling is low, Height is high.

Option **Description**

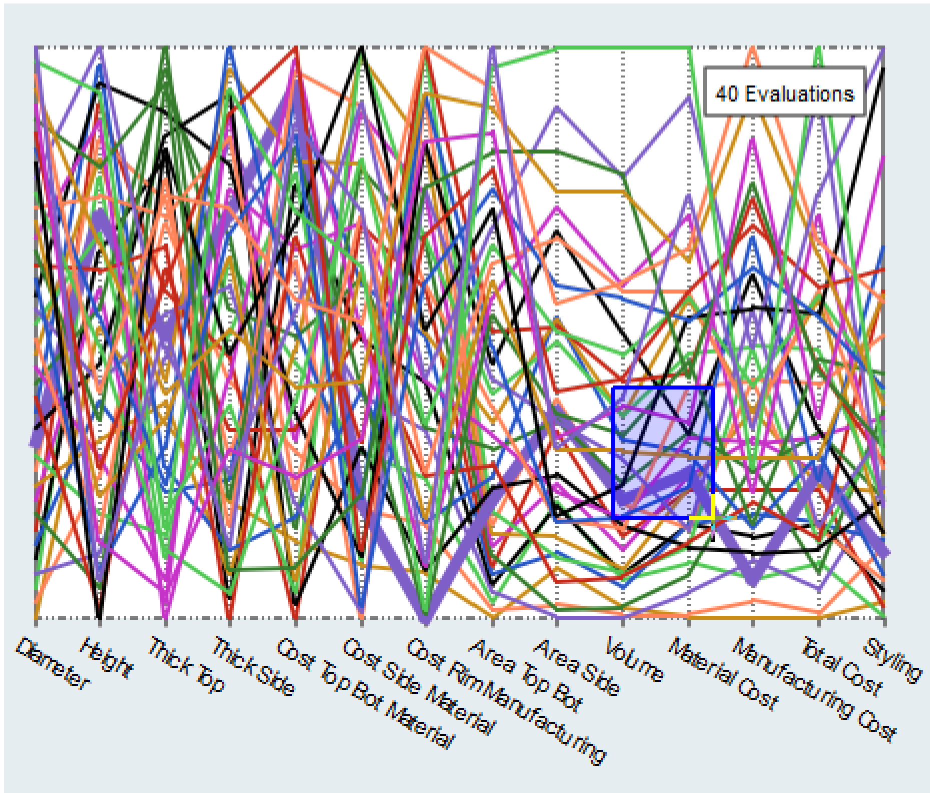
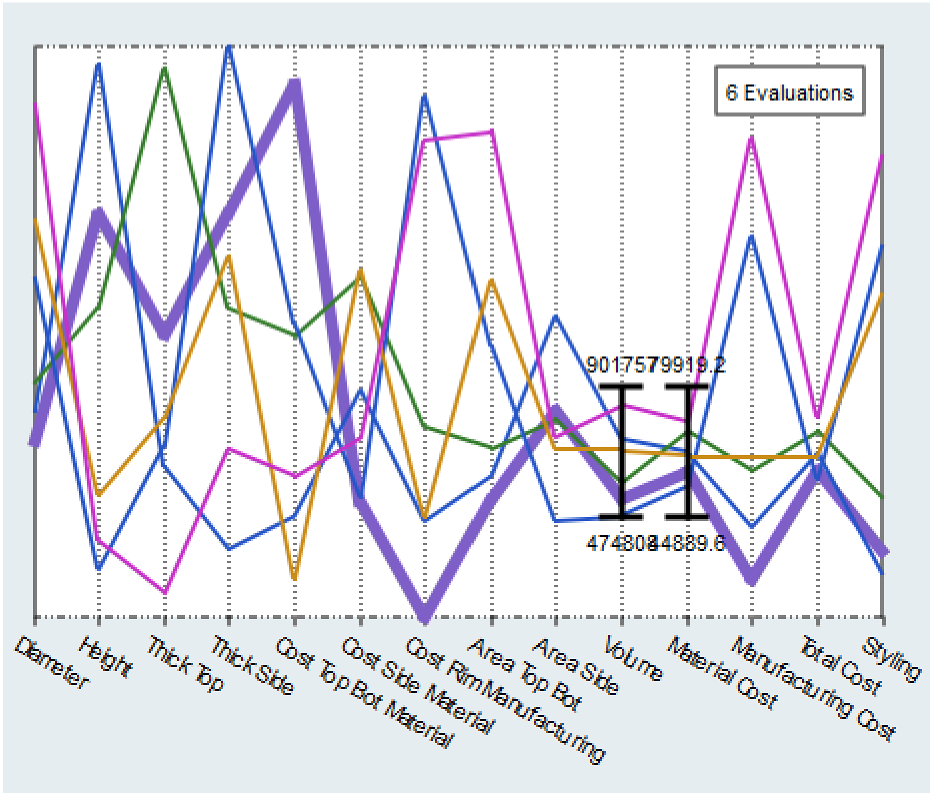


Figure 275:

Option	Description
	 *Figure 276:*
Filter run data	Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog. • From the Filter column, select the input variables and output responses to plot. • From the Filter Min and Filter Max columns, enter values to filter. The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

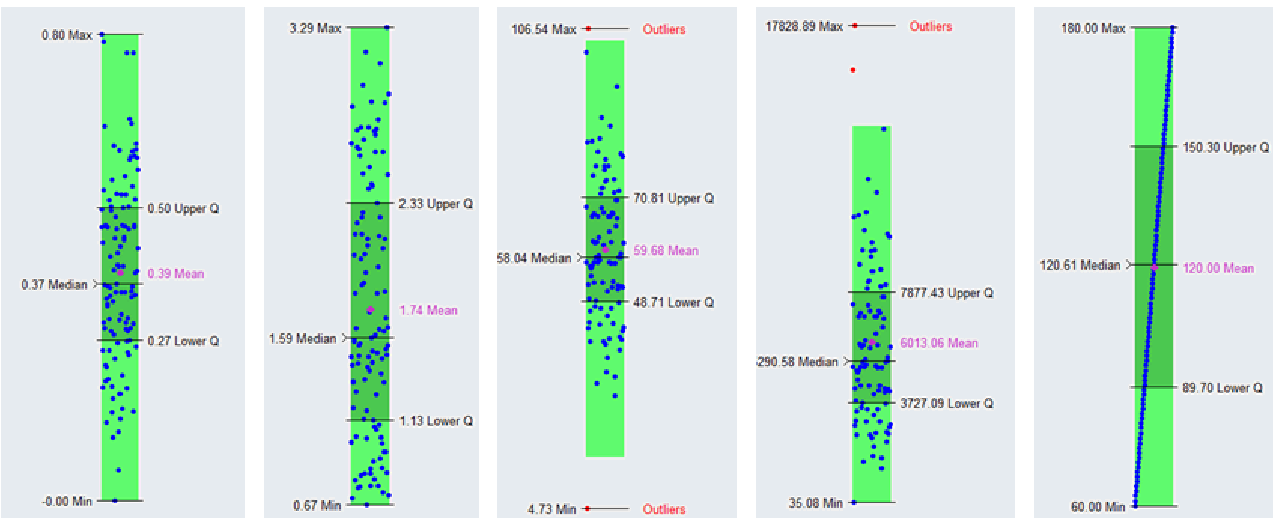


Figure 277:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 278 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

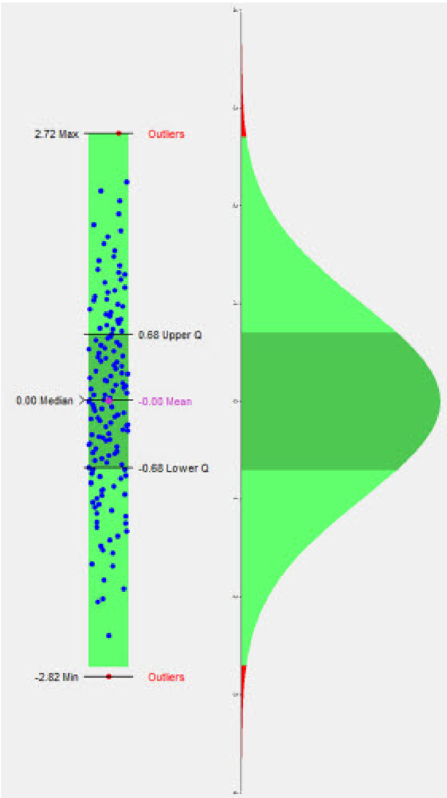


Figure 278:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

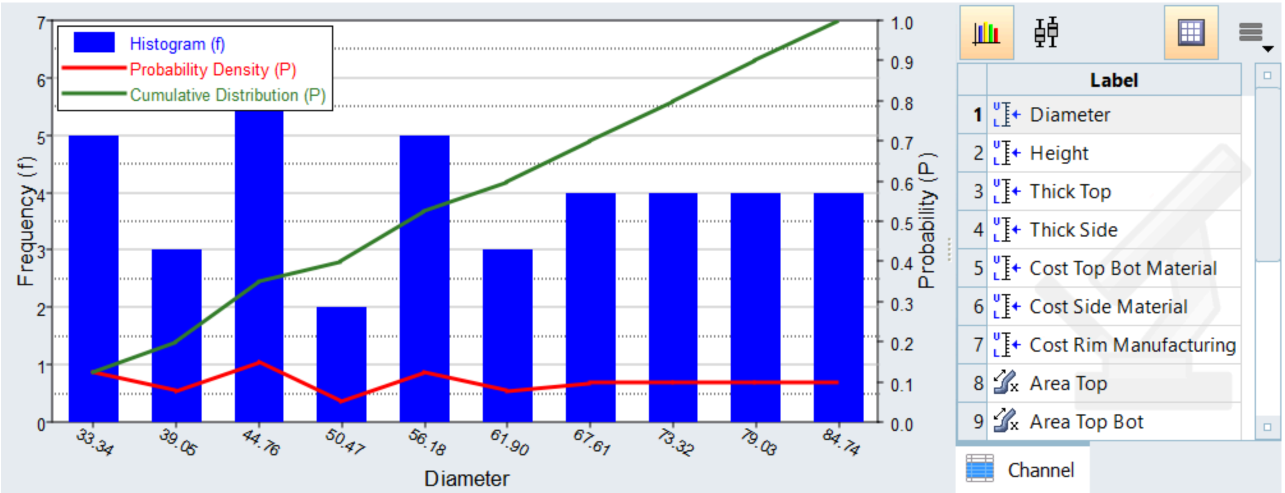


Figure 279:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processing tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Post Processing step, click the **Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

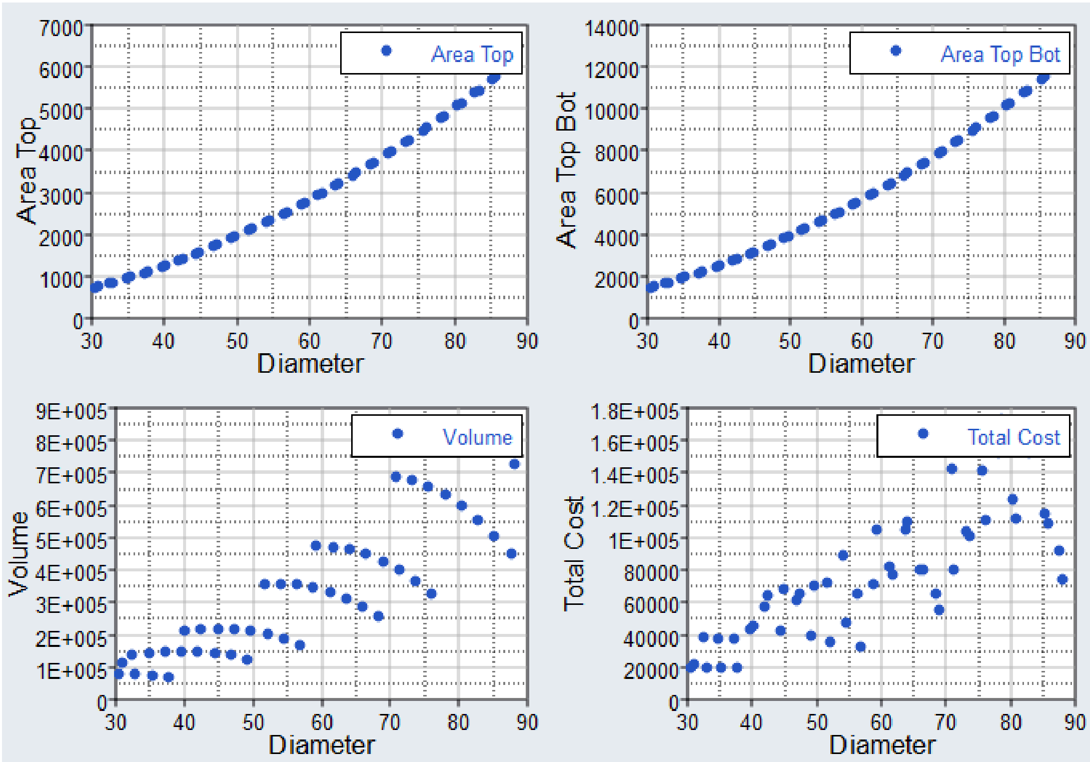


Figure 280:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

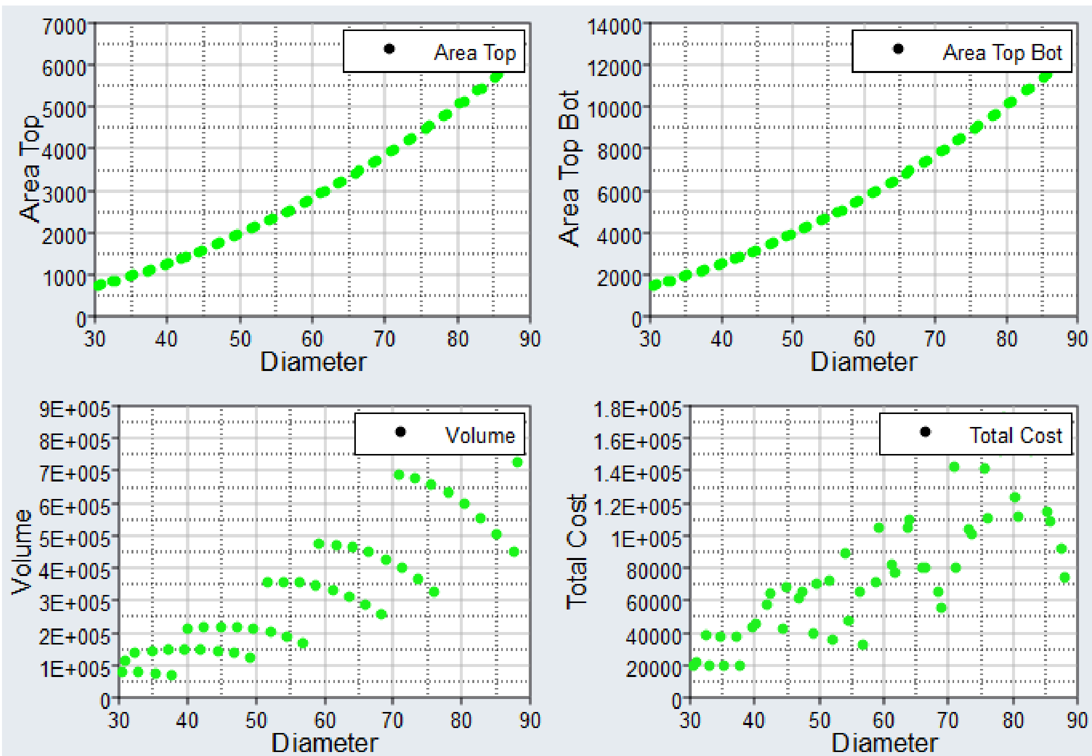


Figure 281:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

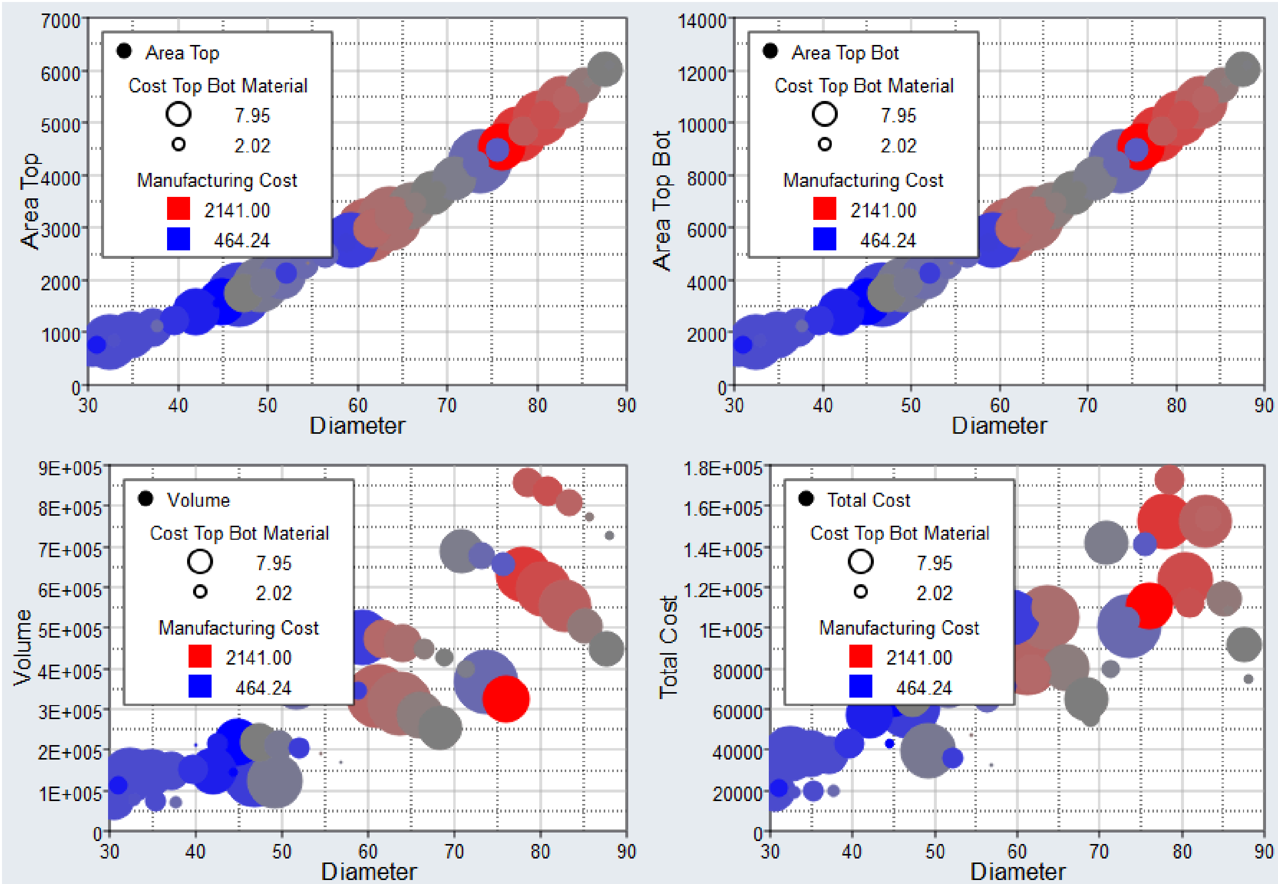


Figure 282:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Scatter Tab Settings

Settings to configure the plots displayed in the Scatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds
Display the X bounds in the plot.

Y-Bounds
Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 283: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 284: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥
Show only the column/rows with cells over the specified threshold.

Show Variables and Responses
Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

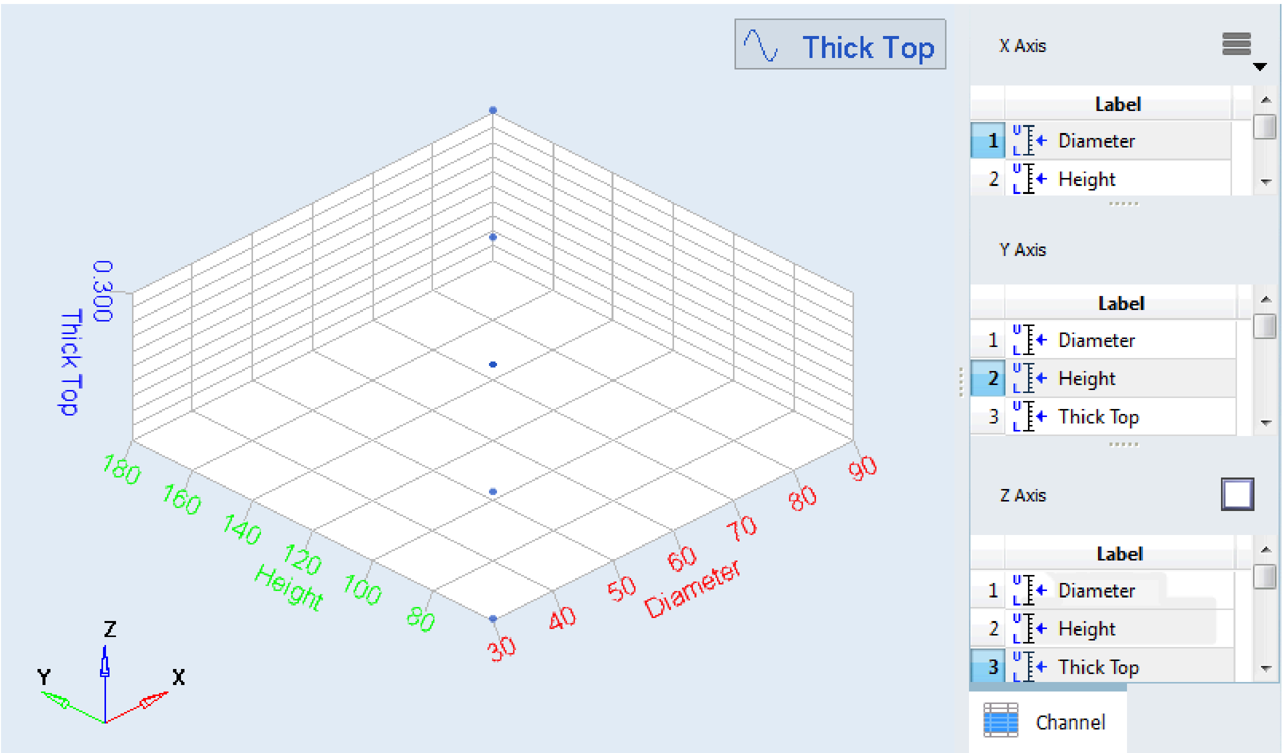


Figure 285:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

> **Tip:** For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the biplot.

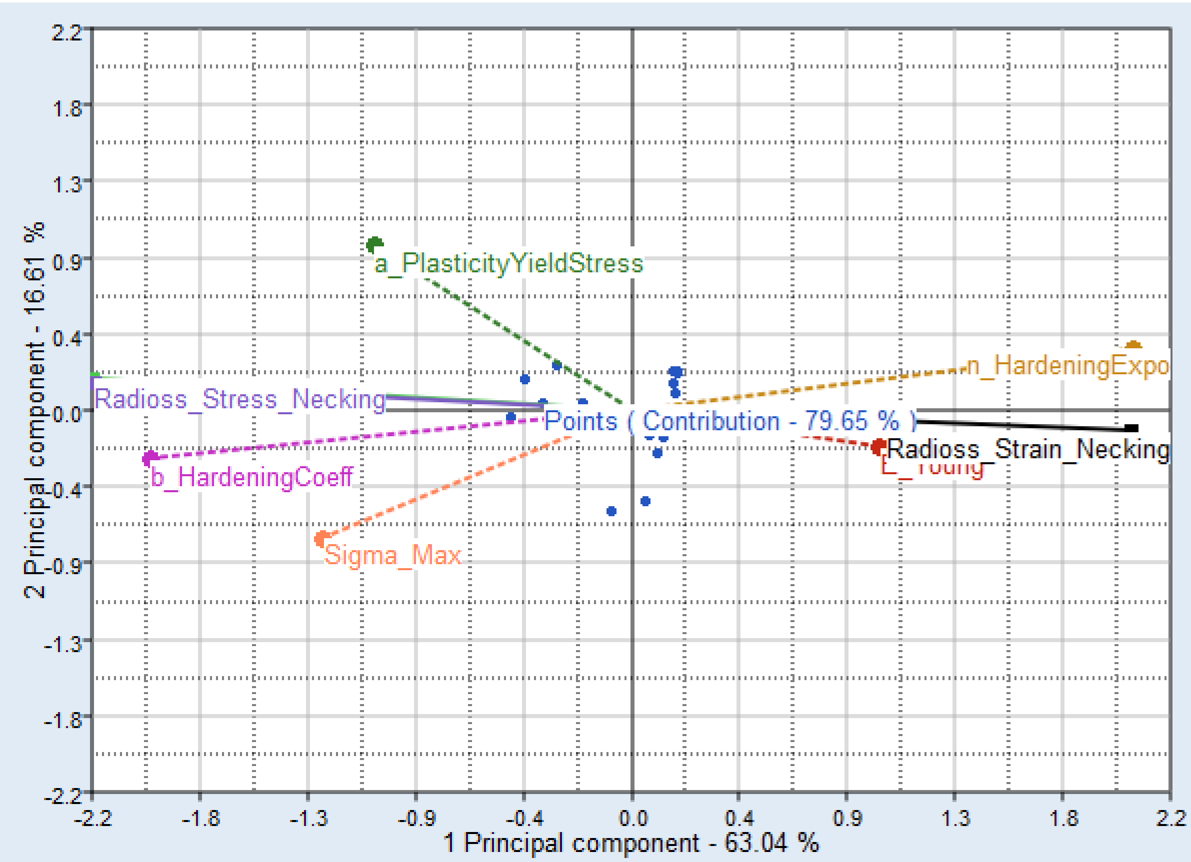


Figure 286:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
    a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

    > **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

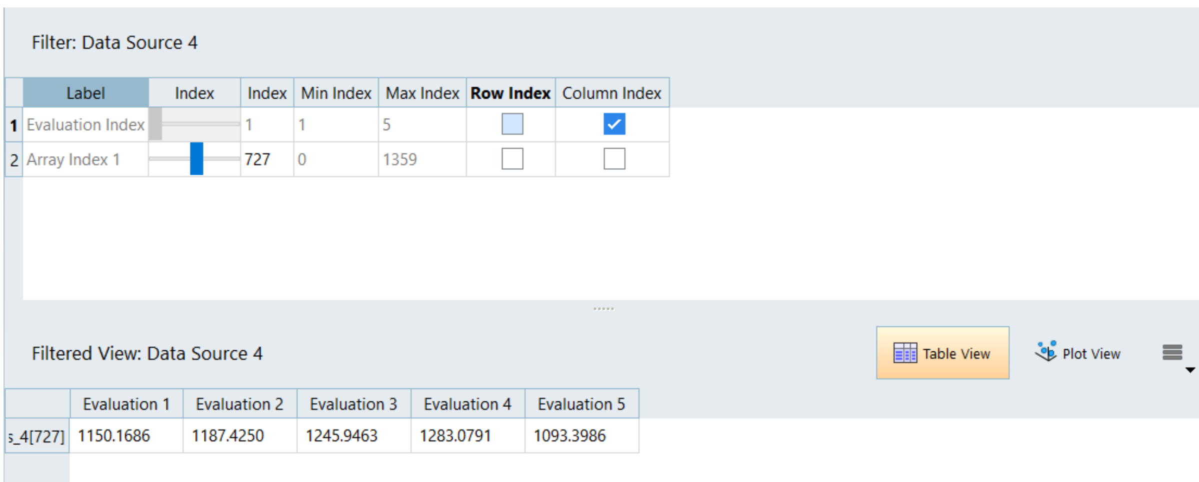
    

    Figure 287:

5. Analyze the table.

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	**Description**
HyperStudy Data	Generates a data report (`*.data`).

Report Type	Description
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel.

3. Click **Create Report**.

4.2.7 Setup Verification Studies

A Verification approach compares two data sets in a side by side comparison.

Add a Verification Approach

Add approach to the study.

1. In the Explorer, right-click and select **Add** from the context menu.
2. In the **Add** dialog perform the following steps:
    a) In the Label field, enter a name for the Verification.
    b) For Definition from, select whether to clone the Definition defined in the study Setup or an existing approach.
       By default, the Definition defined in the study Setup is selected.
    c) Under Select Type, select **Verification**.
    d) Click **OK**.

A new Verification is added to the Explorer.

> **Tip:**
> You can also add a Verification approach by highlighting specific runs in the Evaluate step (Evaluation Tasks and Evaluation Data tabs) or Post-Processing step (Summary tab) and clicking **Verify**.

Define Definition

Define the models, input variables, and output responses to be used in the study.

A Definition is used in the Setup and approaches to define the models, input variables, and output responses used in the study. When creating an approach, you can choose to clone the Definition that was defined in either the Setup or an existing approach.

1. Define Models.
2. Define Input Variables.
3. Test Models.
4. Define Output Responses.
5. Review definitions in the following ways:

To: | **Do this:**

Review status

Review the status of a Definition to verify that each step is complete.

1. Go to the **Definition** step.
2. Click the **Status** tab.

   The work area displays a status of each step in the Definition.
3. Navigate to a step in the Explorer by clicking **Review** from the Navigate column.

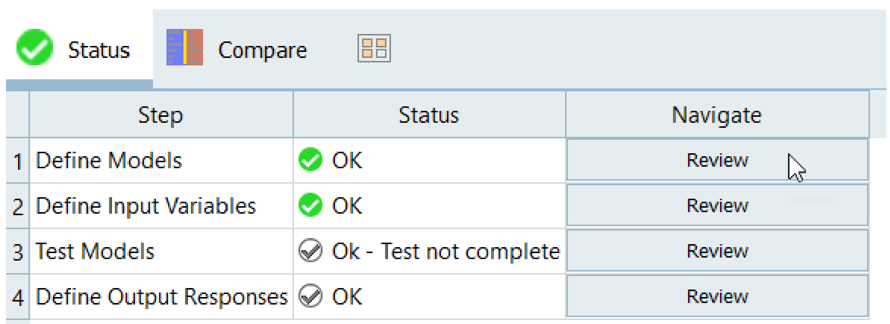


Figure 288:

Compare definitions

Compare a Definition with others in the study to identify which are identical or different.

1. Go to the **Definition** step.
2. Click the **Compare** tab.

   The work area displays a list of Definitions in the study, and indicates which are identical or different.
3. From the Compare to: column, click **Identical** or **Different**.

To:	Do this:

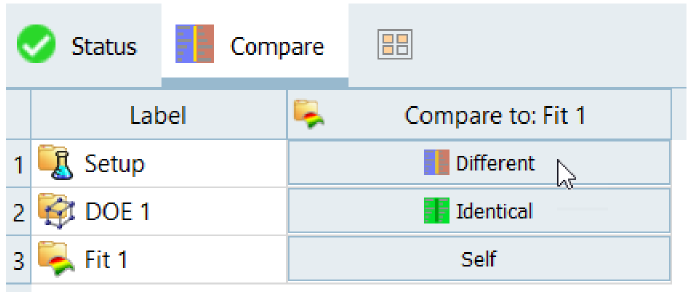
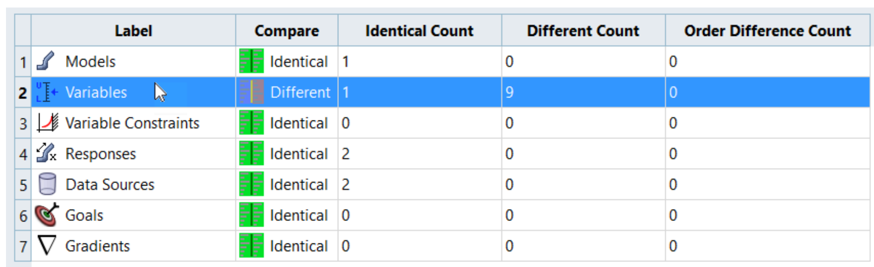


Figure 289:

The **Compare Definitions** dialog opens. A list of the different types of channels used in the study is displayed, along with a count of all instances found to be identical and different.

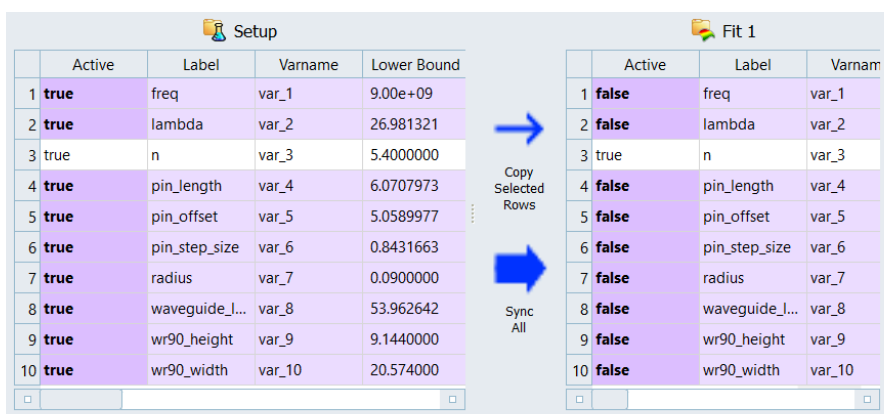
4. Click a channel to display a detailed comparison.



	Label	Compare	Identical Count	Different Count	Order Difference Count
1	Models	Identical	1	0	0
2	Variables	Different	1	9	0
3	Variable Constraints	Identical	0	0	0
4	Responses	Identical	2	0	0
5	Data Sources	Identical	2	0	0
6	Goals	Identical	0	0	0
7	Gradients	Identical	0	0	0

Figure 290:

5. Sync data.
   - Click **Copy Selected Rows** to sync the single row.
   - Click **Sync All** to sync all rows.



Setup

	Active	Label	Varname	Lower Bound
1	true	freq	var_1	9.00e+09
2	true	lambda	var_2	26.981321
3	true	n	var_3	5.4000000
4	true	pin_length	var_4	6.0707973
5	true	pin_offset	var_5	5.0589977
6	true	pin_step_size	var_6	0.8431663
7	true	radius	var_7	0.0900000
8	true	waveguide_l...	var_8	53.962642
9	true	wr90_height	var_9	9.1440000
10	true	wr90_width	var_10	20.574000

Fit 1

	Active	Label	Varnam
1	false	freq	var_1
2	false	lambda	var_2
3	true	n	var_3
4	false	pin_length	var_4
5	false	pin_offset	var_5
6	false	pin_step_size	var_6
7	false	radius	var_7
8	false	waveguide_l...	var_8
9	false	wr90_height	var_9
10	false	wr90_width	var_10

Figure 291:

Select a Numerical Method

Select a numerical method to use when evaluating the Verification.

1. In the Specifications step, Mode column, select a numerical method.
2. Optional: In the Settings tab, change settings as needed.
3. Click **Apply**.
   A run matrix is generated using the numerical method you selected.

Review and edit the run matrix in the **Edit Data Summary** dialog. For more information, see Edit the Run Matrix.

Verification Methods

Numerical methods available for a Verification approach.

Table 27: Verification Methods

Method	Basic Parameter	Comments
Verify Points	Target approach and list of evaluation indices	A general verification definition to compare a list of evaluations.
Verify Pareto	Target Optimization approach	The optimal designs from the approach are automatically selected.
Verify Trade-Off	Target Fit approach and list of input variable values	Intended to compare against input variable values coming from a Fit approach's trade-off tab.

Edit the Run Matrix

Edit the summary of run data stored in the run matrix by editing existing runs or adding new run data.

Before you can edit the Run Matrix you must select a numerical method. For more information, see Test Models.

Edit Run Data

Manually edit existing run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Enter new values in each cell, as necessary.

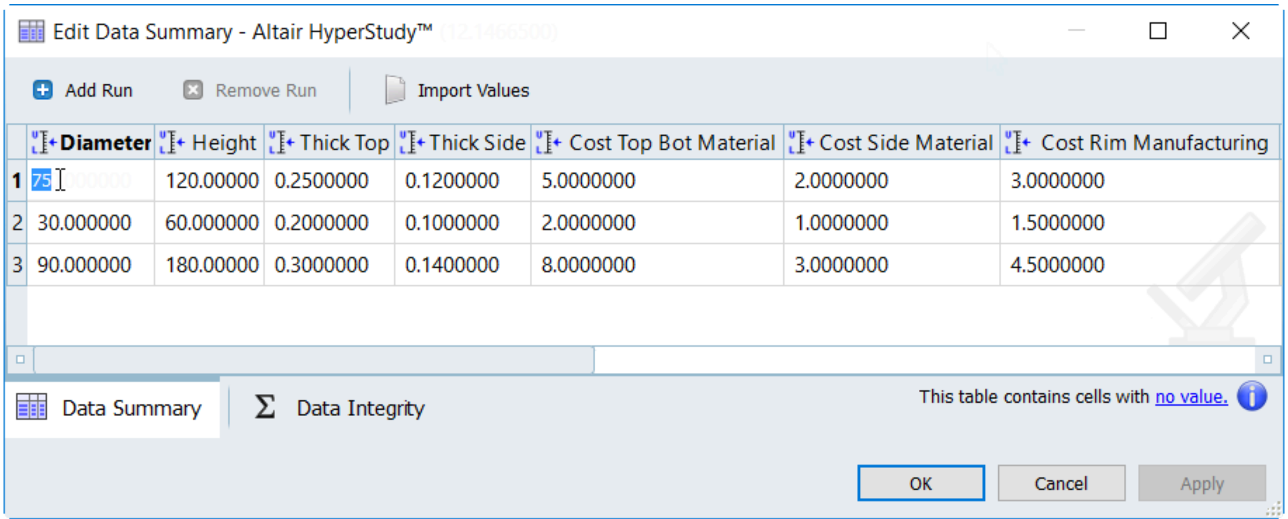


Figure 292:

Add Run Data

Manually enter new run data in the Run Matrix.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Add Run**.
3. Enter run data.
   - Manually enter run data.
   - Copy and paste run data into the run matrix.

     Example: Copy run data from a spreadsheet, then highlight and right-click on the new runs you added in the **Edit Data Summary** dialog and select **Paste** from the context menu.

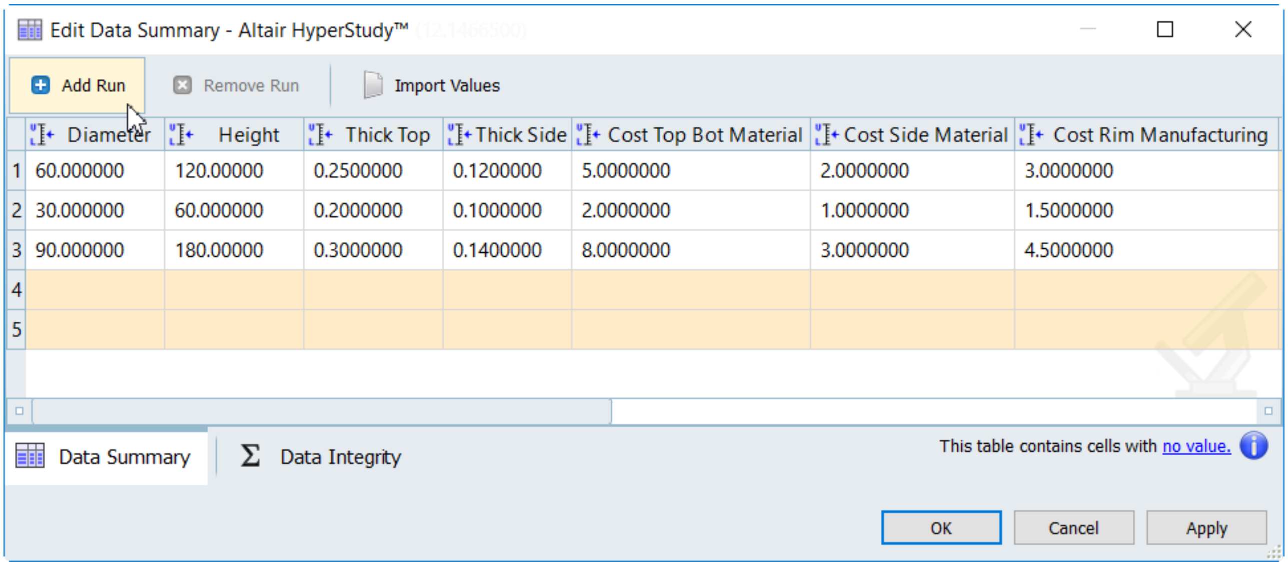
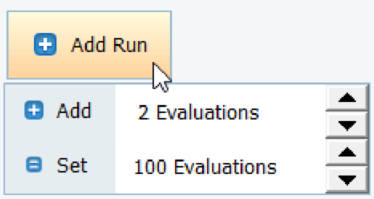


Figure 293:

> **Tip:** Add multiple runs simultaneously by left-clicking and holding the mouse button on **Add Runs**. In the pop-up, enter the number of runs to add and press **Enter**.
>
> 
>
>
> *Figure 294:*

Import Run Data

Import run data into the run matrix from a plain text file, an approaches' evaluation data, or from a HyperStudy post processing file.

1. In the Specifications step, click **Edit Matrix** > **Run Matrix**.
   The **Edit Data Summary** dialog opens.
2. Click **Import Values**.
   The **Import Values** dialog opens.
3. Select a source type.
4. Click **Next**.
5. Select the source that contains run data.
   - For Plain Text File, select the source file and delimiter type, and select whether or not the columns in the source file have labels. Optionally, specify the rows to import by entering the start and end row.
   - For Approach evaluation data, select the approach that contains run data.

- For HyperStudy post processing file, select the source file.

6. Click **Next**.
7. Define the variable to column assignment(s).
   a) From the Variable to Column Assignment table, select a variable to which run data will be assigned.
   b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
   c) Click **Assign**.
8. Click **Finish**.

Reuse Run Data

An Inclusion matrix contains existing data that will be appended into the newly created approach as known data points. This data typically comes from other approaches, such as DOEs or previously run Optimizations.

1. Go to the **Specifications** step for the Verification.
2. In the top, right of the work area, click **Edit Matrix** > **Inclusion Matrix**.
3. In the **Edit Inclusion Matrix** dialog, click **Import Values**.
4. In the **Import Values** dialog, select **Approach evaluation data** and click **Next**.
5. For Approach evaluation data, select the approach that contains run data.
6. Click **Next**.
7. Define the variable to column assignment(s).
   a) From the Variable to ColumnAssignment table, select a variable to which run data will be assigned.
   b) From the Columns in Source File table, select the column that contains run data to assign to the selected variable.
   c) Click **Assign**.
8. Click **Finish**.
9. Review the imported run data.
10. Click **OK**.

Evaluate

Run the approach.

Run Evaluation

Select which runs to evaluate and which tasks to perform.

1. Go to the **Evaluate** step.
2. In the Evaluation Tasks tab, Active column, select the runs to evaluate.

3. In the Run Tasks tab, select the checkboxes of the tasks to perform.

By default, Write Input Files, Execute Analysis, and Extract Output Responses are active.

	Active	Task	Batch
1	☐	Create Design	☐
2	☑	Write Input Files	☐
3	☑	Execute Analysis	☐
4	☑	Extract Output Responses	☐
5	☐	Purge ...	☐
6	☐	Create Reports	☐

Figure 295:

4. Define optional settings.

Setting	Action
Notification of task completion	Click ≡ and activate **Notify**.
Write solver output in Message Log and/or log-file	Click ≡ and activate **Log External Output**.
Change the number of concurrent jobs to run	Click **Multi-Execution** and enter a new value; doesn't have to be a static entry. Enter 0 to stop the submission of new jobs. Click ≡ to select an execution mode.

Multi-execute is a job management setting used to control throughput. Some algortithm's specfication settings can affect the number of jobs created per iteration. To ensure repeatability, the two settings are not tied together. However, it is recommended to coordinate the settings to ensure maximum use of resources.

Each evaluation is independent so multi-execute can be used to run in parallel.

Multi-execution runs jobs in vertical, horizontal, or horizontal (write all first) execution mode.

- Vertical execution mode performs the write, execute, and extract tasks for all designs simultaneously; that is all designs are written, then executed, then extracted.
- Horizontal execution mode sequences the write, execute, and extract task for each run independently.
- Horizontal (write all first) execution mode sequences the write task for each run first, then sequences the execute and extract tasks for each run independently.

5. Click **Evaluate Tasks**.

HyperStudy creates run files in `approaches` directory.

Verification Output Files

Output files generated from the a Verification.

<verify_variable_name>.hstds

File Creation

This file is created when **Apply** is selected during the Specifications step.

File Location

`<study_directory>/approaches/<verify_variable_name>.hstds`

File Contents

Result	Format	Description
Run Matrix Data	hstds, binary	Hstds files stores the retained data sources; direct access data using the `.hstds` file is not suggested.

<verify_variable_name>.hstdf

File Creation

This file is created when **Apply** is selected during the Specifications step.

File Location

`<study_directory>/approaches/<verify_variable_name>.hstdf`

File Contents

Result	Format	Description
Run Matrix Data	hstdf, binary	Hstdf files store the run data; however, direct access to the data using the hstdf files are not suggested.

Evaluation Parameters

Modify the run environment settings for the Evaluation tasks.

1. From the Evaluation step, click the **Evaluation Parameters** tab.
2. In the Value column, modify settings accordingly.

> **Note:** Review the Effectuation column to determine the scope at which each setting takes effect.

Review Evaluation Results

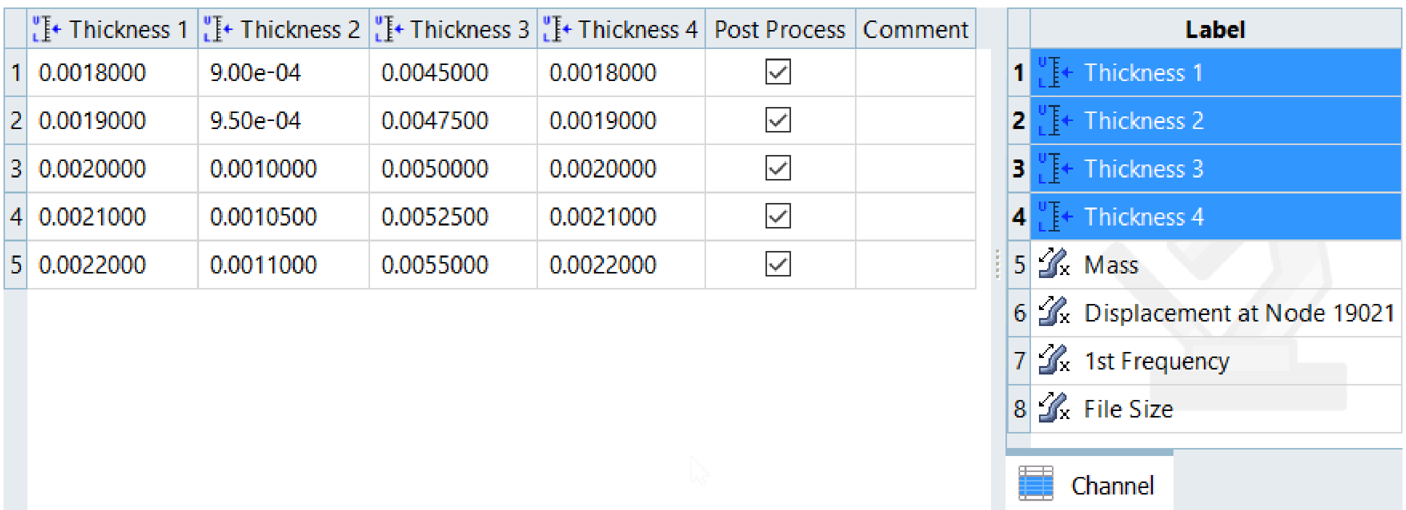
Review the input variable and output response values for each run, as well as review the run files.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Evaluation Data tab.

1. From the Evaluate step, click the **Evaluation Data** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.
4. Optional: Disable run data from post processing without deleting it entirely from the study by clearing a run's corresponding checkbox in the Post Process column.

   When a run is disabled, it will be removed from all plots, tables, and calculations in the Post Processing step.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Figure 296:

Analyze Evaluation Plot

Plot a 2D chart of the input variable and output response values for each run using the Evaluation Plot tool.

1. From the Evaluate step, click the **Evaluation Plot** tab.
2. From the Channel selector, select the input variable and/or output response to plot along the y-axis.

   The x-axis represents the run numbers.
3. Analyze the plot.

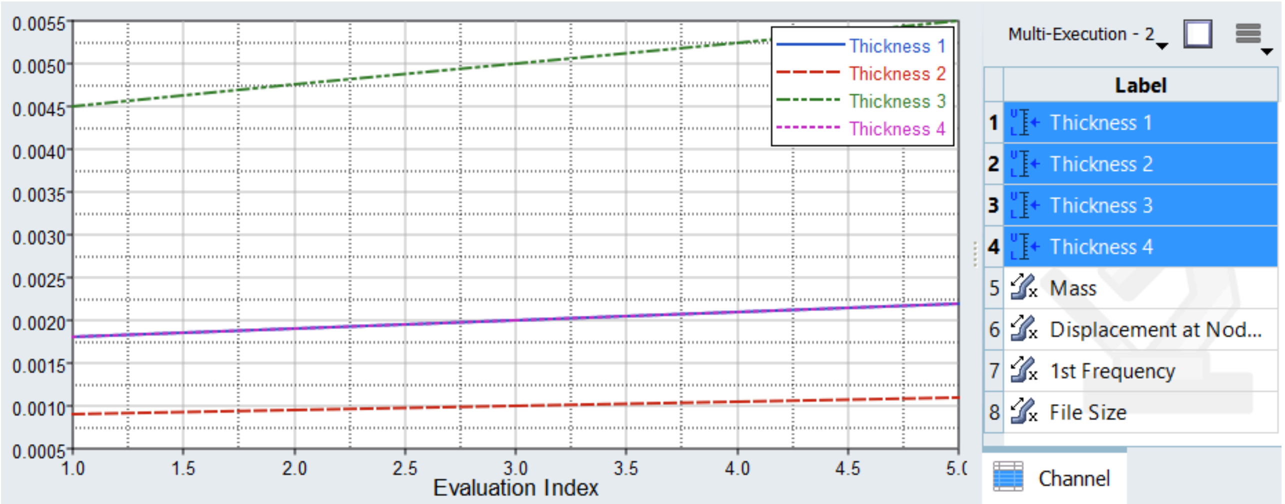


Figure 297:

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Evaluation Scatter tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Evaluate Step, click the **Evaluation Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

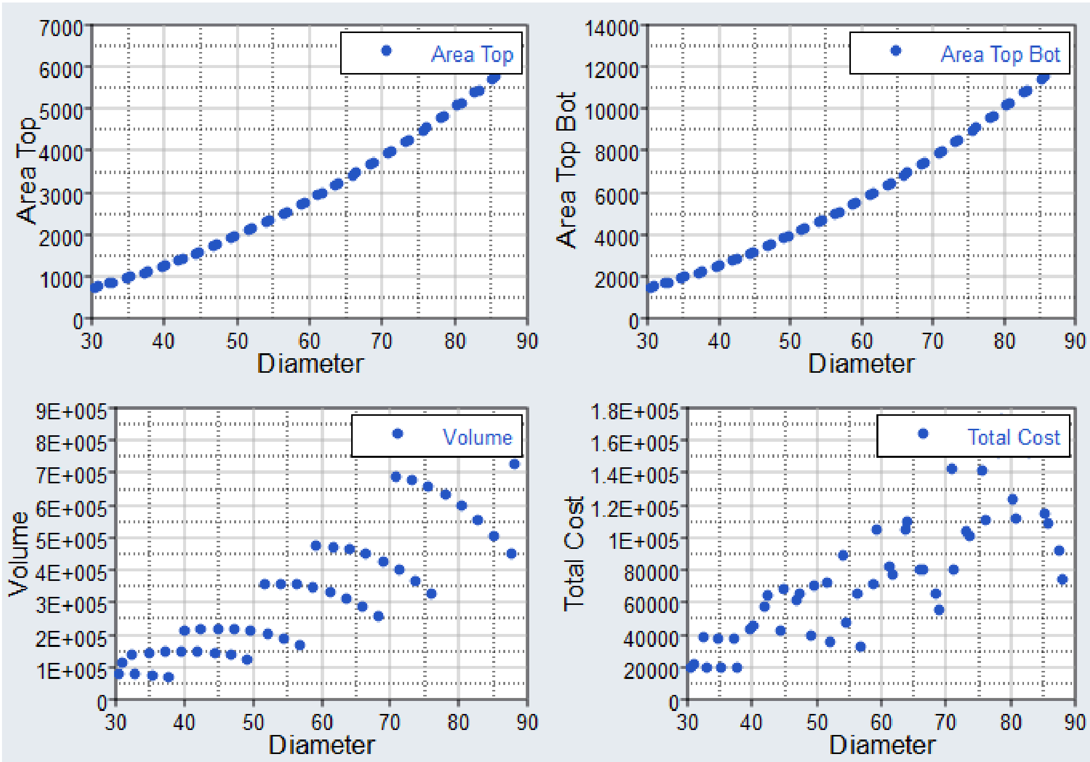


Figure 298:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

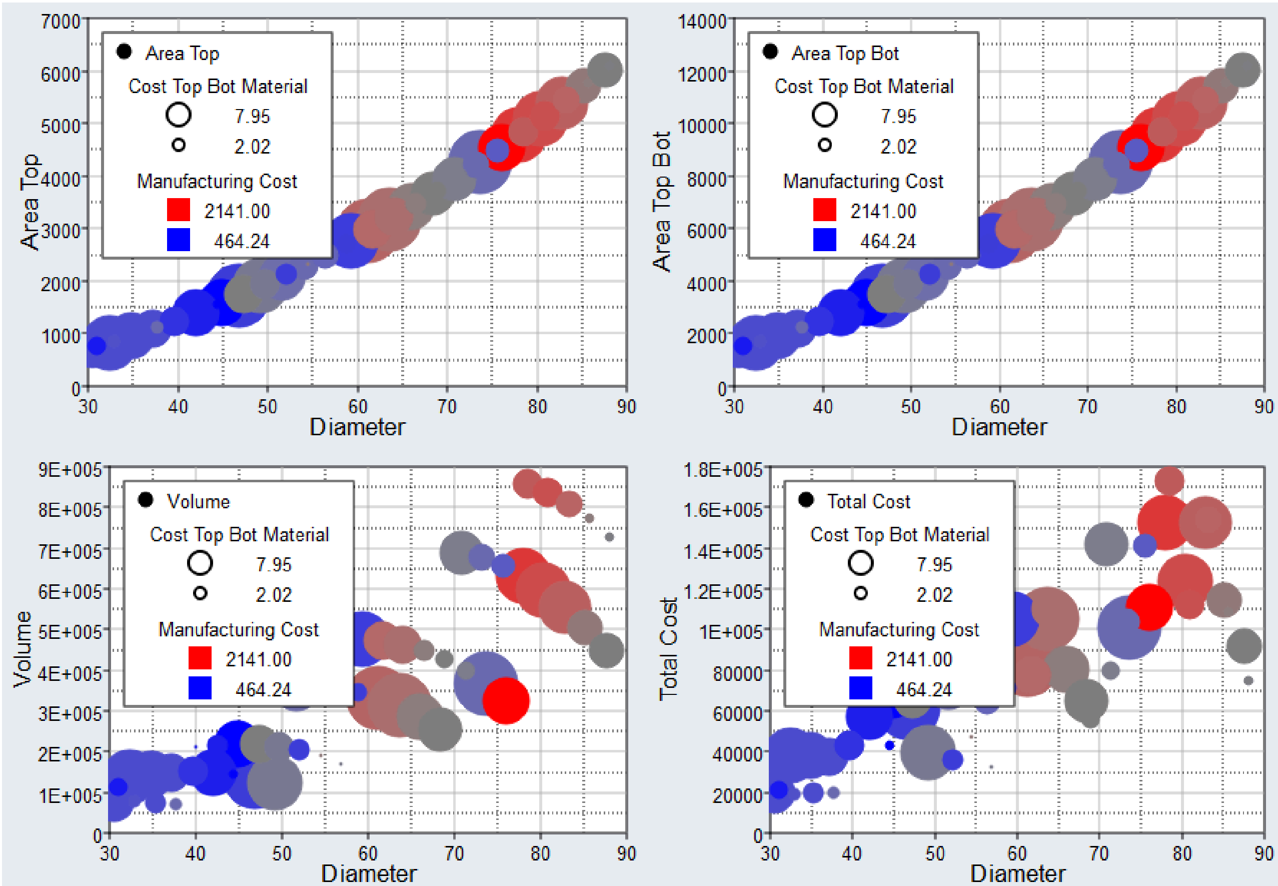


Figure 299:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Evaluation Scatter Tab Settings

Settings to configure the plots displayed in the Evaluation Scatter tab.

In the Evaluation Scatter tab, there are two methods for selecting data to display in the scatter plot: Channel and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds	Display the X bounds in the plot.

Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Review Evaluation Time

Inspect task wall-clock times.

Review the time spent in each task within the Evaluation Time tab. Identify bottlenecks in tabular or plot form.

1. From the Evaluate step, click the **Evaluation Time** tab.
2. Use the top channel selector to select the model(s) to review.
3. Use the bottom channel selector to identify the time categorises for review.

Option	Action
Write	Time spent in the write task.
Execute	Time spent in the execute task.
Extract	Time spent in the extract task.
Model Total	Total time of the write, execute, and extract tasks.
All Models Total	Summation of all Model Totals.

Option	Action
	Note: This category is independent of the selected models.

4. Switch the view between table and plot by clicking Table or Plot, located above the Channel selector.

Evaluation Time Settings

Settings to configure the plots and tables displayed in the Evaluation Time tab.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Cumulative Rows
: Each row entry is a summation of its own wall time and the wall times preceding it with a lower evaluation index.

Plot Time-Unit
: Controls the units of time when plotting the wall times.

Post Processing

View the computational results from the Verification.

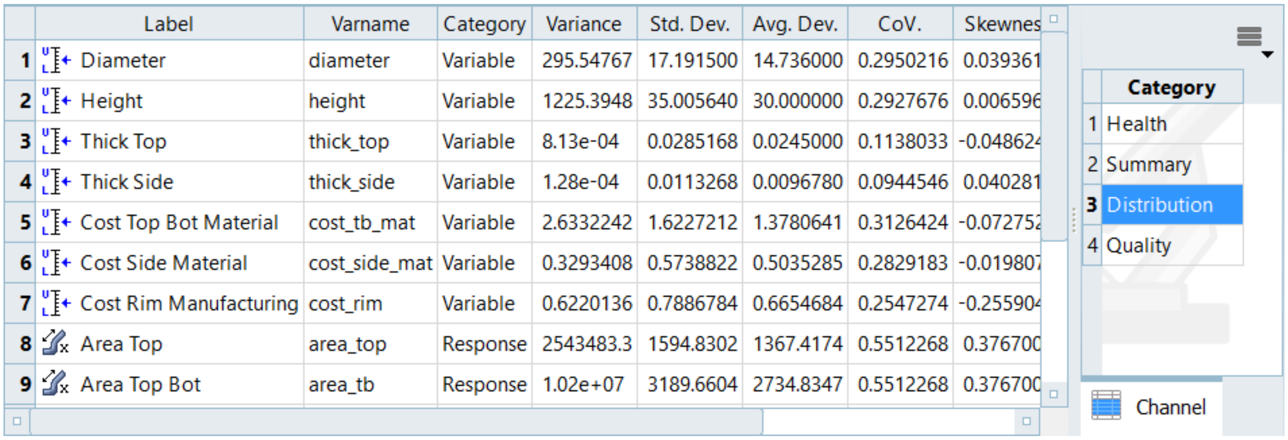
Integrity Post Processing

Check the integrity of data.

Check Integrity of Data

Review a series of statistical measures on input variables and output responses in the Integrity post processing tab.

1. From the Post Processing step, click the **Integrity** tab.
2. From the Channel selector, select a category of information to display in the table.
    - **Health** High level summary of statistics used to easily spot inconsistent, non-changing, or missing data.
    - **Summary** Basic descriptive statistics that presents information on the data in groups such as quartiles or ranges.
    - **Distribution** Detailed descriptive statistics used to quantitatively describe the distribution of data points.
    - **Quality** Values typically used in Quality Engineering.



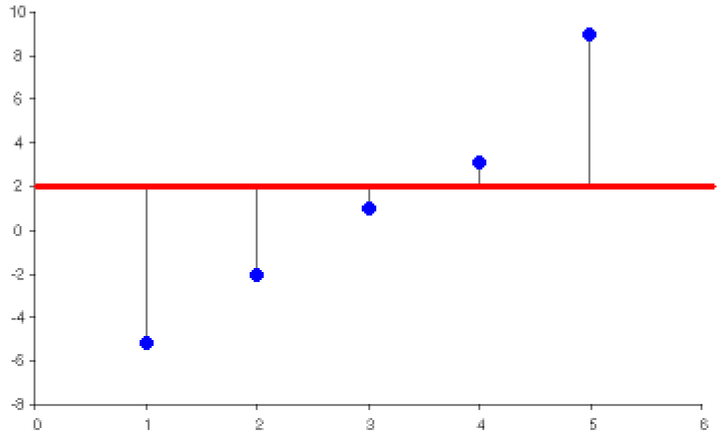
	Label	Varname	Category	Variance	Std. Dev.	Avg. Dev.	CoV.	Skewnes
1	Diameter	diameter	Variable	295.54767	17.191500	14.736000	0.2950216	0.039361
2	Height	height	Variable	1225.3948	35.005640	30.000000	0.2927676	0.006596
3	Thick Top	thick_top	Variable	8.13e-04	0.0285168	0.0245000	0.1138033	-0.048624
4	Thick Side	thick_side	Variable	1.28e-04	0.0113268	0.0096780	0.0944546	0.040281
5	Cost Top Bot Material	cost_tb_mat	Variable	2.6332242	1.6227212	1.3780641	0.3126424	-0.07275
6	Cost Side Material	cost_side_mat	Variable	0.3293408	0.5738822	0.5035285	0.2829183	-0.01980
7	Cost Rim Manufacturing	cost_rim	Variable	0.6220136	0.7886784	0.6654684	0.2547274	-0.25590
8	Area Top	area_top	Response	2543483.3	1594.8302	1367.4174	0.5512268	0.376700
9	Area Top Bot	area_tb	Response	1.02e+07	3189.6604	2734.8347	0.5512268	0.376700

Figure 300:

Integrity Tab Data

Each column in the Integrity tab displays a statistical indicator for output responses.

Column	Description
Avg Dev (Average Deviation)	Average deviation is evaluated using: $$\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$$ In Figure 301, the horizontal line represents the average of the values in the vector. The vertical lines represent the differences between the values of the vector and the average of the values. The average deviation is the average difference between the vector elements and the average of the vector elements. The sign of each element is not taken into consideration when calculating the deviation. The sign of each element is taken into consideration when calculating the average of the elements.

Column	Description
	 *Figure 301:*
CoV (Coefficient of Variation)	Measure of the relative dispersion given by: $$CoV = \frac{Standard\ Deviation}{Mean}$$ The use of variation lies partly in the fact that the mean and standard deviation tend to change together in many experiments. The higher the CoV, the higher the variability. The lower the CoV, the less the variability of the data. CoV is seldom of interest where the mean is likely to be near zero.
Kurtosis	Measure of the flatness of a distribution.
LCL (Lower Control Limit)	Mean - 3*standard_deviation
Maximum	The largest of all output response values.
Mean	The most probable value the output response would take.
Median	The median of a scalar is that value itself. The median of a vector with an odd number of elements is a scalar that is the element at the center of the ordered vector (element (N+1)/2, where N is the number of elements). The median of a vector with an even number of elements is a scalar that is the average value of the two elements closest to the center of the ordered vector (elements N/2 and (N+2)/2, where N is the number of elements).
Minimum	The smallest of all output response values.

Column	Description
Outliers	Outliers are data points that fall outside the whiskers of a box plot. To learn more about outliers, refer to About Box Plots.
RMS	The square root of the mean of the sum of the squares of all output response values is calculated using: $\sqrt{\frac{\sum x_i^2}{N}}$
Skewness	Indicates whether the probability distribution is skewed to the right or to the left. If the skewness is zero, the probability distribution is symmetric about the mean of the distribution. If the skewness is less than zero, the probability distribution is skewed to the left of the mean of the distribution. If the skewness is greater than zero, the probability distribution is skewed to the right of the mean of the distribution.
Standard Deviation	Square root of the variance. Commonly used in the measure of dispersion.
UCL (Upper Control Limit)	Mean + 3*standard_deviation
Variance	Evaluated using: $\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$

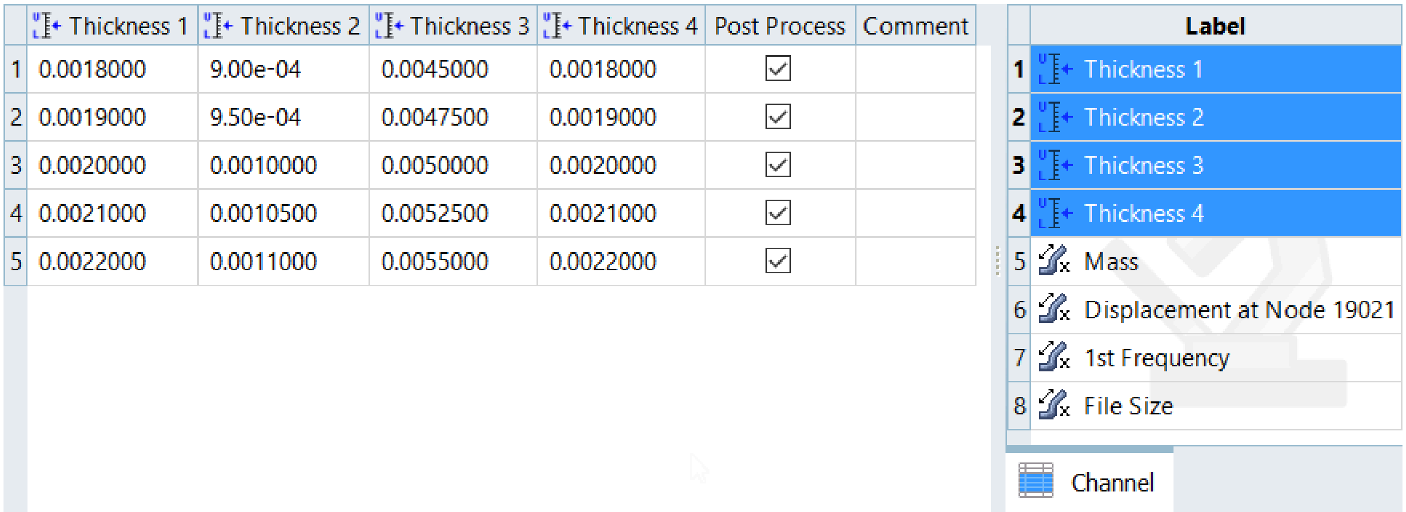
Summary Post Processing

View summary of run data.

View Run Data Summary

View a detailed summary of all input variable and output response run data in a tabular format from the Summary post processing tab.

1. From the Post Processing step, click the **Summary** tab.
2. From the Channel selector, select the channels to display in the summary table.
3. Analyze the run data summary.



	Thickness 1	Thickness 2	Thickness 3	Thickness 4	Post Process	Comment
1	0.0018000	9.00e-04	0.0045000	0.0018000	☑	
2	0.0019000	9.50e-04	0.0047500	0.0019000	☑	
3	0.0020000	0.0010000	0.0050000	0.0020000	☑	
4	0.0021000	0.0010500	0.0052500	0.0021000	☑	
5	0.0022000	0.0011000	0.0055000	0.0022000	☑	

	Label
1	Thickness 1
2	Thickness 2
3	Thickness 3
4	Thickness 4
5	Mass
6	Displacement at Node 19021
7	1st Frequency
8	File Size

Figure 302:

Parallel Coordinate Post Processing

Visualize data trends.

Visualize Data Trends

Visualize all run data across multiple channels on a single plot in the Parallel Coordinate post processing tab.

A parallel coordinate plot is also known as a snake plot.

1. From the Post Processing step, click the **Parallel Coordinates** tab.
2. From the Channel selector, select the channel(s) to plot.
   Each channel is represented by a vertical line, or axis. By default, the min and max range for each selected channel is displayed at the top and bottom of an axis.

   Run data is represented as a horizontal, colored line passing through the axes.
3. Analyze run data.

Option	Description
Display evaluation index and run data	Hover over a run line. The evaluation index and additional run data is displayed as tooltips.
Highlight run line	Left-click a run line in the plot. or Click **Show Table** (located above the Channel selector) to open the **Parallel Coordinate Table** dialog. Each run displayed in the plot is represented in a table row. Select the rows which contain the run to highlight in the plot.

Option	Description
	Note: Highlighting is disabled when a large number of runs is displayed. **Tip:** The **Show Table** option enables you to control the table channels independent of the plotted channels. This can be useful, for example, if you are plotting objective or constraint values and want to only see the variables that correspond to them.
Review trends in run data	Click-and-drag your mouse to draw boxes around sets of lines. All of the lines included in the box remain displayed, while unselected lines disappear. A visual indicator appears, and displays the minimum and maximum values for the selected set of lines. Multiple boxes can be drawn around sets of line to review. To display all of the lines, right-click in the plot and select **Reset Filter** from the context menu. In Figure 303 run data was selected for a set of lines. In Figure 304, you can see that when Styling is low, Height is high.

Option	Description

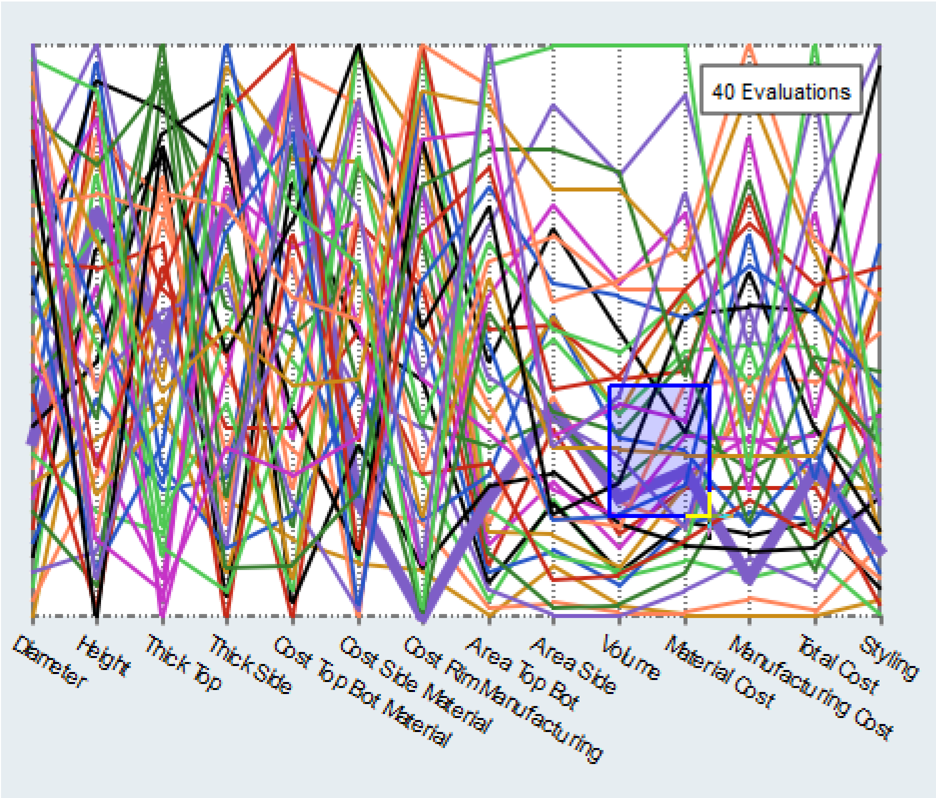
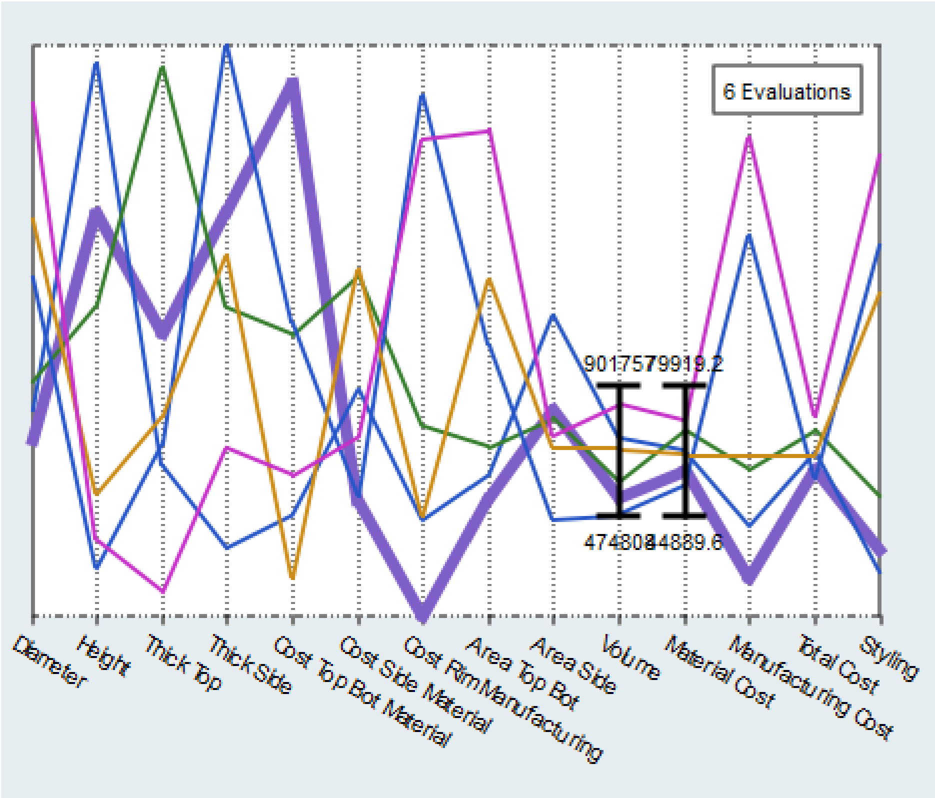


Figure 303:

Option	Description
	 *Figure 304:*
Filter run data	Click **Show Filter** (located above the Channel selector) to open the **Coordinate Filter** dialog. • From the Filter column, select the input variables and output responses to plot. • From the Filter Min and Filter Max columns, enter values to filter. The filtering mechanisms used in the Parallel Coordinate tab are interoperable, meaning the run data you have selected using box selection in the work area will be selected in the **Coordinate Filter** dialog, and visa versa.

Configure the parallel coordinate plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Parallel Coordinate Tab Settings.

Parallel Coordinate Tab Settings

Settings to configure the parallel coordinate plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Absolute Scale	Enable an absolute scale versus a relative scale which is used by default.
Show min/max	Turn the display of min and max ranges on and off.

Distribution Post Processing

Analyze distributions of run data.

Analyze Distributions of Run Data

Analyze all the distributions of run data in a histogram or box plot from the Scatter post processing tab.

1. From the Post Processing step, click the **Distribution** tab.
2. From the Channel selector, select the channels to plot.
3. Switch the view between histogram and box plot by clicking or , located above the Channel selector.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

Configure the plot's display settings by clicking ≡ (located above the Channel selector). For more information about these settings, refer to Distribution Tab Settings.

Distribution Tab Settings

Settings to configure the plots displayed in the Distribution post processing tab.

Access settings for the histogram from the menu that displays when you click ≡ (located above the Channel selector).

Histogram	Turn the display of histogram bins on and off.
Probability density (PDF)	Turn the display of PDF curves on and off.
Cumulative distribution (CDF)	Turn the display of CDF curves on and off.
Bins	Change the number of bins that displays.

About Box Plots

A box plot sorts data and draws a box from the lower quartile (1st quartile, Q1, 25%) to the upper quartile (3rd quartile, Q3, 75%).

Quartiles of a sorted data set consist of the three points (Q1, Q2 which is also the median, and Q3) that divide the data set into four groups, each group comprising a quarter of the data. The median and mean of the data are also marked in the box. In HyperStudy, this box is painted dark green.

Box plots may also have lines extending vertically from the box to indicate the data outside the lower and upper quartiles. Furthermore, to identify outliers, these lines may extend only to the "whiskers" as opposed to the minimum and maximum of the data. Whisker location is calculated as a function of lower and upper quartile and the difference between them (this difference is known as interquartile range, IQR) as:

Lower whisker	Q1 – 1.5*IQR
Upper whisker	Q3 + 1.5*IQ

Any data that is not within the whiskers are identified as "outliers." In HyperStudy, whiskers are displayed as a light green box instead of as a vertical line, and data points are indicated by blue dots. Horizontal scale is their run number and vertical scale is their value.

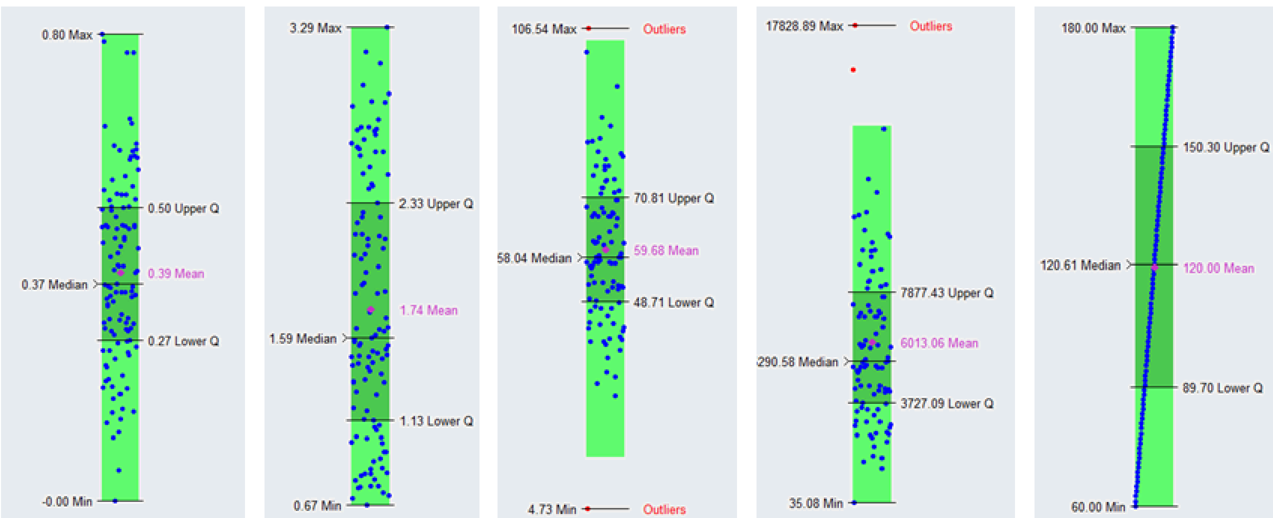


Figure 305:

Box plots display the distribution of data. Use box plots to find the range, mean, median, quartiles, whiskers and outliers. This information tells you the spread and skewness of the data and helps you identify outliers. It is important that you understand the spread and skewness in order to understand and improve the variations in the data. Identifying the outliers gives you an opportunity to investigate these data points and resolve possible issues that you may have missed.

Figure 306 is a comparison of a box plot of data sampled from a normal distribution to the theoretical probability distribution function of the normal distribution. The dark green color indicates the interquartile range, the Light green color indicates the range of the whiskers, and the red color indicates outliers.

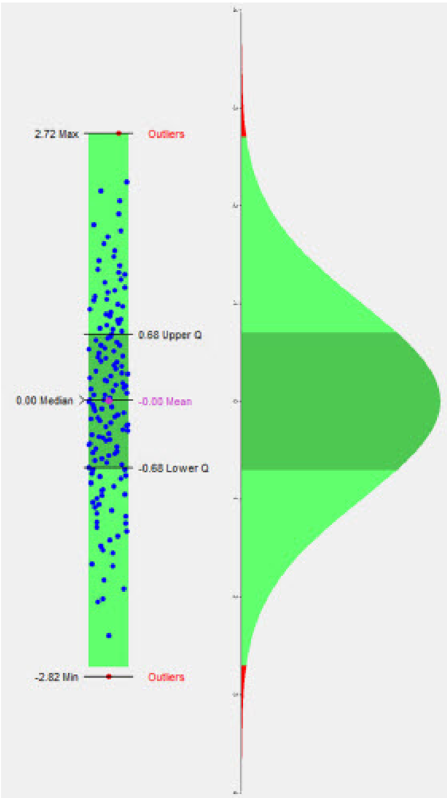


Figure 306:

About Histograms

A histogram displays the frequency of runs yielding a sub-range of output response values.

The size of the sub-range is defined as the total range of the output response value, divided by the number of bins. Histograms are displayed by blue bins.

PDF (Probability Density Function) curves illustrate the probability of the output response being equal to a particular value. PDF is displayed as a red curve.

CDF (Cumulative Density Function) curves illustrate the probability of the output response being less than or equal to a particular value. CDF is displayed as a green curve.

The accuracy of the PDF and CFD curves depend on the number of bins selected.

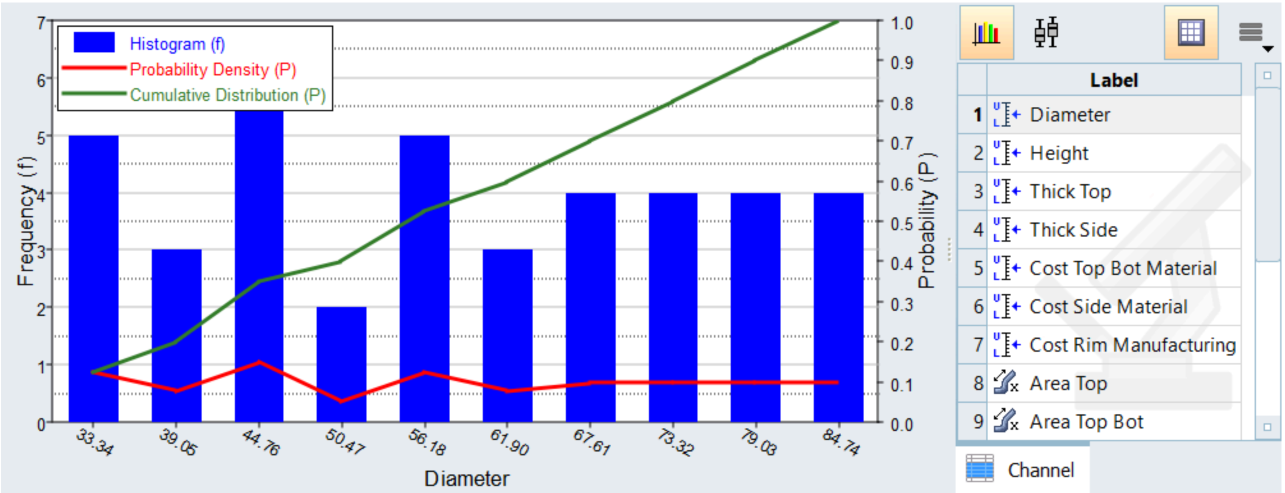


Figure 307:

Scatter Post Processing

Analyze dependency between two sets of data.

Analyze Dependency Between Two Sets of Data

Analyze the dependency between two sets of data in a scatter plot from the Scatter post processing tab. Visually emphasize data in the scatter plot by appending additional dimensions in the form of bubbles.

1. From the Post Processing step, click the **Scatter** tab.
2. Select data to display in the scatter plot.
    - Use the Channel selector to select two dimensions of data to plot.

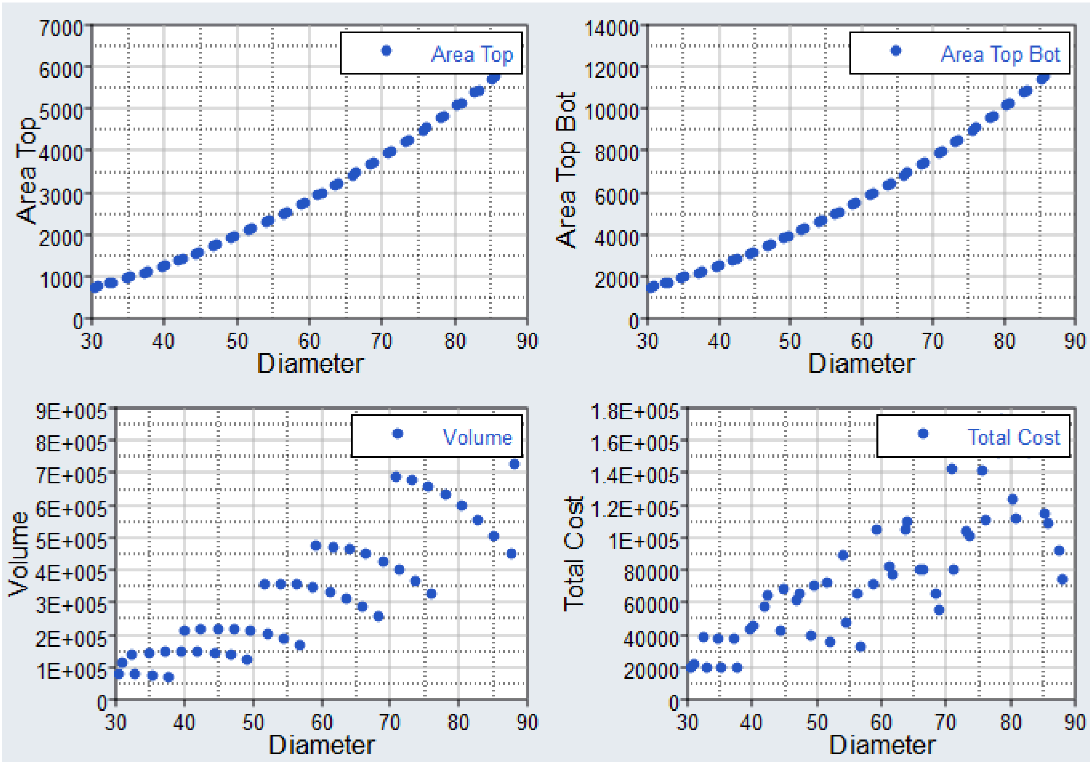


Figure 308:

- Use the Correlation selector to select one or more values from the correlation map to plot.

  Correlation measures the strength and direction between associated variables. Correlation coefficients can have a value from -1 to 1; -1 indicates a strong but negative correlation and 1 indicates a strong and positive correlation.

> **Note:** Data points are colored according to their corresponding cell in the correlation map when there are no selections active in the Bubbles selector.

	1	2	3	4	5	6	7	8	9	10
Cost Top Bot Material (5)	0.09	0.01	0.10	0.04	1.00	0.11	0.18	0.07	0.07	0.03
Cost Side Material (6)	0.22	0.09	0.05	-0.03	0.11	1.00	-0.08	0.18	0.18	0.24
Cost Rim Man...cturing (7)	-0.10	-0.18	-0.17	0.25	0.18	-0.08	1.00	-0.10	-0.10	-0.17
Area Top (8)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Top Bot (9)	0.99	0.01	0.11	-0.02	0.07	0.18	-0.10	1.00	1.00	0.71
Area Side (10)	0.71	0.68	0.06	0.13	0.03	0.24	-0.17	0.71	0.71	1.00
Volume (11)	0.86	0.45	0.09	0.13	0.02	0.22	-0.13	0.87	0.87	0.95
Material Cost (12)	0.82	0.34	0.12	0.03	0.32	0.54	-0.06	0.80	0.80	0.82
Manufacturing Cost (13)	0.72	-0.09	-0.03	0.14	0.22	0.19	0.59	0.71	0.71	0.46
Total Cost (14)	0.82	0.34	0.12	0.03	0.32	0.54	-0.05	0.80	0.80	0.82
Styling (15)	0.66	-0.70	0.13	-0.15	0.09	0.04	0.06	0.66	0.66	-0.03

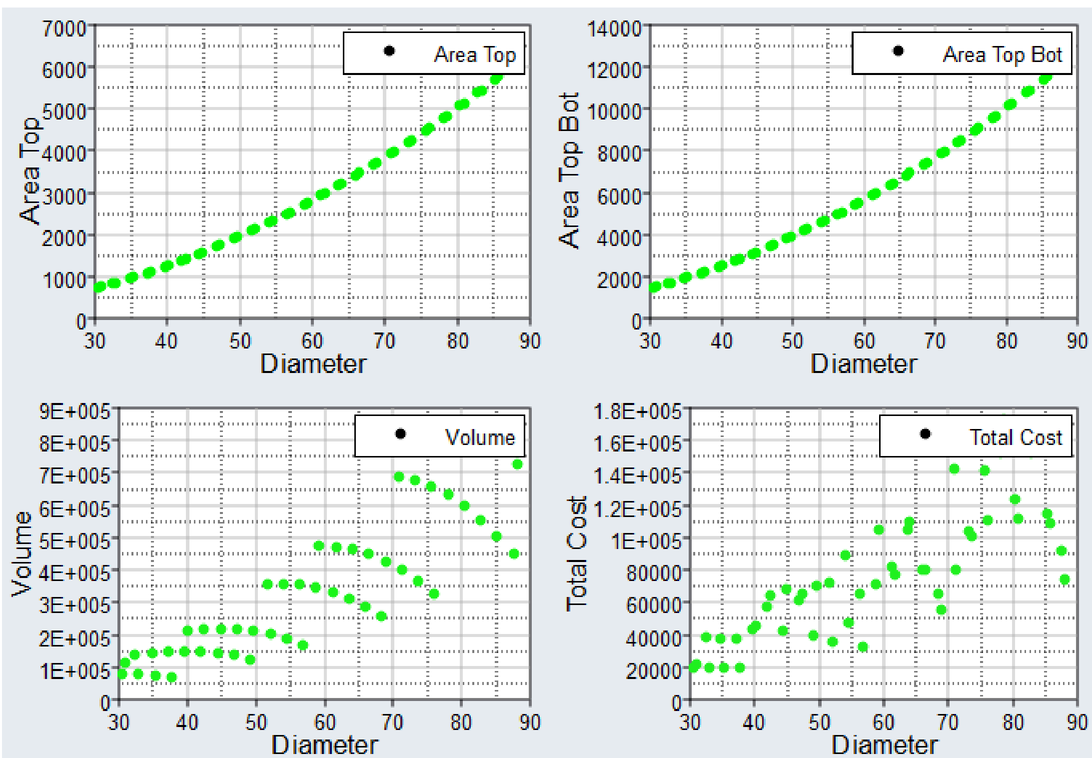


Figure 309:

- Use the Bubbles selector to select additional dimensions of data to visually emphasize in the scatter plot. The selected input variables/output responses are represented by varying sizes and colors of bubbles.

  The size and color of bubbles is determined by values in the run data for the selected input variable/output response. For size, larger bubbles equal larger values. For color, different shades of red, blue, and gray are used to visualize the range of values. The darker the

shade of red, the larger the value. The lighter the shade of blue, the smaller the value. Gray represents the median value.

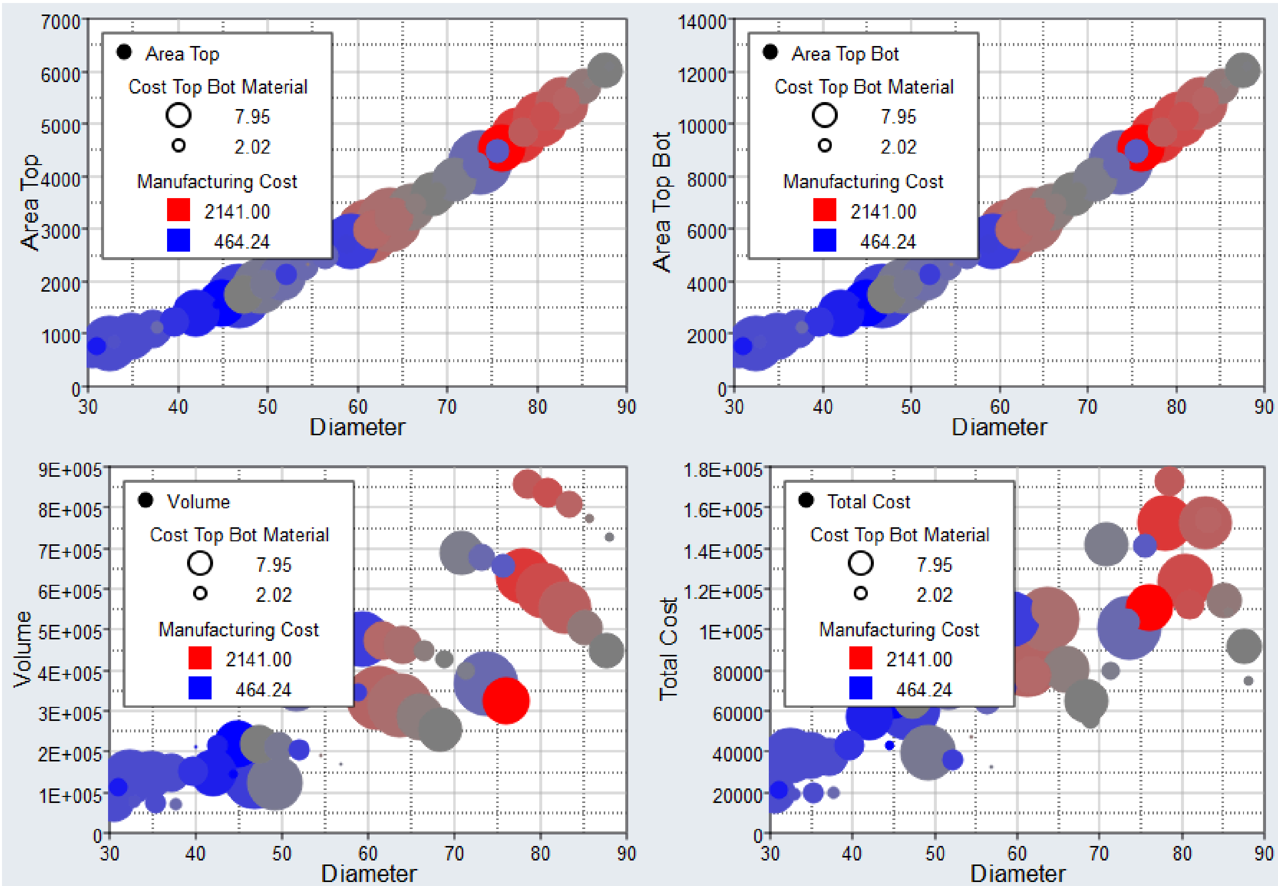


Figure 310:

3. Analyze the dependencies between the selected data sets.

> **Tip:** Display selected data in a single plot or separate plots by switching the Multiplot option between ☐ (single plot) and ▦ (multiple plots).

Configure the scatter plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Evaluation ScatterScatter Tab Settings.

Scatter Tab Settings

Settings to configure the plots displayed in the Scatter post processing tab.

In the Scatter post processing tab, there are three methods for selecting data to display in the scatter plot: Channel, Correlation, and Bubble.

Access settings from the menu that displays when you click ≡ (located in the top, right corner of the work area).

Channel Settings

X-Bounds
Display the X bounds in the plot.

Y-Bounds
Display the Y bounds in the plot.

Correlation Settings

Pearson Product-Moment / Spearman's Rank

Pearson Product-Moment (default)
Assumes a linear association, and the coefficient values indicate how far away all of the data points are from a line of best fit through the data.

Spearman's Rank
Assumes a monotonic association, and the coefficient values indicate the degree of similarity between rankings.

Pearson and Spearman's correlation coefficients are shown in the following data set:

-12.00000	1.0000000
10.000000	800.00000
40.000000	1200.0000
1000.0000	2000.0000

Figure 311: Pearson's Product-Moment Correlation Coefficient
Correlation coefficient is 0.82. There is a correlation but it is not perfectly linear.

Figure 312: Spearman's Rank Correlation Coefficient
Correlation coefficient is 1.0. It is perfectly monotonic

Correlation ≥
Show only the column/rows with cells over the specified threshold.

Show Variables and Responses
Restrict the view of the entire correlation matrix to input variables only, output responses only, input variables and output responses, or input variables versus output responses.

Include Gradients

X-Bounds	Display the X bounds in the plot.
Y-Bounds	Display the Y bounds in the plot.

Bubble Settings

Size	
Scale	Adjust the overall size of all bubbles.
Focus	Adjust the size of bubbles so that smaller bubbles become smaller, while larger bubbles remain fixed, enabling the view to be directed at larger bubbles.
Invert	Reverse the size of bubbles so that smaller values are represented by larger bubbles.
Color	
Discrete Steps	Change the level of color shading applied to bubbles.
Bins	Specify the number of red, blue, and gray shades used to color bubbles.
Invert	Reverse the color of bubbles so that red represents smaller values and blue represents larger values.

Scatter 3D Post Processing

Analyze dependency between three sets of data.

Analyze Dependency Between Three Sets of Data

Analyze the dependency between three sets of data from a scatter plot in the Scatter 3D post processing tab.

1. From the Post Processing step, click the **Scatter 3D** tab.
2. Using the Channel selector, select three dimensions of data to plot.

   > **Tip:** For the Z-Axis, multiple input variables/output responses can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between (single plot) and (multiple plots).

3. Analyze the dependencies between the selected data sets.

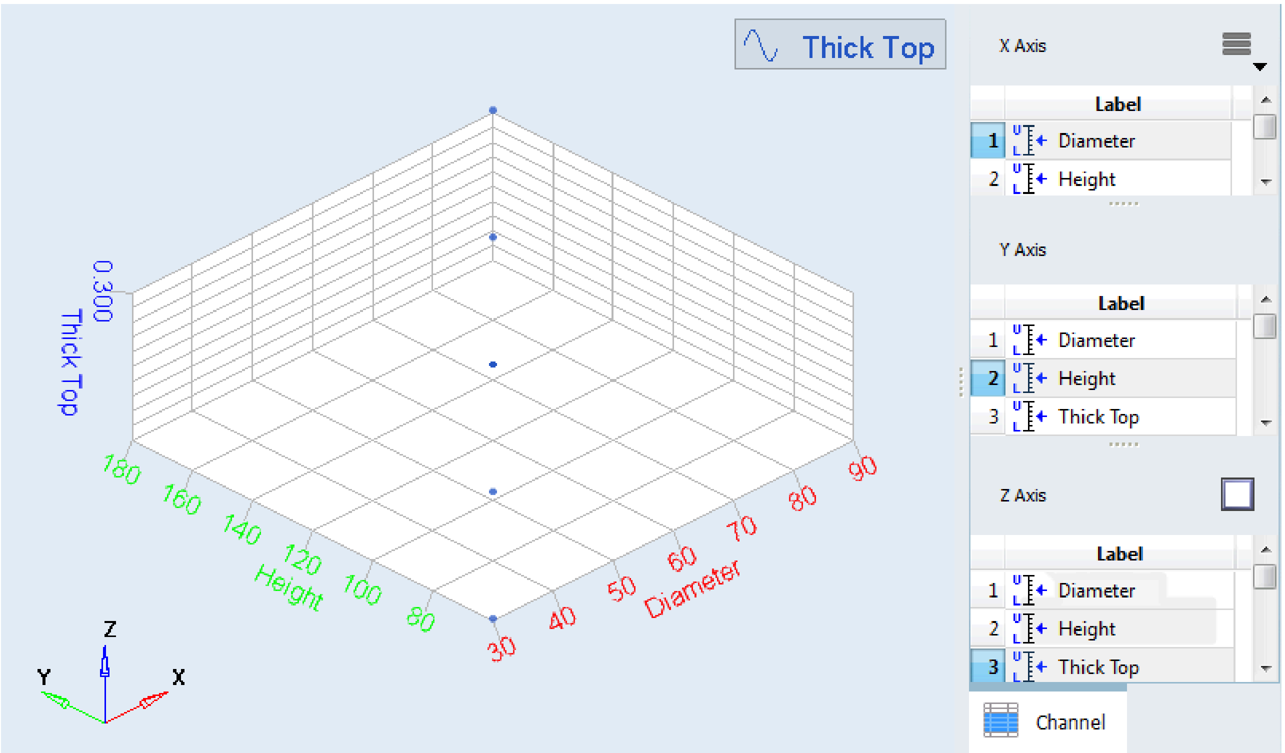


Figure 313:

Ordination Post Processing

Visualize dimension reduction.

Visualize Dimension Reduction

Analyze a biplot from a Principle Component Analysis (PCA) in the Ordination post processing tab. The PCA transforms the source data into different coordinate systems known as the principal coordinates.

Principle coordinates are ordered in terms of decreasing contribution to the data's overall variance; this means that trends in the data can typically be observed by looking at only the first few principal coordinates.

Data is represented as scatter points. Each input variable and output response in the biplot is represented by a line. The relative angle and the angle between lines can be interpreted to determine which are correlated. Lines that point in the same direction have strong correlations (positive or negative depending on whether the lines point in the same or opposite directions). The relative length of the lines also indicates a strong correlation.

1. From the Post Processing step, click the **Ordination** tab.
2. Using the Channel selector, select the principle components to plot.

> **Tip:** For the Y Principle Component, multiple components can be selected. Display selected data in a single plot or separate plots by switching the Multiplot option between ☐ (single plot) and ▦ (multiple plots).

3. Analyze the biplot.

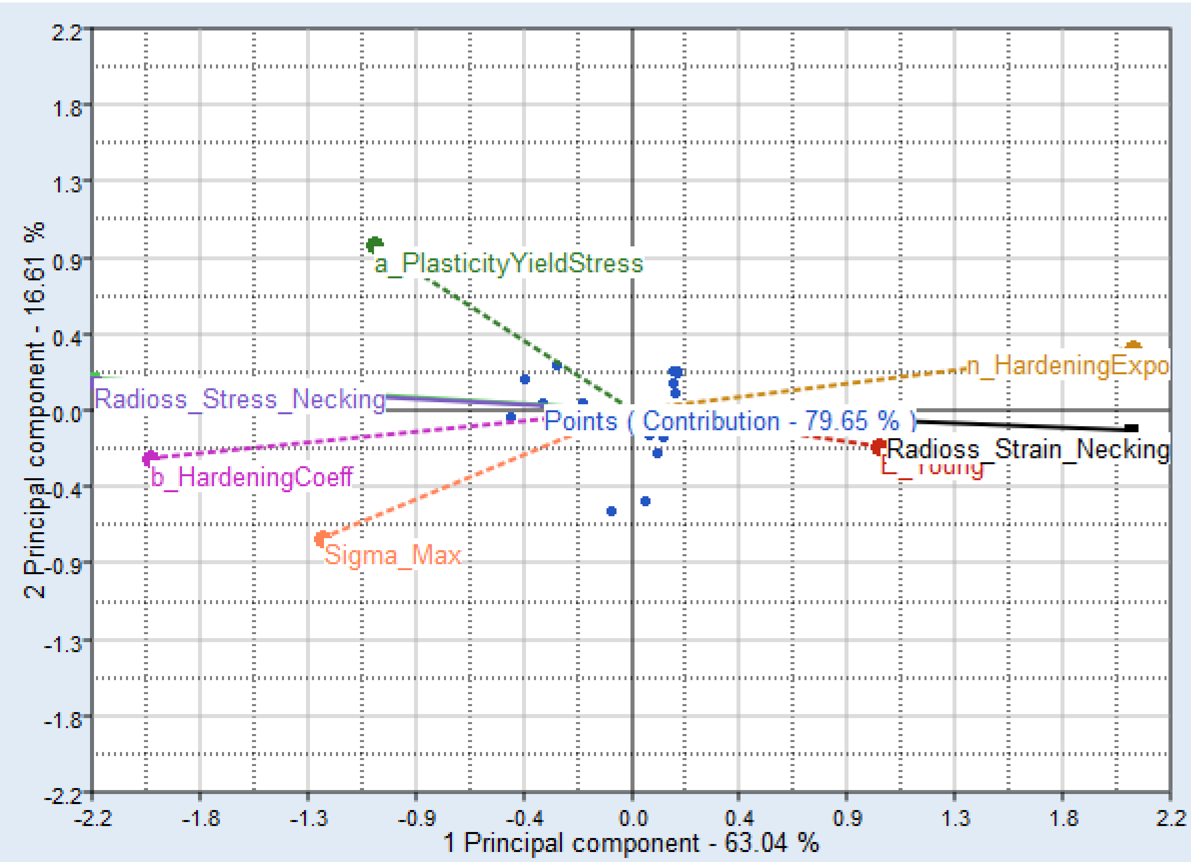


Figure 314:

Configure the plot's display settings by clicking ≡ (located in the top, right corner of the work area). For more information about these settings, refer to Ordination Tab Settings.

Ordination Tab Settings

Settings to configure the plots displayed in the Ordination post processing tab.

Access settings from the menu that displays when you click ≡ (located above the Channel selector).

Labels	Show labels in the biplot.
Points	Show scatter points in the biplot.
Legend	Show legend in the biplot.

Data Sources Post Processing

Analyze data sources.

Analyze Data Sources

Build arrays of information based on data sources using the row and column index.

1. From the Post-Processing step, click the **Data Sources** tab.
2. From the Channel selector, select a data source.
3. Select the **Table View**.
4. Build a table using the Index column, Row Index checkbox, and the Column Index checkbox.
   a) Enable the **Row Index** and **Column Index** checkboxes to display the content of the desired label in the rows or columns respectively.

   > **Tip:** To analyze the data for a specific run or array number, enable the Row Index or Column Index checkbox and enter the desired run or array number in the Index column.

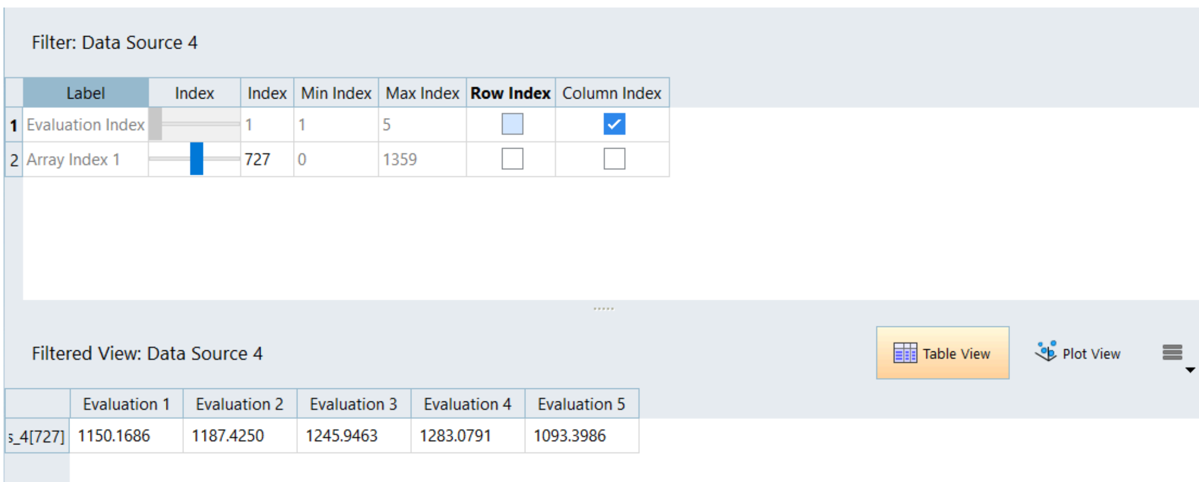


Figure 315:

5. Analyze the table.

Delta Summary Post Processing

View the difference between run and reference data.

View Delta Data Summary

View a detailed summary of the differences between input variable and output response run and reference data in a tabular format from the Delta Summary post processing tab.

1. From the Post Processing step, click the **Delta Summary** tab.

2. Using the Channel selector, select the channels to display in the delta summary table.
3. Analyze the summary.

Variable 1	Response 1	Reference Row
0.0000000	0.2291635	1
0.0000000	-0.0122360	2
0.0000000	0.0312425	3
0.0000000	0.1256404	4

Figure 316:

Delta Plot Post Processing

Compare the difference between run and reference data.

View Delta Plot

Compare differences between input variable and output response run and reference data in a bar chart format from the Delta Plot post processing tab.

1. From the Post Processing step, click the **Delta Plot** tab.
2. Using the Channel selector, select the channels to display in the plot.
3. Analyze the plot.

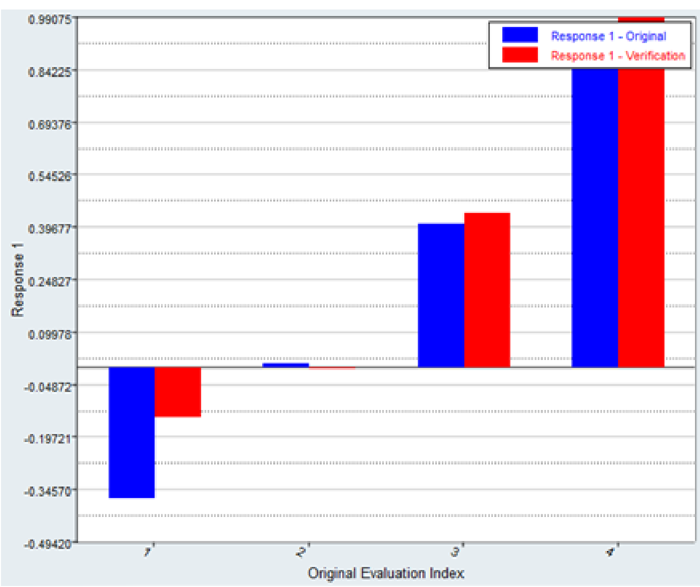


Figure 317:

Delta Scatter Post Processing

Visualize the difference between run and reference data.

View Delta Scatter

Visualize the differences between input variable and output response run and reference data in a scatter plot format from the Delta Scatter post processing tab.

This can be very useful to visualize the trade-off front after a multi-objective optimization.

1. From the Post Processing step, click the **Delta Scatter** tab.
2. Using the Channel selector, select the channels to display in the plot.
3. Analyze the scatter plot.

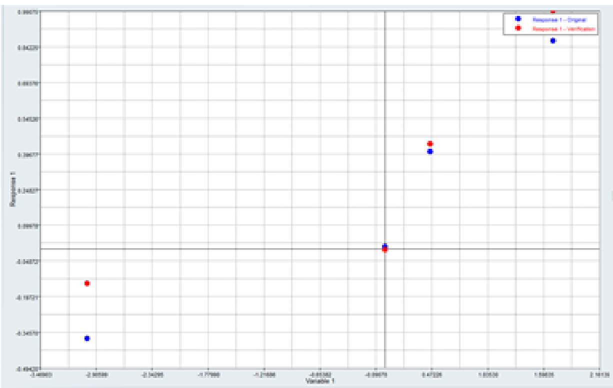


Figure 318:

Create Reports

Package reports for data generated during the approach.

1. In the study Setup, go to the Report step.
2. Select the type of report to generate.

Report Type	Description
HyperStudy Data	Generates a data report (`*.data`).
HyperStudy HTML	Generates a HTML report and opens it in your default web browser.
HyperWorks Session	Generates a HyperWorks report (`*.mvw`) and opens it in HyperWorks Desktop.
Knowledge Studio Text	Generates data compatible with the Altair Knowledge Studio text import node.
HyperStudy Fit	Generates an input file for HyperStudy Fit model (`*.pyfit`).
HyperStudy Spreadsheet	Generates a spreadsheet report and opens it in Excel.

3. Click **Create Report**.

Learn the Approaches

5

Each approach in HyperStudy serves a different purpose in the design study.

This chapter covers the following:

5.1 DOE

A DOE is a series of tests in which purposeful changes are made to the input variables to investigate their effect upon the output responses and to get an understanding of the global behavior of a design problem. By running a DOE, you can determine which factors are most influential on an output response.

5.1.1 DOE Variables

Common terminology in a DOE.

Variables

System parameters that can be changed to improve the system performance.

Reduced Variables

In DOE terminology, it is standard practice to work with reduced variables that have a range of -1 to 1 for each real variable. Reduced variables are associated to real variables with the following formula:

$$y = \frac{x - \frac{(x_{\max} + x_{\min})}{2}}{\frac{x_{\max} - x_{\min}}{2}}$$

where x is the initial variable, and y is the reduced variable.

Levels

Values taken by y in the range [-1; +1]. The number of levels per variable to be considered depends on the level of non-linearity in the problem; for a linear model two levels are sufficient; for a quadratic model three levels are needed.

5.1.2 Confounding and Resolutions

Confounding occurs when two factors are associated with each other or "travel together" and the effect of one is confused with the effect of the other. Resolution describes the degree to which estimated main effects are confounded with estimated 2-level interactions, 3-level interactions, and so on.

Confounding

Confounding occurs when two factors are associated with each other or "travel together" and the effect of one is confused with the effect of the other. For example, in order to improve team performance, a soccer coach asks his team to run two miles a day while the players decide to take vitamins. In this case the effects of running two miles a day and taking vitamins will be confounded since it will not be possible to identify the effect of them on team performance independently.

Confounding occurs whenever a Fractional Factorial design is chosen instead of a Full Factorial design. The consequence of confounding in DOEs is that the values calculated for main effects will have error coming from inclusion of higher order interactions in the calculation and interaction effects will be unknown. However Fractional Factorials also cost a fraction of the Full Factorial designs and therefore in the trade-off between cost and accuracy, they are preferred.

Table 28 shows a Full Factorial DOE for an output response that is a function of three input variable.

Table 28: Full Factorial Design Matrix

	DV 1	**DV 2**	**DV 3**	**output response of 3 input variables** **200 + 3 * dv1 - 12 * dv2 + 8 * dv3 + 2 * dv1 * dv2 - dv1 * dv3**
1	0	0	0	200
2	0	0	10	280
3	0	10	0	80
4	0	10	10	160
5	10	0	0	230
6	10	0	10	210
7	10	10	0	310
8	10	10	10	290

The effects are calculated as:

Label	Varname	Response of Three Variables
DV 1	dv_1	40.000000
DV 2	dv_2	-10.000000
DV 3	dv_3	15.000000

Figure 319:

Table 29 shows a Fractional Factorial DOE for the same output response.

Table 29: Fractional Factorial Design Matrix

	DV 1	**DV 2**	**DV 3**	**output response of 3 input variables** **200 + 3 * dv1 - 12 * dv2 + 8 * dv3 + 2 * dv1 * dv2 - dv1 * dv3**
1	0	0	10	280

	DV 1	DV 2	DV 3	output response of 3 input variables 200 + 3 * dv1 - 12 * dv2 + 8 * dv3 + 2 * dv1 * dv2 - dv1 * dv3
2	10	0	0	80
3	0	10	0	230
4	10	10	10	290

The effects are calculated as:

Label	Varname	Response of Three Variables
DV 1	dv_1	40.000000
DV 2	dv_2	-35.000000
DV 3	dv_3	65.000000

Figure 320:

The main effect values are not accurate since they include the interactions effects and consequently interactions are not captured. However, the results are still accurate enough for practical purposes (Full Factorial showed A, C and AC to be the most important contributors to the output response Y; Fractional Factorial showed A and C to be the most important contributors) and furthermore this DOE required half the runs of a Full Factorial design.

Resolution

Resolution describes the degree to which estimated main effects are confounded with estimated 2-level interactions, 3-level interactions, and so on. The design resolution tells us how badly the design is confounded. Resolution III designs confound main effects with two-factor interactions. Resolution IV designs confound main effects with three-factor interactions (A+BCD), as well as two-factor interactions with other two-factor interactions (AB+CD). Resolution V designs confound main effects with four-factor interactions, or two-factor interactions with three-factor interactions.

Higher resolution designs have less severe confounding, but require more runs. A resolution IV design is "better" than a resolution III design because we have less-severe confounding pattern in the `IV' than in the `III' situation. However in most cases higher-order interactions are less significant than low-order interactions and therefore there is not much benefit to higher resolution designs that come at the cost of additional computational expense.

5.1.3 DOE Methods

Numerical methods available for a DOE approach.

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
Box Behnken	Space Filling	3	Click **Apply** for AutoSelect or select a table using the Design pull-down menu.	Use to build quadratic response surfaces if the responses are known to be quadratic and predictions are not required at the edge of the design space. Number of points can be 13, 25, 41, 49. 57. Selecting Autoselect will pick bbdgn13 if N < 4, where N is the number of design variables; bbdgn25 if N = 4, bbdgn41 if N = 5, etc. Limited to 7 design variables. Discrete variable must have at least 3 levels. Categorical variables must have exactly 3 levels.
Central Composite Design (CCD)	Space Filling	5		Use when the responses are known to be quadratic. Limited to 20 design variables.

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
D-Optimal	Space Filling	Any	You can either accept the default number of runs or enter a different value. You can also select the appropriate regression model.	Use when the known goal is to build a regression. This method is also useful when corner coverage is important, and you have problems with input variable constraints.
Fractional Factorial	Screening	Any	Select the appropriate resolution.	Resolution indicates the level of accuracy of the interactions. Interactions should not be used with Resolution III.
Full Factorial	Screening	Any		Requires a high number of simulations and is therefore unsuitable for most studies. Total number of runs should be less than 1,000,000.
Hammersley	Space Filling	Any	You can either accept the default number of runs or enter a different value.	Use when the response surface is highly nonlinear. This method is a better space filler than Latin HyperCube. The default number of runs is 1.1*((N+1)*(N+2))/2, where N

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
				is the number of design variables.
Latin HyperCube	Space Filling	Any	You can either accept the default number of runs or enter a different value.	Use when the response surface is highly nonlinear. The default number of runs is 1.1*((N+1)*(N+2))/2, where N is the number of design variables. You must maintain the value of the random seed in order to get repeatable designs.
Modified Extensible Lattice Sequence (Mels)	Space Filling	Any	You can either accept the default number of runs or enter a different value.	Use when the response surface is highly nonlinear. This method is a better space filler than Latin HyperCube. The default number of runs is 1.1*((N+1)*(N+2))/2, where N is the number of design variables.
Plackett Burman (PB)	Screening	Any		Computationally least expensive. Number of points can be 12, 20, 24, 28 or 36. Selecting Autoselect will pick pbdgn12 if N < 12, where N is the number of design variables; pbdgn20

Method	Type	Input Variable Levels	Basic Parameters	Properties and Comments
				if 12 <= N < 20, etc. Limited to 35 design variables. Categorical variables must have exactly two levels.
Run Matrix	Custom	Any	Select the perturb file.	Use to create a design matrix using literal variable values.
Taguchi	Screening	Varies	You can either choose AutoSelect or a specific design matrix.	The levels of each variable must be set accordingly to ensure compatibility with a specific design matrix.
User Defined Design	Custom	Any	Select the perturb file.	Use to create a design matrix using abstract variable levels.

5.2 Fit (Approximation)

A Fit is a mathematical model that is trained by data and is capable of predicting output response variables for a given set of input variables.

A Fit model can then be used as an inexpensive surrogate in lieu of an actual solver in another HyperStudy approach, it can be exported for use in an external application, or it can be used in its own right to conduct what-if analyses to learn more about the system being modeled.

Some simulations are computationally expensive which makes it impractical to rely on them exclusively for design studies. In these cases, the use of Fits leads to substantial savings of computational resources. Additionally, the use of a Fit can smooth out noisy functions.

When using approximations, the issue of a tradeoff between accuracy and efficiency is ever present. The challenge is how approximate the representation of the design space can be while remaining accurate enough. The answer to this question depends on the nature of the problem as well as the resources; type of output responses, number of design parameters, and how many runs can be afforded.

5.2.1 Best Fit Selection to Prevent Overfitting

A primary desire when creating a Fit is to construct it with high predictive accuracy. HyperStudy provides several metrics which can be used to quantitatively judge the quality of a Fit. Selecting a Fit based on observing how the metrics perform on the input data is simple, but may result in overfitting the model.

> **Tip:** These metrics are presented in the Post Processing step, Diagnostic tab of the Fit.

Overfitting describes the phenomena of a Fit with very high input data diagnostics, but the Fit results in inaccurate predictions when presented with new data. Essentially, the model has been tuned to be too specific to the exact input data.

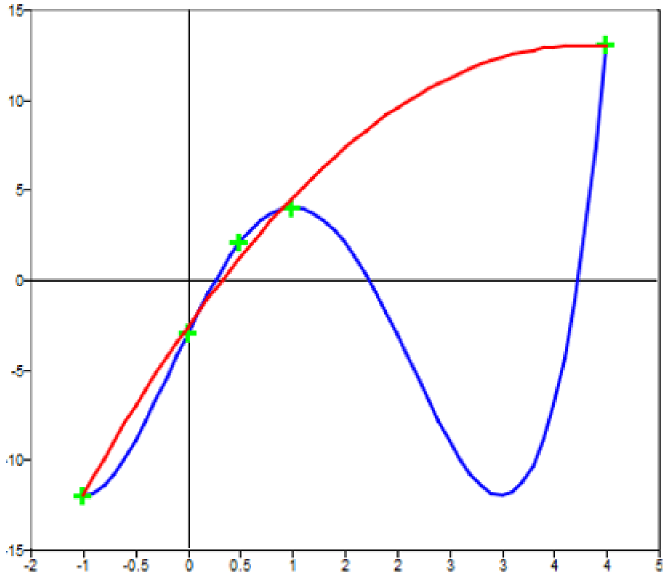


Figure 321: Difference between Two Curves Fitting the Same Data Points

The blue curve produces the exact values of the green data points, while the red curve captures the data trend without capturing small deviations in the original data. In most cases the red curve will generalize to new data better than the overfit blue curve.

To avoid overfitting, a Fit is trained with three conceptually unique sets of data. Input data is used to build a Fit, validation data is used to tune and compare different Fit options, and the testing data is used in a final step to quantify the predictive ability to unseen data.

> **Note:** Test data is never used in the construction and tuning of the Fit.

In HyperStudy, testing data is optional and the validation data is automatically constructed from the input data using a technique known as k-fold cross validation.

This technique begins with the input data and segments it into multiple folds (or groups). Imagine having 10 data points and 3 folds, the folding may look like:

Fold #	Run #
1	1,4,7,10
2	2,5,8
3	3,6,9

A fold is first withheld and a response surface is built using the remaining data. The prediction is then tested on data from the withheld fold. In this example, a Fit is first built using folds 2 and 3 and tested on fold 1. Next, it is built data from folds 1 and 3, while predicted on fold 2. This process continues for each fold. When this process is completed, the predictions on the folded data sets are compared to their known values and traditional diagnostic measures can be evaluated. Selecting a Fit based on cross-validation metrics is good practice to ensure a balance between predictive accuracy and avoiding overfitting. The size of the cross-validation folds can be set via the Cross-Validation option (accessed in the Evaluate step of the Fit); the method Fit Automatically Selected by Training calculates an internal fold size to ensure efficiency.

See Also

Diagnostics Post Processing

5.2.2 Resolve Singularities in a Fit

Singularities in a Fit matrix indicates that there is insufficient data to properly solve the posed problem. A singular matrix means that it cannot be inverted properly, which is similar to dividing by zero in scalar problems.

The most simple case of singularities in a Fit happening occurs when there are three equations but four unknowns in algebra. This system cannot properly be solved. In a Fit, this issue is more likely to occur when the runs are not sufficiently independent, which will lead to a singular matrix.

Resolve a Fit using one of the following solutions:

- Choose a different Fit method.
- Modify the Fit settings.
- Obtain additional data.

Singularities in a Fit

Consider a data set with input variables x and y, and output response z.

x,y,z

1,2,3

5,2,1

6,4,4

Run 3 is a combination of runs 1 and 2 added together. The three runs would produce two pieces of information in a linear regression, which would result in a singular matrix.

It is not possible to determine which of the input variables is responsible for a change in the output response.

The conclusion is that the data set is incompatible with the Fit specification.

5.2.3 Fit Methods

Numerical methods available for a Fit approach.

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
Fit Automatically Selected by Training	General	N/A	N/A	Choose methods for Fit Automatically Selected by	Selects the most appropriate method and settings.

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
				Training to consider.	It it recommended that you use this method unless you desire a specific method and settings.
HyperKriging	Interpolated data	###	##		The time to build the Fit and use the Fit (Evaluate From) increases with both the number of runs and the number of design variables in the input matrix. The number of design variables has more influence than the number of runs if order is larger than 1.
Least Squares Regression	Data trend lines	#	###		Noises can be screened out with this method. Closed form equations are available.
Moving Least Squares Method (MLSM)	General	##	##		The time to build the Fit and use the Fit (Evaluate From) increases with both the

Method	Response Characteristics	Accuracy	Efficiency	Basic Parameters	Comments
					number of runs and the number of design variables in the input matrix. The number of design variables has more influence than the number of runs if order is larger than 1.
Radial Basis Function	Interpolate data	###	##		The time to build the Fit increases with both the number of runs and the number of design variables in the input matrix. The number of runs has more influence than the number of design variables. The run time for using the Fit in another approach (Evaluate From) is very small regardless of the size of the input matrix.

5.3 Optimization

An Optimization is a mathematical procedure used to determine the best design for a set of given constraints, by changing the input variables in an automatic manner.

Typically, Optimization can be used to reduce product weight, improve performance, and meet design targets.

Input variables, objectives and constraints need to be defined in order to formulate a design problem as an optimization problem. There can be a single or multiple objectives. Furthermore, problems can be either deterministic or probabilistic, where the objective is to meet a certain reliability and robustness target.

5.3.1 Criteria for Formulating an Optimization Problem

In order to formulate a design problem as an optimization problem you must identify input variables, objective functions, and constraint functions.

When you put input variables, objectives and constraints together, you get the optimization formulation for a design problem as shown in Table 30.

Refer to Objectives and Constraints for more information.

Table 30:

Type	Formula(s)	Example
Objectives	$\min f(x)$	$\text{mincost}(\$)$
Constraints	$g(x) \leq 0.0$ $h(x) = 0.0$	$\sigma < \sigma allowable$
Design Space	$lower\ xi \leq xi \leq upper\ xi$ $2.5mm < thickness < 5.0mm$	$\text{number of bolts } \in (20, 22, 24, 26, 28, 30)$

Where:

f(x) is the vector of system output responses that are used as objectives.

g(x) and h(x) is the vector or system output responses that are used as inequality and equality constraints.

x is the vector of input variables.

Some of the results that would be reported after an optimization run include but not limited to:

Optimum Design	The point or design that minimized (maximized) the objective function and at the same time satisfy all the constraints.
Violated Constraint	Constraint that is not satisfied.
Active Constraint	Constraint that is satisfied exactly; equality constraints are active for feasible designs.
Inactive Constraint	Constraint that satisfied but not on the bound.
Feasible Design	A point or a design that satisfies all the constraints.
Infeasible Design	A design that violates one or more constraints.

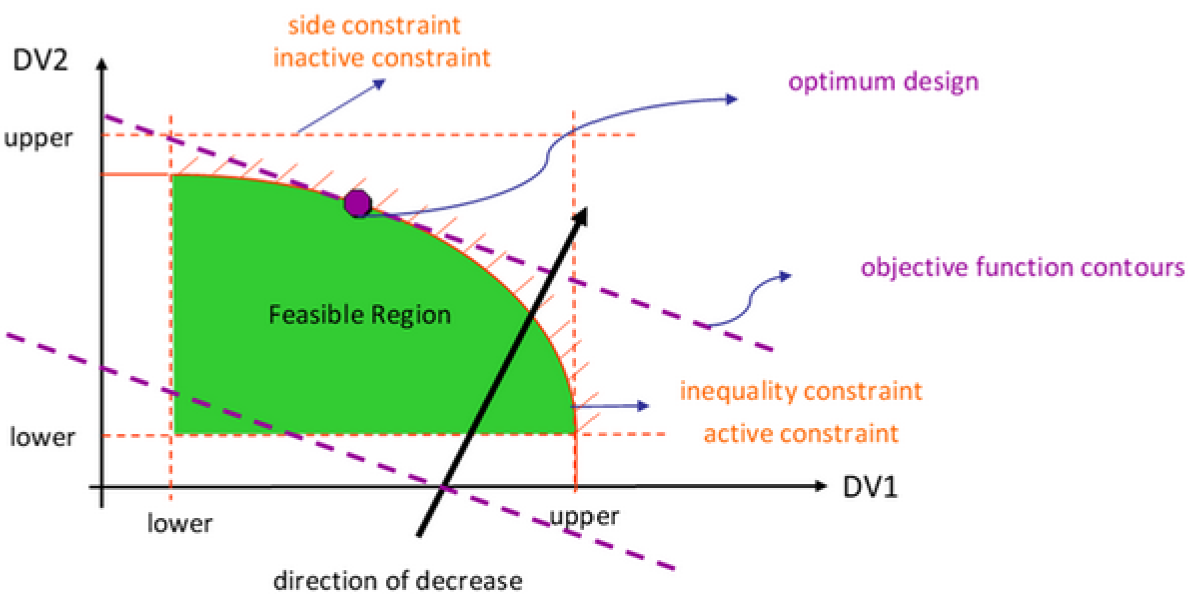


Figure 322: Design Space Definitions

See Also
Objectives
Constraints

5.3.2 Optimization Method Classification

Optimization methods can be categorized, with respect to their search technique, as iterative or exploratory. Iterative techniques can be either a local or global approximation.

Local Approximation Method (Gradient Based)

Local approximation methods are effective when the sensitivities (derivatives) of the system output responses with respect to input variables can be computed easily and inexpensively.

Local approximation methods require design sensitivity analysis (DSA) and are most suitable for linear static, dynamic and multi-body simulations.

Since finite difference calculations are expensive, DSA are preferred to be calculated directly and therefore these methods are mostly integrated with FEA Solvers. These methods are not feasible for non-linear solvers since they are locally-oriented methods.

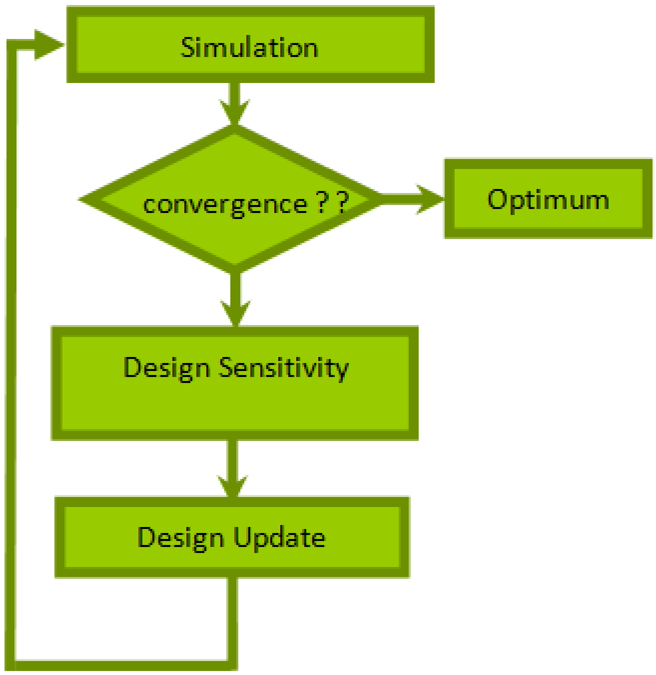


Figure 323: Gradient-Based Optimization Methods Algorithm

Global Approximation Method (Response Surface Based)

Global approximation methods are very efficient and hence they are preferred methods when dealing with noisy non-linear output responses. Global optimization methods use higher order polynomials to approximate the original structural optimization problem over a wide range of input variables.

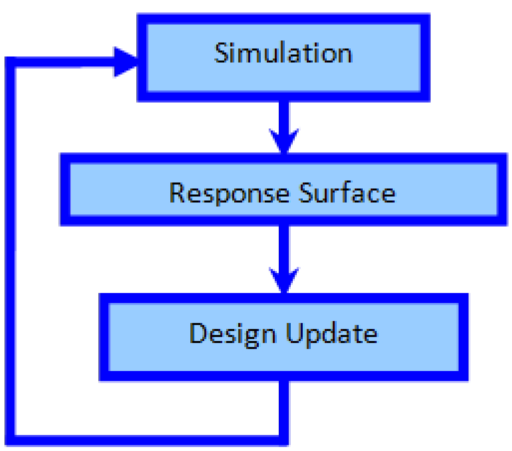


Figure 324: Approximation-Based Optimization Methods Algorithm

Exploratory Methods

Exploratory methods do not show the typical convergence of other optimization algorithms. These algorithms efficiently search the design space, however they are computationally expensive as they require large number of analysis. Rather than exhibiting conventional convergence characteristics, a maximum number of evaluations is defined.

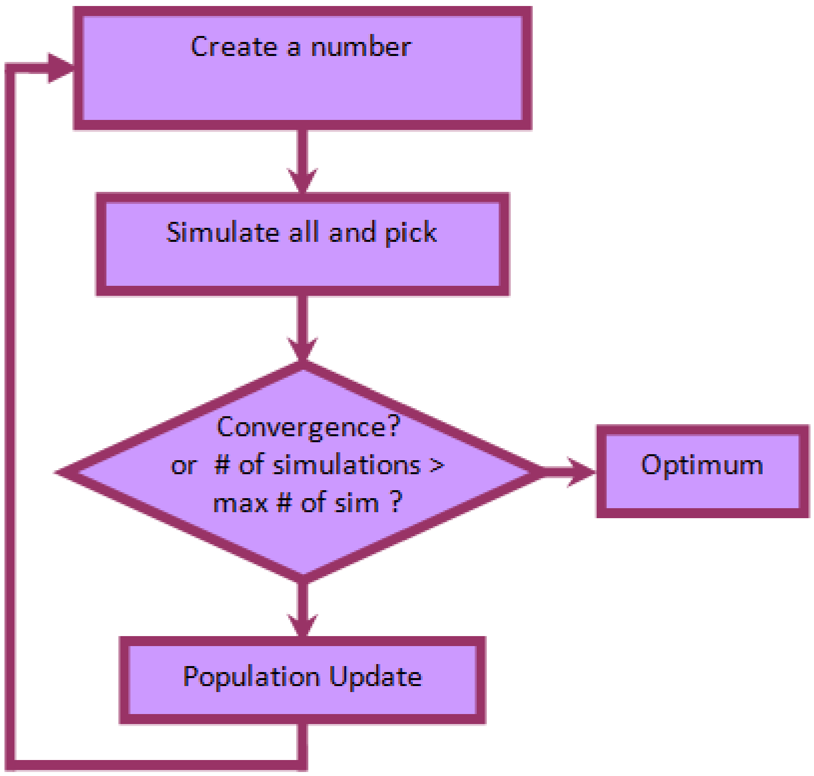


Figure 325: Exploratory Optimization Methods Algorithm

5.3.3 Optimization Methods

Numerical methods available for an Optimization approach.

Constraint violation tolerance and constraint threshold are set in the Objectives/Constraints - Goals tab of the Define Output Responses step within the Definition. For more information, visit Constraints.

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributi Role	# of Objectives	Exploratio Type	Accuracy	Efficiency	Comments
Adaptive Response Surface	None	None	Determinis	Single	Local	#	###	Default method for single

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributio Role	# of Objectives	Exploratio Type	Accuracy	Efficiency	Comments
Method (ARSM)								objective problems.
ARSM-Based Sequential Optimizatio and Reliability Assessmen (SORA_ARS	Continuous only	Input variable constraints are not allowed	Probabilisti	Single	Local	#	###	More efficient than Sequential Optimization and Reliability Assessment, but not as accurate. It is not recommended to use ARSM-Based Sequential Optimization and Reliability Assessment with a Fit.
Genetic Algorithm (GA)	None	None	Determinis	Single	Global	##	#	Significantly expensive. Use Genetic Algorithm if the simulation is affordable or if you have a good Fit.
Global Response Search	None	None	Determinis	Single or Multiple	Global	###	##	Default method for multi

Method	Input Variable Model Restriction	Input Variable Constraint Restriction	Distribution Role	# of Objectives	Exploration Type	Accuracy	Efficiency	Comments
Method (GRSM)								objective problems. Preferred method when the number of design variables is large. Optimizing can start with just a few number of points independent of the number of design variables.
Method of Feasible Directions (MFD)	Continuous only	Input variable constraints are not allowed	Deterministic	Single	Local	##	##	May work more efficiently for problems with a large number of constraints.
Multi - Objective Genetic Algorithm (MOGA)	None	None	Deterministic	Multiple	Global	##	#	Significantly more expensive. Use Multi - Objective Genetic Algorithm if the simulation

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributi Role	# of Objective:	Exploratio Type	Accuracy	Efficiency	Comments
								is affordable or if you have a good Fit.
Sequential Optimizatic and Reliability Assessmen (SORA)	Continuous only	Input variable constraints are not allowed	Probabilisti	Single	Local	###	#	Use if the simulation is affordable or if you have a good Fit.
Sequential Quadratic Programmi (SQP)	Continuous only	Input variable constraints are not allowed	Determinis	Single	Local	###	##	Use if the simulation is affordable or if you have a good Fit.
System Reliability Optimizatic (SRO)	None	None	Probabilisti	Single	Global	###	##	Default method for probabilistic problems. In robust optimization, it provides the trade-off between the nominal value and variance of the objective.

Method	Input Variable Model Restrictio	Input Variable Constraint Restrictio	Distributio Role	# of Objectives	Exploratio Type	Accuracy	Efficiency	Comments
Xopt (User-Defined Optimizatio Engine)	None	None	Determinis	Single				

5.4 Sampling Fit

A Sampling Fit is a combination of space-filling DOE method and mathematical model trained by the data generated.

To build an accurate Fit model, it is crucial to have enough DOE runs. Since it is challenging to define the right number of runs in a single DOE, the solution has been to sample with an initial guess to build a Fit model. Then, if more data is needed, perform more sampling until the expected accuracy is achieved.

Sampling Fit remedies this by continuously sampling the design space and building a Fit at specified, fixed intervals until the required cross-validation R^2 is achieved.

Modified Extensible Lattice Sequence (MELS) is used as DOE Method.

Parameter	**Default**	**Range**	**Description**
Stopping R^2	0.9	> 0 integer	Cross-validation R^2 value. Sampling stops as soon as this value is satisfied.
Maximum Evaluation Count	50	> 0 integer	Maximum number of evaluations allowed.
Evaluations Per Iteration	2	> 0 integer	Controls the fequency of building a fit model to check cross-validation R^2 value.
Sequence Offset	1	Integer 0 to 10000	Controls the starting offset for the Modified Extensible Lattice Sequence sequence. 0 Random (non-repeatable). > 0 Triggers a new sequence of pseudo-random numbers, repeatable if the same number is specified.

Parameter	Default	Range	Description
Filter Rungs with Bad Runs	On	Off or On	Filters runs with missing/invalid results per response.

5.5 Stochastic

A Stochastic approach is a method of probabilistic analysis where the input variables are defined by a probability distribution, and consequently the corresponding output responses are not a single deterministic value, but a distribution.

A Stochastic approach is used to study the influence of these uncertainties on the design. Uncertainty is inevitable in any system. Monte Carlo and Quasi-Monte Carlo methods are used to study the influence of these uncertainties on the design.

5.5.1 Uncertainty in Design

Many factors can be a source for variations in design parameters changing the nature of a design problem from deterministic to probabilistic.

These factors can be due to various types of uncertainty.

Physical uncertainty:

- Loads
- Boundary and initial conditions
- Material properties
- Geometry

Numerical simulation uncertainty:

- Conceptual modeling
- Mathematical modeling

Manufacturing:

- Sheet metal thickness
- Welds
- Random design (controlled) variables

Loads:

- Direction
- Magnitude
- Random noise (uncontrolled) variables

Material data:

- Elastic properties
- Failure
- Random noise or input variables

Uncertainties can affect input variables, also called controlled parameters, such as thickness, stiffness. They can also affect design parameters, also called uncontrolled or noise parameters, such as temperature, humidity. The resulting variations in these parameters are usually modeled by one of the many probability distribution functions based on their nature. In HyperStudy, normal, uniform, weibull, triangular and exponential distributions are available.

Corresponding to the variations in controlled or uncontrolled parameters, the design performance will also have variations. In Figure 326 probabilistic characteristics of both the parameter types and output response are shown with a typical probability density function, PDF, curve. PDF curve is a plot of variable values and corresponding probabilities. PDF describes the range of values that a probabilistic variable can attain along with the occurrence probability of each value.

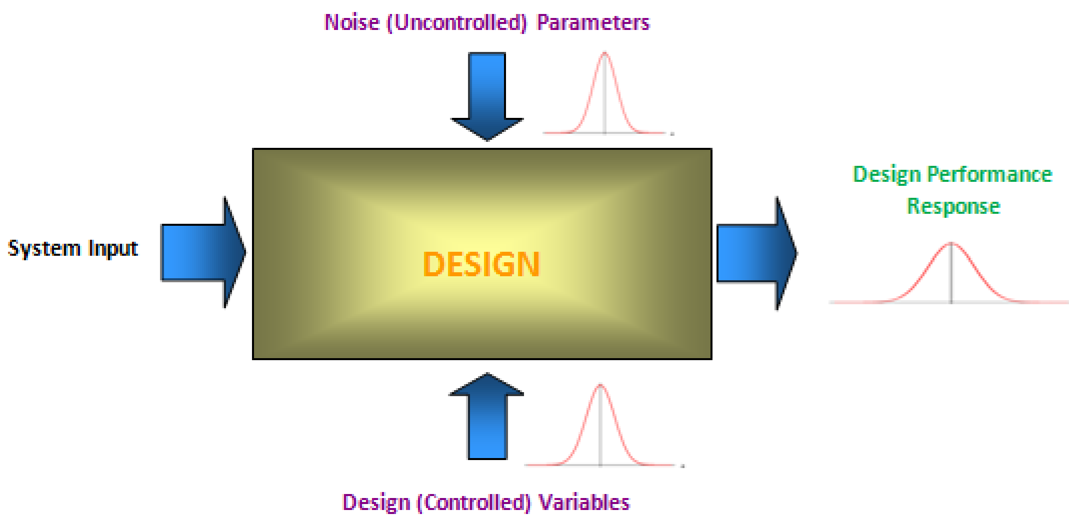


Figure 326: Uncertainty in Design

5.5.2 Reliability and Robustness

The objective in probabilistic design is to reduce the effects of probabilistic characteristics of design parameters onto design performance. Generally these effects are grouped as reliability, robustness, and reliability and robustness.

Reliability

In engineering, reliability is the ability of a system or component to perform its required functions under stated conditions for a specified period of time. It is often reported in terms of a probability. During the design process, one of the requirements can be a minimum level of reliability on a design specification such as the probability of strength values to be greater than stress values have to be greater than Po; such as, 95%; meaning that the design has to be at least 95% reliable with respect to strength requirements.

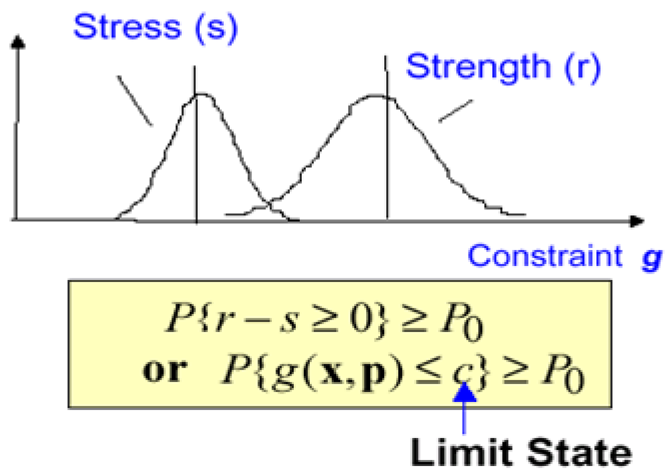


Figure 327: Example: Reliability

In Figure 328, two PDF curves are given. The PDF in solid corresponds to a design with a large area under its curve on the right curve tail violating the g constraint. In order to increase the reliability of this performance, this area needs to be reduced; meaning possible number of failures needs to be reduced. This can be achieved by shifting the mean of performance away from the constraint. The dotted PDF corresponds to such a design and it can be seen that the area under the curve in the infeasible area is much smaller than the previous one.

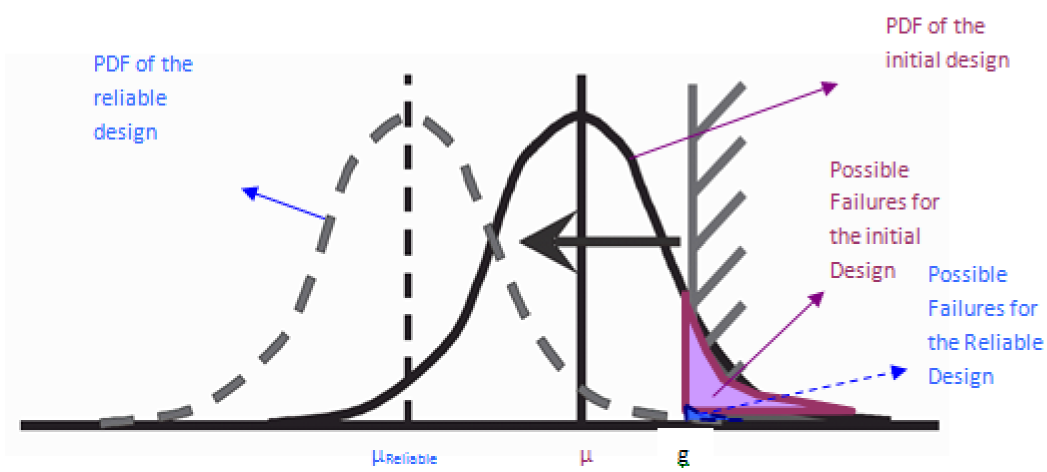


Figure 328: Improving Reliability of a Performance

Robustness

A system or design is said to be "robust" if it has minimal change of performance when subjected to variations in its design; for example, its performance is consistent within the variations.

Robustness of a product can be improved by shrinking the “variation of performance”.

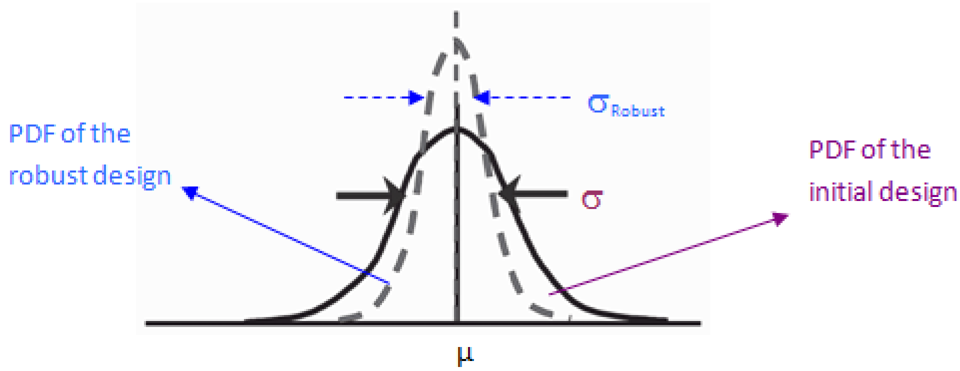


Figure 329: Improving Robustness of a Performance

Reliability and Robustness

Simultaneously shifting the mean of performance and shrinking the variation of performance, leads to both reliability and robustness improvement.

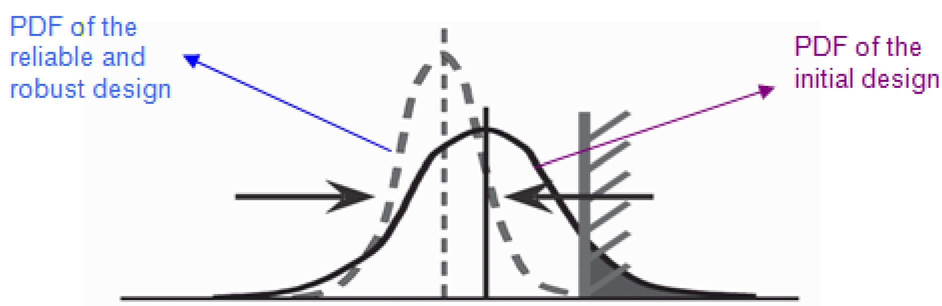


Figure 330: Improving Reliability and Robustness Simultaneously

Stochastic Assessment

Sampling-based methods generate many random samples and evaluate whether performance function is violated. They typically use random numbers; the ones that do not use random numbers are called quasi Monte Carlo methods. Sampling-based methods are also known as Monte Carlo methods.

In HyperStudy, the following sampling-based methods for reliability and robustness assessment can be used.

- Simple Random
- Latin HyperCube
- Hammersley
- Modified Extensible Lattice Sequence

Simple Random and Latin HyperCube are based on pseudo-random numbers, whereas Hammersley and Modified Extensible Lattice Sequence are based on deterministic points.

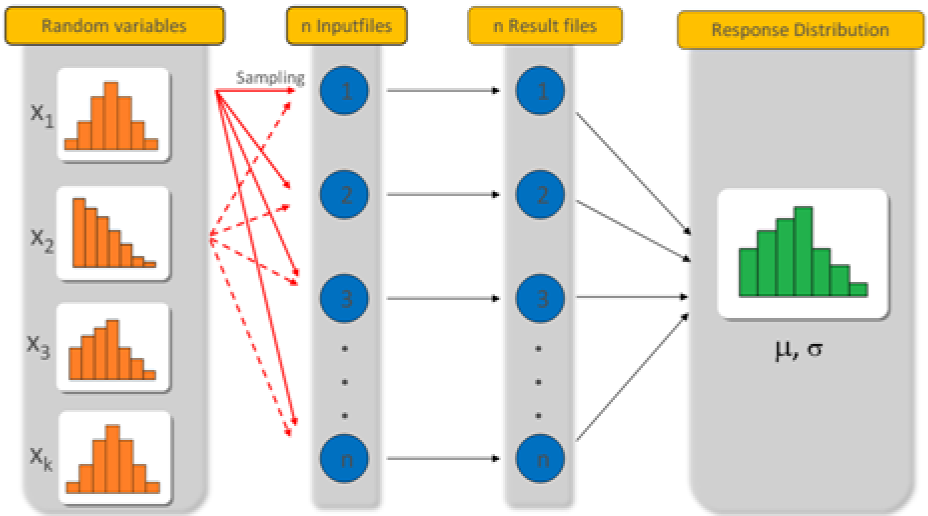


Figure 331: Position of the Sampling in the Stochastic Analysis

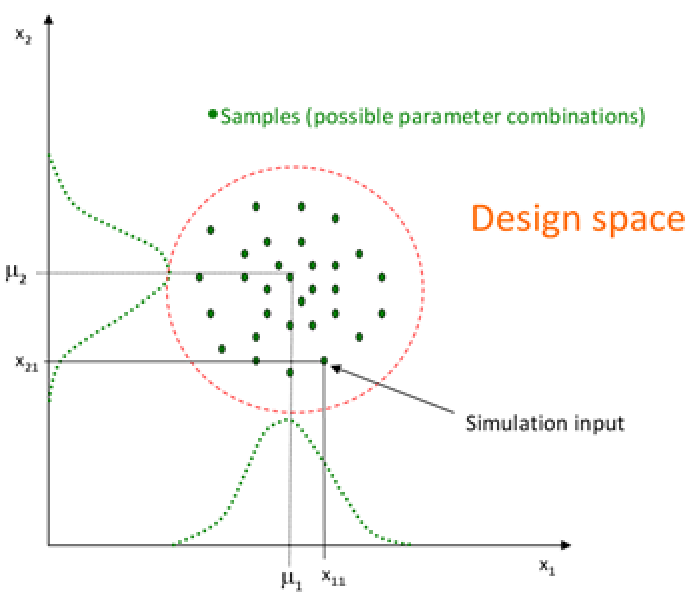


Figure 332: Illustration of the Sampling

5.5.3 Stochastic Methods

Numerical methods available for an Stochastic approach.

Method	Efficiency	Basic Parameter	Comments
Hammersley	##	Number of runs	

Method	Efficiency	Basic Parameter	Comments
Latin HyperCube	##	Number of runs	Maintain the value of the random seed to get repeatable designs.
Modified Extensible Lattice Sequence	##	Number of runs	Maintain the value of the random seed to get repeatable designs. **Note:** Modified Extensible Lattice Sequence can be extended upon itself to add points to a previously completed MELS DOE.
Simple Random	#	Number of runs	Maintain the value of the random seed to get repeatable designs.

5.6 Basic

A Basic approach can be used to test nominal values and bounds by performing a nominal run, system bound check, or sweep.

5.6.1 Basic Methods

Numerical methods available for a Basic approach.

Table 31: Basic Methods

Method	Description
Nominal Run	Runs one simulation, and sets the input variable's values to the initial values.
System Bound Check	Checks the study setup and the design space using three runs. The first run sets all of the input variables to their nominal values, the second run sets all of the values to their lower bounds, and the third run sets all of the values to their upper bounds.
Sweep	The values for a discrete input variable are iterated by index. If the number of runs exceeds the maximum index of the discrete list, the indexing scheme returns to the first index in a periodic sense.

5.7 Verification

A Verification approach compares two data sets in a side by side comparison.

This is most frequently done to investigate the differences between a Fit based prediction and a result from the approximated solver.

5.7.1 Verification Methods

Numerical methods available for a Verification approach.

Table 32: Verification Methods

Method	Basic Parameter	Comments
Verify Points	Target approach and list of evaluation indices	A general verification definition to compare a list of evaluations.
Verify Pareto	Target Optimization approach	The optimal designs from the approach are automatically selected.
Verify Trade-Off	Target Fit approach and list of input variable values	Intended to compare against input variable values coming from a Fit approach’s trade-off tab.

Customize HyperStudy 6

Customize HyperStudy by registering solver scripts, functions, and optimizers, and defining user preferences files.

This chapter covers the following:

6.1 Write and Register Solver Scripts

Learn how to write and register solver script files in HyperStudy.

6.1.1 About Solver Scripts

In HyperStudy, each model is associated with a solver execution script. The role of the solver script is to provide the name and location of the file that HyperStudy uses to execute the model.

A HyperStudy model is a construct that maps a set of independent input variables to a set of dependent output responses. The model consists of three separate steps: writing, executing, and extracting. The writing of the model files is the process of getting the independent input variable out of HyperStudy and into another format. Most of the time this corresponds to the writing of a file in some format. The opposite is true in the extraction step: data exists in some output files and must be absorbed into HyperStudy. The solver script is responsible for bridging the gap in moving the process along from the input file to the output files. So the question to ask is: "Given an input file, how do I generate the output files?".

Knowing how to generate the output files will provide the directions on how to construct the solver script. Imagine being given the input file and asked to get the results file. If the process is as simple as submitting the file as direct input to some program's executable, then the solver script could be as simple as a single line. The solver script should not finish and return control back to HyperStudy until both the process is completed and the output files are generated. If there are many steps involved, such as uploading the file to a server, submitting the file to a queue, waiting for it to finish, and then downloading the file, then the solver script must perform all these actions, too.

A solver script must be able to do everything you would do manually to create the outputs given the input file, but non-interactively.

6.1.2 Write Solver Scripts

Solver scripts can be written in any language, and the contents can be as simple as a single line or a detailed set of commands. This generality is intentional so HyperStudy remains flexible enough to be wrapped around any non-interactive process.

Suggestions for writing solver scripts:

- For each run, HyperStudy creates a separate run folder. In the case of multiple models, a separate model folder is also created. These folders are called Study Run folders.
- For each run, HyperStudy writes the solver input file to the Study Run folder.
- If any other files need to reside in the Study Run folder, they will need to be copied. For example, Radioss needs a starter file and an engine file to reside in the Study Run folder. A file can be copied from the study directory by adding its name to the solver input file. Separate names in the solver input file with a semi-colon.
- In order for HyperStudy to execute properly, verify that the solver returns control back to HyperStudy only after the execution is finished. Otherwise, HyperStudy will attempt to extract

results before all files are finished writing and the study will fail. To avoid this, the solver should be run in interactive mode if possible. Otherwise, you will need to include a wait command in your batch file. Refer to Table 33 for solver wait command examples.

- In a study that uses more than one model, the models are executed in a sequence determined by HyperStudy. To control the sequence of runs, specify the priority option for the model. The results are extracted after the solvers have finished, or earlier depending on any model dependencies.
- A failure during the script execution can be noted by creating a file titled `task__exe_err.txt`. If this file is present in the run directory, HyperStudy will detect the execution as a failure. A similar error file can be created for other task failures: write (`task__wri_err.txt`), extraction (`task__ext_err.txt`) and purge (`task__pur_err.txt`).

Table 33:

Solver	Wait Command
Abaqus	`<PATH>/abaqus.exe job=jobname.inp`
PBS	`#PBS -W block=true`
OptiStruct	`-nobg`

Access Process Environment Variables

View a list of process environment variables set by HyperStudy, which can be useful when writing solver scripts.

1. In the Message Log window, right-click and select **Verbose** > **Level 3** from the context menu.
2. Evaluate any approach.

The values of the environment variables for the current study approach will be displayed in the Message Log window.

Common CAE Solver Script Inline Commands

Solver script inline commands commonly used in HyperStudy.

> **Note:** For a complete list of available options, refer to the corresponding solver specific documentation.

Abaqus

To be used with solver input file `$filebasename`.

Windows	`<PATH>/abaqus.exe job=%1 interactive`
Linux	`<PATH>/abaqus.exe job=$1 interactive`

ANSYS

Windows	`<PATH>/ansysXXX.exe -b -i %1`
Linux	`<PATH>/ansysXXX.exe -b -i $1`

Compose/OML

Windows	`<PATH>/hwx/Compose_batch.bat -f %1`
Linux	`<PATH>/Compose_Batch -f $1`

LS-DYNA

Windows	`<PATH>/dyna.exe i=%1`
Linux	`<PATH>/dyna.exe i=$1`

OptiStruct

Windows	`<PATH>/optistruct.bat %1`
Linux	`<PATH>/optistruct.bat $1`

Example Solver Script Files Run Locally

> **Note:** The following examples are structured to become progressively more complicated. It is recommended that the examples are read in sequence and not treated independently. The examples show the use of OptiStruct, but the script concepts highlighted are general enough to be applied to many processes.

Direct Call to an Executable

Many programs already have a simple command line executable to begin a batch solution. This command generally takes some additional arguments. In general, the command line syntax to perform these operations takes the form of an executable command followed by space separated arguments.

```
[Command] [ Arguments]
```

For example, the file to call the OptiStruct solver may be located, on windows, at `C:\Program Files\Altair\14.0.120\hwsolvers\scripts\optistruct.bat`. This command can take several optional arguments, for example the name of the input file and a selection of the number of cpus. In this case, the command line syntax would be:

```
"C:\Program Files\Altair\12.0.110\hwsolvers\scripts\optistruct.bat" test.fem -ncpu 4
```

The `[Command]` is `“C:\Program Files\Altair\14.0.120\hwsolvers\scripts\optistruct.bat”`, and `[Arguments]` is `test.fem -ncpu 4`. The command is stored in HyperStudy as the solver execution script, and the solver input arguments entry would take the text `test.fem -ncpu 4`.

Another HyperStudy entry is the solver input file, which is the name of the file that will be written by HyperStudy during the model writing phase. The value of this entry is accessible from the internal variable `$file`, which in turn can be accessed in the solver input arguments, letting HyperStudy substitute this entry for you. In this case the solver input arguments would take the form `$file -ncpu 4`.

Figure 333:

Basic Solver Script with Arguments (System Native)

Rather than having to type out a set of input arguments, you can create a general purpose script that has preferred options hard coded. A general purpose script can still take a variable input file. The windows batch file (`*.bat`) that will perform this operation contains the line:

```
“C:\Program
        Files\Altair\14.0.110\hwsolvers\scripts\optistruct.bat” %1 -ncpu 4 -core
      in
```

The syntax `%1` is a substitution of the first argument to this script, which should be the input file. The argument `%2` would be second argument, and so on. This batch file can be registered as the solver script, and can have the solver input arguments as `$file`.

A similar syntax is available on linux/unix shells. The exact syntax depends on the type of shell script being written, but for Bash shells the syntax is `$n` to access the `nth` argument.

Figure 334:

Basic Solver Script with Arguments (Python)

Writing a script to run in the system native command layer can be instructive, but these languages are limited. Other languages can be used to create more detailed and featured scripts. This example uses Python, a full featured and platform neutral language. See any python reference for details on the specifics of this language.

> **Tip:** It is recommend that you use a language that you are comfortable.

The script in Basic Solver Script with Arguments (System Native) can be re-written to work with python. This script has some additional features such a functionality that corrects operating system path differences and logs the output into text files for better transparency when debugging.

```
# import statements
import os
import subprocess
import sys

f = open('logFile.txt', 'w')

#set path to file
file_exe = 'C:/Program Files/Altair/14.0/hwsolvers/scripts/optistruct.bat'
```

```
#set arguments
file_args = sys.argv[1] + ' -ncpu 4' + ' -core in'

#correct the path for the operating system, concatenate, and execute
file_exe = os.path.normpath(file_exe)
lstCommands = [file_exe, file_args]
f.write('Running the command:\n' + ' '.join(lstCommands) + '\n')
p1 = subprocess.call(' '.join(lstCommands), stdout=subprocess.PIPE ,
 stderr=subprocess.PIPE)

#write output and close
f.write('\nStandard out:' + '\n' + p1.communicate()[0] + '\n')
f.write('\nStandard error:' + '\n' + p1.communicate()[1] + '\n')
f.close()
```

Script with Simple Logic

Sometimes a script must perform several operations in sequence. This script copies a file into the current directory. This script is also designed to take a second argument, which is a file that should be removed if it exists. The same solution used in Basic Solver Script with Arguments (Python) can be executed.

Figure 335:

```
# import statements
import os
import subprocess
import sys
import shutil

f = open('logFile.txt', 'w')

#set path to files
file_exe = 'C:/Program Files/Altair/14.0/hwsolvers/scripts/optistruct.bat'
file_to_copy = 'C:/Users/jmpajot/Documents/HMath_solutions/HST/documentation_changes/
solver_script/target.txt'

#set arguments
file_args = sys.argv[1] + ' -ncpu 4' + ' -core in'

#copy the file
f.write('\nCopying the file: ' + file_to_copy + '\n')
shutil.copy(os.path.normpath(file_to_copy),os.getcwd())

#remove the file if it exists
if os.path.isfile(sys.argv[2]):
  f.write('\nRemoving the file: ' + sys.argv[2] + '\n')
  os.remove(sys.argv[2])

#correct the path for the operating system, concatenate, and execute
file_exe = os.path.normpath(file_exe)
lstCommands = [file_exe, file_args]
f.write('Running the command:\n' + ' '.join(lstCommands) + '\n')
p1 = subprocess.call(' '.join(lstCommands), stdout=subprocess.PIPE ,
 stderr=subprocess.PIPE)

#write output and close
f.write('\nStandard out:' + '\n' + p1.communicate()[0] + '\n')
f.write('\nStandard error:' + '\n' + p1.communicate()[1] + '\n')
```

```
f.close()
```

Script with Environment Variables

Dynamic scripts can be used to maintain portability by taking advantage of HyperStudy's environment variables. For example, consider the script in Script with Simple Logic. The file to be copied is defined with a fully qualified path. Instead you may require that the file be located in the HyperStudy study directory's `_usr` directory, and be copied to the current run directory. The `_usr` directory is always in the study directory, and that path is stored in the environment variable named HST_STUDY_PATH. HyperStudy will write a complete list of available environment variables to a file in the run directory if the verbose level is increased to level 3. Because these are HyperStudy environment variables, a script with these variables can only be run successfully from within HyperStudy.

```
# import statements
import os
import subprocess
import sys
import shutil
import time

f = open('logFile.txt', 'w')

#length of sleep command (seconds)
time_to_sleep = 1
time_out =  10

#set path to files
file_exe = 'C:/Program Files/Altair/14.0/hwsolvers/scripts/optistruct.bat'
file_to_copy = 'target.txt'

#set arguments
file_args = sys.argv[1] + ' -ncpu 4' + ' -core in'

#copy the file
file_to_copy = os.path.join(os.getenv('HST_STUDY_PATH'),'_usr',file_to_copy)
f.write('\nCopying the file: ' + file_to_copy + '\n')
shutil.copy(os.path.normpath(file_to_copy),os.getcwd())

#remove the file if it exists
if os.path.exists(sys.argv[2]):
  f.write('\nRemoving the file: ' + sys.argv[2] + '\n')
  os.remove(sys.argv[2])

#correct the path for the operating system, concatenate, and execute
file_exe = os.path.normpath(file_exe)
lstCommands = [file_exe, file_args]
f.write('Running the command:\n' + ' '.join(lstCommands) + '\n')
p1 = subprocess.call(' '.join(lstCommands), stdout=subprocess.PIPE ,
 stderr=subprocess.PIPE)

#write output and close
f.write('\nStandard out:' + '\n' + p1.communicate()[0] + '\n')
f.write('\nStandard error:' + '\n' + p1.communicate()[1] + '\n')
f.close()
```

Waiting for an Output File

While most commands wait until they are completed to give control back to the script, some commands return control immediately. This is common, for example, with queuing systems. In this case, additional logic is required in the script to make the script wait before moving onto the next steps. One solution to this problem is to wait for a particular file that only exists when the process in complete.

For example, HyperStudy checks for the `task__exe_err.txt` file, which indicates a failure in the execution when present.

```
# import statements
import os
import subprocess
import sys
import shutil
import time

f = open('logFile.txt', 'w')

#length of sleep command (seconds)
time_to_sleep = 1
time_out =  10

#set path to files
file_exe = 'C:/Program Files/Altair/14.0/hwsolvers/scripts/optistruct.bat'
file_to_copy = 'target.txt'

#set arguments
file_args = sys.argv[1] + ' -ncpu 4' + ' -core in'

#copy the file
file_to_copy = os.path.join(os.getenv('HST_STUDY_PATH'),'_usr',file_to_copy)
f.write('\nCopying the file: ' + file_to_copy + '\n')
shutil.copy(os.path.normpath(file_to_copy),os.getcwd())

#remove the file if it exists
if os.path.exists(sys.argv[2]):
  f.write('\nRemoving the file: ' + sys.argv[2] + '\n')
  os.remove(sys.argv[2])

#correct the path for the operating system, concatenate, and execute
file_exe = os.path.normpath(file_exe)
lstCommands = [file_exe, file_args]
f.write('Running the command:\n' + ' '.join(lstCommands) + '\n')
p1 = subprocess.Popen(' '.join(lstCommands), stdout=subprocess.PIPE ,
 stderr=subprocess.PIPE)

#wait for the file to appear
init_time = time.time()
while not os.path.exists(sys.argv[2]):
  f.write('\n ... Waiting for file to appear: ' + sys.argv[2] + '\n')
  time.sleep(time_to_sleep)
  time_delta = time.time() - init_time
  if time_delta > time_out:
     f2 = open('task__exe_err.txt', 'w')
     f2.write('Time of waiting for ' + sys.argv[2])
     f2.close()
     break

#write output and close
f.write('\nStandard out:' + '\n' + p1.communicate()[0] + '\n')
f.write('\nStandard error:' + '\n' + p1.communicate()[1] + '\n')
```

```
f.close()
```

Example Solver Scripts Run on Distributed Machines

Run Solvers Using a Queuing System on Unix

This example script runs a solver using a queuing system on Unix.

The sample file is set up for running LS-DYNA with the NQS queuing system.

The script contains a ‘wait loop’, which makes sure that the optimization process stops until the solver has completely finished. This is necessary for an Optimization study where all analyses must be performed in sequence. The wait loop can be omitted when performing a DOE study. This allows for multiple analyses to be carried out in parallel, with the network queuing system controlling the allocation of resources. The wait loop is checking the `d3hsp` file for the string `N o r m a l t e r m i n a t i o n`, which signals the end of the solution process. If another solver or queuing system is used, the script needs to be changed accordingly. For Optimization studies, the respective ASCII output file should be screened for a string signaling the proper end of the solver run.

The sample script also shows how to include certain post-processing tasks such as translating the results into a HyperMesh result file and deleting files not needed to save disc space.

```
#!/bin/sh
#
#  Set base filename for the optimization study
#
 base=filename
#
#  Set name of the que
#
 que =quename
#
#  Set environment variables for the solver
#
LSTC_FILE=/soft/usr/dyna/pass/v940_902
export LSTC_FILE
#
#  Submit solution to the que
#
(
echo "cd $PWD"
echo "/soft/usr/dyna/v940_902 i=$base.bdf x=99 memory=50000000"
) | qsub -q $que
#
#  Wait for the solver run to be finished. This can be omitted when
#  running a DOE Study. Thereby allowing multiple runs to be performed
#  in parallel, with the queuing system controlling the allocation of
#  resources.
#
MSG=""
while [ "$MSG" = "" ] ; do
 MSG=`grep "N o r m a l    t e r m i n a t i o n" d3hsp 2>/dev/null`
 sleep 30
done
#
```

```
#   Post-process data, if necessary (Create HyperMesh result file)
#
/soft/net/hmdyna d3plot $base.res
#
#   Delete data not needed
#
/bin/rm -f d3p* d3d*
#
#   End of script
#
```

Run Solvers using PBS Professional on Unix

This example script runs a solver using PBS Professional on Unix.

The sample file is set up for running LS-DYNA.

The script contains a 'wait loop,' which makes sure that the optimization process stops until the solver has completely finished. This is necessary for an Optimization study where all analyses must be performed in sequence. The wait loop can be omitted when performing a DOE study. This allows for multiple analyses to be carried out in parallel, with the network queuing system controlling the allocation of resources. The wait loop is checking for the existence of the `.pbslock` file. Its disappearance signals the end of the solution process.

If another solver or queuing system is used, the script needs to be changed accordingly.

```
#!/bin/sh
#################################################
# Script to run pbs-submit from HyperStudy
#################################################
cd $HOME/$2;
pbs-submit -c dyna -v 970_5434asingle_smp -m 300 -i $1
#################################################
# Don't return until the job completes.
# Monitor the existence of the .pbslock file
#################################################
echo "Waiting for job to complete...";
while [ -f .pbslock ]
do
continue;
done;
echo "Job completed. Returning to HyperStudy"
#################################################
```

Run Solvers on Unix, While Study is on a Unix Mapped Drive

This example demonstrates how files on Unix can be accessed from the PC by mapping a network drive to the Unix side of the network.

In order to facilitate running HyperStudy on a PC while running a solver on Unix, HyperStudy sets a process environment variable called STUDY_UNIX_PATH. The value of this process environment variable is set to the full path of the current run directory, excluding the drive letter of the mapped drive. This environment variable keeps changing with each run.

A batch file must be created to facilitate the execution of the solver on Unix from HyperStudy running on PC.

Below is a simple example of such a batch file:

```
rsh unix_mc -l user_name solver_script $HOME%STUDY_UNIX_PATH%/%1
```

- The batch file uses the `rsh` command to log into the Unix machine and execute the solver on the iterative designs created by HyperStudy.
- The variable `solver_script` is the script to run the solver on Unix.
- The batch file assumes that the mapped drive is the user's home directory on the Unix machine, therefore, the $HOME UNIX environment variable is used in the path to the input file.
-
- The variable `unix_mc` is the host name of the Unix computer where the solver is executed.
- The variable `user_name` is a Unix login name.
- In order to be able to execute the `rsh` into the Unix side, the file `.rhosts` needs to be modified by adding the host name of the PC and your login. You can do that using vi or any other editor on Unix. The `.rhosts` file must have read and write permissions for the user that is specified. If a secure shell (ssh) is installed on a user's system, the ssh can be configured with host and user keys in order to avoid password requirements. A user authentication key needs to be created on the machine running HyperStudy. This gets passed on to the Unix machine (where the solver resides) when the ssh is invoked. The solver machine needs to have a copy of the authentication key so that it can verify the user.

Run Solvers on Unix, Without a Unix Mapped Drive

This example demonstrates how to use a PC batch file and Unix shell script to run solvers on Unix, without a Unix mapped drive.

When you are running solvers on Unix, without a Unix mapped drive, the files are kept locally on the PC. Files are copied to a temporary location on the Unix machine, where the solver is executed, result files are then copied back to the study_directory on the PC where output responses are evaluated. HyperStudy sets a process environment variable called STUDY_PC_PATH, which facilitates this process. The value of this process environment variable is set to the full path of the current run directory on the PC. This environment variable changes with each run.

For this scenario, a PC batch file and a Unix shell script are required.

- When entering the values for the variables, it is important not to leave spaces after the last character or surrounding the '=' symbol.
- The file uses `rcp` and `rsh` commands.

  > **Note:** Check with your system administrator to see if these are enabled in your network.

- The batch file was created for Windows NT 4.0.
- The Unix shell script was created on IRIX 6.5.

- You want to test to see that your batch file and script are working correctly before using HyperStudy. You will need to replace `%STUDY_PC_PATH%` in the batch file with the full path for the directory containing your test file, then at the command prompt enter: `batch_filename.bat input_filename`.
- It is possible to adapt the Unix script to run with a queuing system.

Batch File

In the example batch file you should only need to edit the USER INPUT SECTION. However, depending on your system, you may need to alter other parts of the file.

Certain variables must be defined in the USER INPUT SECTION of the batch file.

username
: Username for the Unix account used.

unix_root
: Home directory on the Unix machine for the define username.

unix_tmp_dir
: Subdirectory of the user's Unix home directory to which the input file will be written.

unix_script
: Complete path, including file name, of the Unix shell script shown in Unix Shell Script, which is called upon to execute the solver on the Unix machine.

UNIX_machine
: Name of the Unix machine where the solver is executed.

```
@echo off



::::#### USER INPUT SECTION ##########################################

:: Please Input relevant information after '=' sign on each of the
:: lines in this section.
::
:: For information on what information is required on each line, go to
:: the 'Interacting with your system' page of the HyperStudy manual.


set username=user1

set unix_root=/home/user1

set unix_tmp_dir=study_dir/scratch

set unix_script=/home/user1/study_dir/scripts/study1.sh

set UNIX_machine=UNIX1


::::#### END OF USER INPUT SECTION ###################################

cls
echo.
```

```
echo.
echo ---------------------------------------------------
echo User Defined Shell Parameters on PC
echo ---------------------------------------------------
echo User name....................: %username%
echo UNIX home of user............: %unix_root%
echo UNIX Temp Dir................: %unix_root%/%unix_tmp_dir%
echo Unix run script..............: %unix_script%
echo Remote Unix Machine..........: %UNIX_machine%
echo.
echo.


set pc_dir=%STUDY_PC_PATH%

set input_deck=%1%

echo.
echo.
echo -------------------------------------------------------------
echo Copying input deck from PC to Unix ...
echo.
echo.

cd %pc_dir%

c:/winnt/system32/rcp.exe "%input_deck%" "%UNIX_machine%.%username%:%unix_tmp_dir%"


echo.
echo.
echo ...Done.
echo -------------------------------------------------------------

echo.
echo.
echo -------------------------------------------------------------
echo Launching Unix Shell...
echo.
echo.
c:/winnt/system32/rsh.exe %UNIX_machine% -l %username% "%unix_script% %unix_tmp_dir%/
%input_deck%"
echo.
echo.
echo ...Unix Shell Done.
echo -------------------------------------------------------------

echo.
echo.
echo ------------------------------------------------------------
echo Moving all files back from unix to PC...

c:/winnt/system32/rcp.exe -b "%UNIX_machine%.%username%:%unix_tmp_dir%/*" "."

echo.
echo Deleting all files from %unix_root%/%unix_tmp_dir%...
echo.
c:/winnt/system32/rsh.exe %UNIX_machine% -l %username% "rm -f %unix_tmp_dir%/*"

echo ...Done.
echo ------------------------------------------------------------
```

Unix Shell Script

In the example Unix Shell script you should only need to edit the USER INPUT SECTION. However, depending on your system, you may need to alter other parts of the file.

Certain variables must be defined in the USER INPUT SECTION of the Unix shell script.

unix_root
: User's home directory on the Unix machine.

unix_tmp_dir
: Subdirectory of the user's Unix home directory to which the input file will be written.

UNIX_machine
: Name of the Unix machine where the solver is executed.

exe_path
: Complete path, including file name, of the solver executable to be used on the input file.

```
#!/bin/sh



######## USER INPUT SECTION ###########################################

# Please Input relevant information after '=' sign on each of the
# lines in this section.
#
# For information on what information is required on each line, go to
# the 'Interacting with your system' page of the HyperStudy manual.

unix_root=/home/user1

unix_tmp_dir=study_dir/scratch

UNIX_machine=UNIX1

exe_path=/soft/solver/solver.exe


######## END OF USER INPUT SECTION ####################################

echo -----------------------------------------------
echo User Defined Shell Paramters on UNIX
echo -----------------------------------------------

echo Unix home of User...................$unix_root

echo Unix Temp Dir.......................$unix_tmp_dir

echo Remote Unix Machine.................$UNIX_machine

echo Path of SOLVER Executable...........$exe_path


input_deck=$1

echo
echo
echo
echo -------------------------------------------------
```

```
echo "Stripping cntrl M's ....."

to_unix $unix_root/$input_deck $unix_root/$input_deck.tmp
mv -f $unix_root/$input_deck.tmp $unix_root/$input_deck

echo
echo
echo ...Done.
echo -------------------------------------------------
echo
echo


echo
echo -------------------------------------------------
echo "Launching SOLVER......"
echo
echo
echo "$exe_path $unix_root/$input_deck"
$exe_path $unix_root/$input_deck

exit
```

6.1.3 Register Solver Scripts

You need to register solver scripts when the solver is not a HyperWorks solver and hence is not registered by default, or when you need to perform a series of actions such as copying files, running one or more solvers, and extracting data.

Register solver scripts in HyperStudy or by editing the preferences file.

Register Solver Scripts in HyperStudy

Register solver scripts in HyperStudy with the **Register Solver Scripts** dialog.

1. Open the **Register Solver Script** dialog.
   - From the menu bar, click **Edit** > **Register Solver Script**.
   - In the Define Models step, click the **Solver Execution Script** cell and select **Register new Solver** from the drop-down menu.
2. Add a solver script.
   a) Click **Add Solver Script**.
   b) In the **Add - HyperStudy** dialog, select the type of script to create and click **OK**.
   HyperStudy diagnostic messages are richer for HyperWorks solvers. Select the solver specific script types to use a different version of Feko, Flux, MotionSolve, OptiStruct, Radioss, or Workbench other than the default scripts provided and still be able to receive the same rich, diagnostic information.
3. In the Path column, enter the location of the solver script to be registered.
4. Optional: In the Arguments column, enter a solver script argument for the solver script.
5. Click **Close** to exit the solver registration.

Register Solver Scripts by Editing the Preferences File

Register solver scripts in the `*BeginSolverDefaults` section of the `preferences_study.mvw` file.

1. In a text editor, open the `preferences_study.mvw` file.

   You can find the `preferences_study.mvw` file in the HyperWorks installation directory under `<install directory>/hw/prefinc/`.
2. Search for the syntax `*BeginSolverDefaults`.

   Below this syntax there is an 'if-else-loop.' The first part of the loop registers the three solvers listed above when the operating system is Windows. The second part (else) registers these solvers when the operating system is not Windows.
3. If you are working with a Windows operating system, append the first list of `*RegisterSolverScript` statements. Otherwise, you will need to append the second list of `*RegisterSolverScript` statements.
4. Immediately following the existing `*RegisterSolverScript` statements, register further solver scripts using the following syntax:

```
*RegisterSolverScript(script_name,"script_label","executable","solver_type ",
 "arguments")
```

   where:

script_name	Unique name for the script.
"script_label"	Name used within HyperStudy to reference the script. **Note:** This name must be enclosed in double quotes.
"executable"	Full path of the solver script, including the file name and extension.
"solver_type"	Indicates to HyperStudy which solver is used.
"arguments"	Solver input arguments.

5. Save the `preferences_study.mvw` file.

Example: Register LS-DYNA in the Preferences File

In this example, a LS-DYNA executable is added to the solver defaults section of the `preferences_study.mvw` file for a Windows operating system. The added line is in bold.

```
*BeginSolverDefaults()
*RegisterSolverScript(radioss,"RADIOSS",{ getenv("radioss_launch") },
 HST_SolverRadioss)
*RegisterSolverScript(os,"OptiStruct",{ getenv("opti_launch") },
 HST_SolverOptiStruct)
*RegisterSolverScript(templex,"Templex",{getenv("templex_launch") },
 HST_SolverGeneric)
*RegisterSolverScript(hx,"HyperXtrude",{getenv("hx_launch") },
 HST_SolverGeneric)
```

```
*RegisterSolverScript(ms,"MotionSolve - standalone", { getenv("ms_launch") },
 HST_SolverMotionSolve)
*RegisterSolverScript(tcl,"TCL",{ getenv("tclsh_fullpath") },
 HST_SolverGeneric)
*RegisterSolverScript(lsdyna, "Ls-Dyna", "C:\Solvers\dyna\dyna.exe",
 "HST_SolverGeneric")
*EndSolverDefaults()
```

6.2 Register User Utilities

Run external programs from within the HyperStudy interface, such as within the Directory.

Register user utilities in HyperStudy or by editing the preferences file.

Register User Utilities in HyperStudy

1. From the menu bar, click **Edit** > **Register User Utility**.
   The **Register User Utility** dialog opens.
2. Add a user utility.
   a) Click **Add User Utility**.
   b) In the **Add** dialog, enter a label for the user utility and click **OK**.
3. In the Path column, enter the location of the user utility to be registered.
4. Optional: In the Arguments column, enter any arguments for the user utility.
5. Click **Close** to exit the user utility registration.

Register User Utilities by Editing the Preferences File

Register user utilities in the `*BeginUserUtilityDefaults` section of the `preferences_study.mvw` file.

1. In a text editor, open the `preferences_study.mvw` file.

   You can find the `preferences_study.mvw` file in the HyperWorks installation directory under `<install directory>/hw/prefinc/`.
2. Search for the syntax `*BeginUserUtilityDefaults`.
3. If you are working with a Windows operating system, append the first list of `*RegisterUserUtility` statements. Otherwise, you will need to append the second list of `*RegisterUserUtility` statements.
4. Immediately following the existing `*RegisterUserUtility` statements, register further user utilities using the following syntax:

```
*RegisterUserUtility (script_name,"script_label","utility interpreter", "utility
 path")
```

   where:

script_name	Unique name for the utility.
script_label	Name used within HyperStudy to reference the utility. **Note:** This name must be enclosed in double quotes.

utility interpreter	Full path of the interpreter of the utility, including file name and extension. For example, `wish.exe` for a Tcl script.
utility path	Full path of the utility, including the file name and extension.

5. Save the `preferences_study.mvw` file.

Example: Register Tcl Script in the Preferences File

In this example, a Tcl script is added to the user utility defaults section of the `preferences_study.mvw` file for a Windows operating system.

```
*BeginUserUtilityDefaults()
*RegisterUserUtilityScript(u_1,"U1","C:\wish.exe","C:\param.tcl")
*EndUserUtilityDefaults()
```

6.3 Register External Optimizers

In order to use HyperStudy with other external optimizers, they must be registered.

Register external optimizers in HyperStudy or by editing the preferences file.

Register External Optimizers in HyperStudy

1. From the menu bar, click **Edit** > **Register Optimizer**.
   The **Register External Optimizer** dialog opens.
2. Add an external optimizer.
   a) Click **Add External Optimizer**.
   b) In the **Add** dialog, enter a label for the optimizer and click **OK**.
3. In the Path column, enter the location of the external optimizer to be registered.
4. Click **Close** to exit the external optimizer registration.

Register External Optimizers by Editing the Preferences File

Register external optimizers in the `*BeginExternalOptimizerDefaults()` section of the `preferences_study.mvw` file.

1. In a text editor, open the `preferences_study.mvw` file.

   > **Tip:** You can find the `preferences_study.mvw` file in the HyperWorks installation directory under `<install directory>/hw/prefinc/`.

2. Search for the syntax `*BeginExternalOptimizerDefaults()`.
3. If you are working with a Windows operating system, append the first list of `*RegisterExternalOptimizer` statements. Otherwise, you will need to append the second list of `*RegisterExternalOptimizer` statements.
4. Immediately following the existing `*RegisterExternalOptimizer` statements, register further external optimizers using the following syntax:

```
*RegisterExternalOptimizer (opti_name,"opti_label","executable")
```

   where:

opti_name	Unique name for the optimizer.
opti_label	Name used within HyperStudy to reference the optimizer.

> **Note:** This name must be enclosed in double quotes.

executable Full path of the external optimizer, including file name and extension.

5. Save the `preferences_study.mvw` file.

Example: Register `sqp.exe` in the Preferences File

In this example, the `sqp.exe` executable is added to the default list of registered external optimizer for a Windows operating system. The added line is in bold.

```
*BeginExternalOptimizerDefaults()

       {machtype = sysid()}

       {if (machtype == "windows")}

              *RegisterExternalOptimizer(xopt, "Xopt",
 { getenv("HST_ALTAIR_HOME") + "/hw/bin/WIN64/xopt.exe" })
*RegisterExternalOptimizer(sqp, "SQP", {"C:/My_SQPOptimizer/sqp.exe" })

         {else}

              *RegisterExternalOptimizer(xopt, "Xopt",
 { getenv("HST_ALTAIR_HOME") + "/hw/bin/" + getenv("ALTAIR_PROD_ARCH") + "/
xopt" })

         {endif}

   *EndExternalOptimizerDefaults()
```

6.4 Set User Preferences File

A preferences file is an ASCII file that configures the application or its clients, and specifies default user settings such as the readers, functions, and solver script locations.

A standard preferences file is created in the program installation directory and is executed every time a license is activated. An additional preferences file can be created in your working directory, and can contain personal settings that will either overwrite the standard preferences file or be added to the existing settings in the standard preferences file.

By default, HyperStudy uses the default preferences file (`$HST_ALTAIR_HOME/hw/preferences_hst.mvw`).

> **Note:** To use the custom preferences file in batch mode, you need to use the option `–preffile` “preferences file `name.mvw`”. For more information, refer to HyperStudy Start Options.

1. From the menu bar, select **File** > **Use Preferences File**.
2. In the **Set Preferences File** dialog, select your custom preferences file and click **Open**.

HyperStudy reads the default preferences file in the installation directory, followed by the preferences file that you specify. This ensures that all readers and import templates are available.

After you have set a user preferences file, you can select a new preferences file or disable the custom preferences file.

- Select a new custom preferences file by selecting **File** > **Set Preferences File** from the menu bar.
- Disable the custom preferences file by selecting **File** > **Use Preferences File** from the menu bar. The Use Preferences File checkbox should be off.

6.4.1 Preference File Reading Sequence

The locations and order in which HyperStudy will load preference files.

1. Read main installation preference
   - `getenv("ALTAIR_HOME") + "/hw/preferences_hst.mvw`
2. Read user preference files
   - `getenv("USER_PREFERENCES_HST"` (points directly to a file)
   - `getenv("USER_PREFERENCES_HST_DIR")/*.mvw` (all `*.mvw` files in directory)
   - Linux: `(Linux) getenv("HOME")/.preferences.mvw`
   - Linux: `<current working dir>/.preferences.mvw`
   - `preferences.mvw` (single file)

     `getenv("HW_SETTINGS_DIR")/HyperGraph/preferences.mvw`

     Windows: `C:/Users/<user>/Documents/HyperGraph/preferences.mvw`

     Linux: `$HOME/.altair/HyperGraph/preferences.mvw`
   - `preferences/*.mvw` (directory)

`getenv("HW_SETTINGS_DIR")/HyperStudy/preferences/*.mvw`

Windows: `C:/Users/<user>/Documents/HyperStudy/preferences/*.mvw`

Linux: `$HOME/.altair/HyperStudy/preferences/*.mvw`

- `getenv(HW_CONFIG_PATH)`, a list of directories separated by semicolons

  In every directory, HyperStudy tries to read `<dir>/preferences.mvw`
- Read `-preffile` commandline argument
- Read registered GUI preference file

3. Read study preference file (hstbatch only)

- `<studyDir>/_usr/hstudy_pref.mvw`

6.5 Register Reports

Register custom reports in HyperStudy by editing the preferences file.

> **Tip:** Add to your list of reports with custom reports. To learn more, reference Create Reports.

Register Reports by Editing the Preferences File

Register reports in the `*BeginRegisterReportGenerators()` section of the `preferences_study.mvw` file.

1. In the **Editor**, open the `preferences_study.mvw` file.

   > **Tip:** You can find the `preferences_study.mvw` file in the HyperWorks installation directory under `<install directory>/hw/prefinc/`.

2. Search for the syntax `*BeginRegisterReportGenerators()`.
3. Immediately following the existing `*BeginRegisterReportGenerators()` statements, register further reports using the following syntax:

   ```
   *BeginRegisterReportGenerators(report_name)
   ```

   where:

report_name	Full path of the report, including file name and extension.

4. Save the `preferences_study.mvw` file.

Register* `.data` *Report in the Preferences File

In this example, a `.data` report is added to the reports section of the `preferences_study.mvw` file.

```
*BeginRegisterReportGenerators()
*RegisterReportGeneratorFile("C:\v1.py")
*EndRegisterReportGenerators()
```

6.6 Register External Fits

To use HyperStudy with a custom Python fit, it must be registered.

Register user utilities in HyperStudy or by editing the preferences file.

Register External Fits in HyperStudy

1. From the menu bar, click **Edit** > **Register External Fit**.
2. Add an external fit.
   a) Click **Add External Fit**.
   b) In the **Add** dialog, enter a label for the Fit and click **OK**.
3. In the Path column, enter the location of the external fit to be registered.
4. Click **Close** to exit the external fit registration.

Register External Fits by Editing the Preferences File

Register external optimizers in the `*BeginExternalFitDefaults()` section of the `preferences_study.mvw` file.

1. In the text editor, open the `preferences_study.mvw` file.

   You can find the `preferences_study.mvw` file in the HyperWorks installation directory under `<install directory>/hw/prefinc/`.
2. Search for the syntax `*BeginExternalFitDefaults()`.
3. Immediately following any existing `*RegisterExternalFit` statements, register further external optimizers using the following syntax:

   ```
   *RegisterExternalFit(fit_name,fit_script)
   ```

   where

   fit_name
   : Unique name for the fit, enclosed in parenthesis.

   fit_script
   : Full path of the external fit Python file, including file name and extension.

4. Save the `preferences_study.mvw` file.

Register `myfit.py` in the Preferences File

In this example, the `myfit.py` executable is added to the list of registered external fits.

```
*BeginExternalFitDefaults()
            *RegisterExternalFit("myFit", "C:/My_SQPOptimizer/myfit.py")
```

```
*EndExternalFitDefaults()
```

Keyboard Shortcuts

7

Keyboard shortcuts used to access HyperStudy features.

To do this	Press
Invoke contextual help.	F1
Invoke the Editor.	F8
Invoke the HyperStudy Evaluation tool.	F9
Switch the display of HyperStudy between full view and standard view.	F11
Create a new study.	Ctrl + N
Open a saved study.	Ctrl + O
Close the current study.	Ctrl + W
Close HyperStudy.	Ctrl + Q
Invokes the "work button" of each step. Work buttons include Import Variables, Apply, Evaluate Tasks, Evaluate Expressions, and so on.	Ctrl + D
Switch between the Explorer and Directory view.	Ctrl + J
Navigate steps in the Explorer.	Ctrl + Up/Down
Navigate tabs in the work area.	Ctrl + Left/Right