

# MODDE<sup>®</sup>

## MODDE<sup>®</sup> 12 User Guide



**© 1992-2017 Sartorius Stedim Data Analytics AB, all rights reserved**

Information in this document is subject to change without notice and does not represent a commitment on part of Sartorius Stedim Data Analytics AB. The software, which includes information contained in any databases, described in this document is furnished under license agreement or nondisclosure agreement and may be used or copied only in accordance with the terms of the agreement. It is against the law to copy the software except as specifically allowed in the license or nondisclosure agreement. No part of this user guide may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose, without the express written permission of Sartorius Stedim Data Analytics AB.

**Sartorius Stedim Data Analytics AB patents and trademarks**

US-7151976, US-7809450, US-8494798, US-8271103, US-8244498, US-9069345, US-8086327, US-8412356, US-8577480, US-8725469, US-9429939, US-9541471, US-7523384, US-7465417.

COMPUTER-IMPLEMENTED SYSTEMS AND METHODS FOR GENERATING GENERALIZED FRACTIONAL DESIGNS, Pending

<b>Trademark</b>	<b>Descriptor</b>
Umetrics™	Suite of Data Analytics Solutions
MODDE®	Design of Experiments Solution
SIMCA®	Multivariate Data Analysis Solution
OPLS®	Method for improved regression analysis
O2PLS®	Method for data integration
O2PLS-DA®	Method for exhaustive discriminant analysis
OPLS-DA®	Method for group separation
PLS-TREE®	Top down clustering
S-PLOT®	Highlighting discriminatory variables
EZinfo®	Embedded Waters solution
VALUE FROM DATA®	We are value providers

ID #2068

*User guide edition date: June 21, 2017*

Sartorius Stedim Data Analytics AB  
 Östra Strandgatan 24  
 SE-903 33 Umeå  
 Sweden

**Phone:** +46 (0)90 18 48 00

**Email:** [umetrics@sartorius.com](mailto:umetrics@sartorius.com)

# Welcome

Welcome to the user guide for MODDE 12. MODDE is part of the Umetrics™ Suite of Data Analytic Solutions.

This is your guide to MODDE and its capabilities. It explains how to install and use MODDE.

## **Content**

This user guide is divided into 14 chapters, 5 appendices and one reference list. Chapter 1 gives a short introduction of how to use MODDE. Chapter 2 presents an introduction to experimental design. Chapters 3 - 14 provide step-by-step procedures for creating and using experimental designs with MODDE.

The Statistical appendix presents short explanations of statistical methods used by MODDE.

The Design appendix presents short descriptions of the designs available in MODDE.

The Optimizer appendix describes the optimizer feature and the properties of the different optimizer objectives.

The Design space appendix describes the design space estimation feature.

The Generalized subset designs appendix describes how to create and work with Generalized subset designs for different applications such as multivariate calibration and stability testing design in MODDE.

References are available on the references page.



# Table of Contents

<b>01-Getting started</b>	<b>1</b>
Introduction.....	1
Installation .....	1
Getting started with MODDE .....	1
Getting help.....	2
Tutorials.....	2
Starting a new investigation .....	2
Experimental cycle .....	2
Design phase.....	2
Analysis phase .....	2
Explore the data .....	3
Evaluate the design .....	3
Fit.....	3
Review the fit using plots and lists.....	3
Diagnostics.....	3
Interpret the model .....	3
Refine the model .....	4
Prediction phase (using the model).....	4
Help .....	4
MODDE Help .....	4
Registration and activation.....	5
Sartorius Stedim Data Analytics .....	5
About MODDE.....	5
Using MODDE .....	6
MODDE ribbon description .....	6
Keyboard shortcuts.....	7
Mini toolbar .....	9
Quick access toolbar .....	9
Customizing the user interface .....	10
Investigation.....	10
<b>02-Introduction</b>	<b>13</b>
General description.....	13
What is modeling and experimental design? .....	13
Objectives of modeling and experimental design .....	13
Screening models and designs .....	13
Number of factors in screening designs .....	14
Number of factors with split objective .....	14
Response surface modeling (RSM) designs.....	14
Number of factors in RSM designs .....	14
Number of factors in split objective .....	15
Fit methods .....	15
Multiple Linear Regression (MLR).....	15
Partial Least Squares (PLS).....	15
Results.....	16

Analysis phase .....	16
Review the model fit .....	16
Assess model adequacy .....	16
Prediction - using the fitted model .....	17
Conventions .....	17
Limitations in investigation names .....	17
Limitations in factor and response names .....	17
Case sensitivity .....	17
Menu and tab reference syntax .....	17
Select and mark .....	18
Vector and matrix representation .....	18
Suggestions for further reading on experimental designs .....	18
<b>03-Design wizard</b> .....	<b>19</b>
Introduction .....	19
Factors .....	19
Factor definition dialog box .....	20
General tab .....	21
Advanced .....	24
Constraints .....	26
Defining constraints .....	26
Defining constraints in the spreadsheet .....	27
Constraints supported .....	27
Constraints in qualitative or quantitative multilevel factors .....	28
Responses .....	28
Defining responses .....	28
Response definition dialog box .....	28
Regular responses .....	30
MLR scaling .....	30
PLS scaling .....	30
Derived responses .....	31
Linked responses .....	33
Objective .....	33
Select objective .....	33
Model and design .....	35
Selecting model and design .....	35
Design runs .....	37
Center points .....	37
Replicates .....	37
Total runs .....	37
Blocks .....	37
Settings .....	38
How high should the power be? .....	39
D-Optimal .....	39
D-Optimal pages .....	39
Design generation criteria section .....	39
Design alternatives section .....	40
Candidate set section .....	41
D-Optimal results .....	42
D-Optimal onion pages .....	43
Onion on the Design tab .....	45
Reduced combinatorial .....	46

Design generation options.....	46
Reduced combinatorial design selection page.....	47
<b>04-File</b>	<b>49</b>
Introduction.....	49
Info.....	50
Protect investigation.....	51
Report.....	52
New.....	53
Experimental design.....	53
Using existing design.....	54
Specific application design.....	59
Open.....	64
Recent investigations.....	64
Recent folders.....	64
Browse.....	64
Save.....	65
Save as.....	65
Save plot as.....	65
Save list as.....	65
Save or Copy a plot.....	65
Print.....	67
Share.....	68
Send as attachment.....	68
Export to SIMCA.....	68
Close.....	68
Help.....	68
Activate and manage license.....	69
View help.....	69
Sartorius Stedim Data Analytics.....	69
Knowledge base.....	69
About us.....	70
Webshop.....	70
Options.....	70
MODDE options.....	70
Investigation options.....	73
Audit trail.....	76
Coefficients.....	77
Scaled and centered coefficients.....	77
Orthogonal coefficients in PLS.....	77
Unscaled.....	78
Normalized coefficients.....	78
Confidence interval.....	78
Extended or compact format.....	78
Interval type and probability levels.....	78
Residuals.....	80
Raw residuals.....	80
Standardized residuals.....	80
Deleted studentized residuals.....	80
Customize ribbon.....	81
Customize quick access toolbar.....	82
Keyboard shortcuts.....	82

Restore .....	84
<b>05-Home</b>	<b>85</b>
Introduction.....	85
Design wizard .....	86
Analysis wizard and One-Click .....	86
One-Click simple case.....	88
Replicates.....	89
Histogram.....	90
Coefficients .....	92
Summary of Fit .....	95
Residuals Normal Probability .....	96
Observed vs. Predicted .....	97
Specification .....	97
Spreadsheet access .....	98
Edit model.....	98
Edit Model dialog box.....	98
Model list .....	99
Fit model.....	100
Standard fit.....	100
Mixture fit .....	101
Summary of fit.....	101
Overview .....	102
Coefficients.....	104
Residuals.....	104
Observed vs. predicted.....	104
Properties .....	104
Contour .....	105
Sweet spot.....	105
Design space .....	106
Optimizer .....	106
Exclude .....	106
Undo .....	107
<b>06-Design</b>	<b>109</b>
Introduction.....	109
Factors .....	110
Factors spreadsheet .....	110
Responses .....	112
Responses spreadsheet .....	112
Constraints .....	112
Specifying constraints .....	113
Defining a constraint graphically .....	113
Modifying a constraint graphically .....	114
Candidate set with a constraint.....	114
Inclusions.....	115
Opening the inclusions spreadsheet.....	115
Inclusions vs. complement design.....	115
Adding inclusions to the worksheet .....	115
Inclusions as part of the design .....	116
Editing inclusions.....	116

Reference mixture.....	117
Generators.....	117
Objective.....	118
Design region.....	118
Design region properties .....	119
Design matrix.....	119
Design matrix for Stability testing.....	121
Overview tab.....	121
All tab .....	121
A, B, C etc. tabs .....	121
Design summary .....	122
Design summary D-Optimal.....	123
Design summary, GSD and Stability testing .....	124
Confoundings.....	126
To list the confoundings .....	126
Candidate set.....	126
Accessing the Candidate set .....	127
New design .....	127
New D-Optimal design .....	127
Select from already generated D-Optimal designs .....	127
Onion .....	128
Onion Plot.....	128
Onion 3D Plot .....	129
<b>07-Worksheet</b> .....	<b>131</b>
Introduction.....	131
Worksheet.....	131
Accessing the Worksheet .....	132
Description of the worksheet.....	132
Excluding qualitative settings .....	132
Missing values in the worksheet .....	132
Deleting the worksheet.....	133
Adding experiments in the worksheet .....	133
Sorting the worksheet.....	133
Colors in the worksheet.....	133
Sorting spreadsheet.....	133
Custom sort .....	133
Sorting the candidate set .....	134
Scatter .....	134
Run order.....	135
Randomize run order.....	135
Run order to detect curvature .....	135
Curvature diagnostics plot.....	136
Correlation .....	137
Descriptive statistics .....	138
Descriptive statistics properties.....	140
Box Whisker .....	141
Histogram .....	141
Accessing the Histogram.....	141
Replicates.....	142
Accessing the Replicate plot .....	142

Plot information .....	142
<b>08-Analyze</b>	<b>145</b>
Introduction.....	145
Summary of fit.....	145
Summary of fit plot.....	147
PLS total summary plot.....	147
PLS response summary plot.....	148
Summary of fit list.....	149
PLS summary list.....	150
ANOVA.....	150
ANOVA table.....	151
ANOVA plot.....	151
Lack of fit.....	152
Box-Cox plot (MLR only).....	154
Residuals.....	155
Residual type.....	155
Residuals normal probability plot.....	156
Residuals vs. predicted response.....	157
Residuals vs. run order.....	157
Residuals vs. variable.....	158
Residuals list.....	158
Observed vs. predicted.....	159
Properties.....	159
Coefficients.....	160
Coefficient plot.....	161
Coefficient overview plot.....	162
Coefficient list.....	162
Coefficient overview list.....	163
Effects.....	163
Accessing effects plots and list.....	164
Effect plot.....	164
Effects normal probability plot.....	165
Main effect plot.....	165
Effect list.....	167
Interactions.....	167
PLS.....	168
Score plots.....	169
Loading plots.....	169
Distance to model (Y).....	170
VIP - Variable importance in the projection.....	170
<b>09-Predict</b>	<b>173</b>
Introduction.....	173
Prediction spreadsheet.....	173
Prediction scatter plot.....	174
Prediction plots.....	175
Prediction plot.....	176
Overlay prediction plot.....	176
Factor effects.....	177
Factor effects, Stability testing.....	178

Contour .....	179
2D contour .....	179
4D contour .....	180
Response surface .....	181
Contour plot wizard .....	182
Contour plot options .....	183
Sweet spot .....	184
Creating a sweet spot plot .....	184
Design space .....	187
Design space wizard .....	188
First page .....	188
4D axes .....	189
Optimizer .....	190
Setpoint validation .....	190
Setpoint validation plots and lists .....	191
<b>10-View</b> .....	<b>193</b>
Introduction .....	193
Show .....	193
Audit trail .....	194
Favorites .....	195
Introduction .....	195
Full screen .....	197
Window .....	197
Close .....	197
Switch windows .....	198
<b>11-Tools</b> .....	<b>199</b>
Introduction .....	199
Add plot element .....	199
Available plot elements .....	200
Templates .....	200
Select .....	201
Free-form selection .....	201
Rectangular selection .....	201
Select along the X-axis .....	202
Select along the Y-axis .....	202
Move points .....	202
Zoom and zoom out .....	202
Zoom and rotate 3D plots .....	202
Screen reader .....	203
Exclude .....	203
Format plot .....	203
Accessing format plot .....	203
Mini toolbar .....	204
Axis .....	204
Axis Font and Title Font .....	206
Gridlines .....	206
Background .....	207
Titles and Font .....	207
Legend and Font .....	208

Limits and regions.....	208
Labels.....	209
Contour.....	210
Error bars.....	211
Column.....	211
Styles.....	211
List.....	211
Add to favorites.....	211
Add to report.....	211
Plots and lists.....	212
Properties dialog box.....	212
Automatic update of plots and lists.....	212
Generating multiple plots or lists.....	212
<b>12-Optimizer</b>	<b>213</b>
Introduction.....	213
Starting the optimizer.....	213
Optimizer theory.....	213
Optimizer objectives.....	214
Optimizer window.....	214
Objective tab.....	216
Optimizer desirability.....	217
Factor spreadsheet.....	219
CCC constraints.....	219
Setpoint tab.....	220
Alternative setpoints tab.....	221
Optimizer list.....	225
Optimizer tab.....	225
Dynamic profile.....	226
Creating contour plots from the optimizer.....	226
Contour plot wizard from the optimizer.....	227
Creating a sweet spot plot from the optimizer.....	227
Sweet spot plot wizard from the optimizer.....	228
Creating a design space plot from the optimizer.....	229
Design space wizard from the optimizer.....	229
Find robust setpoint.....	231
Setpoint analysis.....	234
<b>13-Setpoint</b>	<b>235</b>
Introduction.....	235
Setpoint analysis plots and lists.....	235
Setpoint window.....	236
Preventing auto update.....	236
Adding more columns.....	236
Setpoint properties.....	237
Response spreadsheet.....	240
Factor distribution.....	241
Response distribution.....	242
Setpoint list.....	242
Individual response analysis.....	242
Setpoint, design space.....	243

<b>14-Report</b>	<b>245</b>
Introduction.....	245
Opening the Report.....	245
Saving a report.....	246
Opening and saving an old report.....	246
Report window.....	247
File tab.....	247
General Windows commands.....	247
Save as.....	248
Home.....	248
Introduction.....	248
Clipboard.....	249
Font and Paragraph.....	249
Insert.....	249
Report.....	249
Formatting.....	250
View tab.....	250
Placeholders window.....	251
Properties window.....	252
Adding plots and lists to the report.....	252
<b>Statistical appendix</b>	<b>253</b>
Introduction.....	253
Fit methods.....	253
Multiple linear regression (MLR).....	253
Partial least squares regression (PLS).....	254
Cross-validation significance rules.....	256
Model.....	257
Hierarchy.....	257
Scaling.....	257
Scaling the factor matrix, X.....	257
Scaling the response matrix, Y.....	258
Missing data.....	258
Missing data in the design matrix, X.....	258
Missing data in Y with multiple linear regression.....	258
Missing data in Y with PLS regression.....	258
Number of observations.....	259
Summary of fit.....	259
R <sup>2</sup> .....	259
Q <sup>2</sup> .....	259
Model validity.....	260
Reproducibility.....	260
Residual standard deviation (RSD).....	260
Analysis of variance, ANOVA.....	261
Measures of goodness-of-fit.....	262
Q <sup>2</sup> .....	262
R <sup>2</sup> .....	263
Adjusted R <sup>2</sup> .....	263
F-test.....	263
Degrees of freedom and saturated models.....	264

Coefficients .....	264
Scaled and centered coefficients .....	264
Orthogonal coefficients in PLS .....	264
Unscaled .....	265
Normalized coefficients .....	265
Confidence interval .....	265
Extended or compact format .....	265
Analysis wizard features .....	266
Tukey's and variability tests .....	266
Auto transform and Auto tune .....	266
Square and interaction tests .....	267
Qualitative factors with more than 2 levels .....	268
Residuals .....	268
Raw residuals .....	268
Standardized residuals .....	268
Deleted studentized residuals .....	268
Condition number .....	269
Condition number definition .....	269
Condition number with mixture factors .....	270
PLS and Cox reference mixture model .....	270
MLR regression and the Cox reference mixture model .....	270
MLR and Scheffé models .....	270
Interval estimates .....	271
PLS plots .....	272
Loadings - WC plots .....	272
Scores - TT, UU, and TU plots .....	273
PLS regression coefficients .....	273
Box-Cox plot (only MLR) .....	273
Mixture data in MODDE .....	274
Mixture factors only, Model types .....	274
Slack variable model .....	275
Cox reference mixture model .....	275
Scheffé model .....	275
Mixture factors only, fitting a model .....	275
Analysis and fit method .....	275
MLR and the Cox reference mixture model .....	277
PLS and the Cox reference mixture .....	277
Scheffé models derived from the Cox reference mixture model .....	278
Scheffé models .....	278
Using the model .....	279
Process and mixture factors .....	279
MODDE plots .....	279
Optimizer .....	280
Desirability .....	280
Overall desirability .....	283
Overall distance to target .....	283
Starting simplex .....	284
Sensitivity analysis .....	284
Factor contribution .....	285
Orthogonal blocking .....	286
Blocking screening designs .....	287
Blocking RSM designs .....	288

Blocking D-Optimal designs.....	289
Random versus fixed block factor.....	289
Setpoint statistics .....	290
Monte Carlo simulations .....	291
Probability of failure and process capability indices .....	292
DPMO and Cpk definition .....	293
Predictions including model error.....	294
General subset designs algorithm and principles.....	294
<b>Design appendix</b> .....	<b>297</b>
Designs for process factors.....	297
Screening designs.....	297
Generalized Subset Designs .....	298
Stability testing design .....	299
RSM designs .....	301
Designs for mixture factors.....	304
Mixture and process factors .....	304
Mixture factors definition.....	304
Mixture constraint.....	305
Mixture experimental region.....	305
Classical mixture designs.....	306
D-Optimal designs.....	308
What are D-Optimal designs? .....	308
Candidate set.....	309
When do I use D-Optimal designs?.....	309
D-Optimal algorithm.....	310
Implementation of the D-Optimal algorithm in MODDE .....	310
Potential terms.....	310
Design evaluation.....	311
Inclusions and design augmentation.....	312
Irregular region .....	312
D-Optimal Onion designs .....	314
Screening onion designs.....	314
RSM onion designs .....	314
Candidate set.....	315
Definition OA .....	316
Definition of NOA .....	316
References.....	316
Power of the design .....	316
Power of the model .....	316
Post-hoc power analysis.....	318
How high should the power be?.....	318
References.....	319
<b>Optimizer appendix</b> .....	<b>321</b>
Introduction.....	321
Search function .....	321
Optimizer objectives .....	322
Accessing the individual desirability functions.....	322
Limit optimization.....	323
Target optimization .....	324
Customized desirability function.....	324

Focus optimization.....	325
Define optimizer specifications .....	325
Example of response specification in the optimizer .....	326
Optimizer search function.....	326
Robust setpoint.....	327
<b>Design space appendix</b> .....	<b>329</b>
Introduction.....	329
What is a design space? .....	329
Design space visualization.....	330
In-depth assessment of Design Space .....	332
Proven acceptable ranges.....	334
1 – Based on the robust setpoint.....	334
2 – Based on the dotted hypercube frame.....	335
3 – Based on setpoint analysis.....	335
Correlation in probability of failure.....	335
Setpoint analysis .....	336
Properties - Setpoint analysis.....	337
Monte Carlo simulations .....	337
Evaluate the results and make necessary adjustments .....	337
How to find the best Design Space.....	338
Setpoint validation for robustness testing.....	338
Setpoint validation example .....	339
Evaluation of the setpoint validation.....	339
Final adjustments .....	340
<b>Generalized subset designs appendix</b> .....	<b>343</b>
Introduction.....	343
Generalized Subset Designs example .....	344
Factor and response setup .....	344
Define subset designs.....	345
Replicates and Center points .....	346
Starting with balanced subset.....	347
Conclusion .....	349
Evaluation after excluding setting .....	349
Multivariate calibration, GSD.....	350
Stability testing design.....	351
Setting up a stability test .....	352
Define stability testing design setup time point design sets .....	353
Replicates and Center points .....	356
Stability testing design worksheet.....	356
Early stage data analysis – trajectory trending, Stability testing design .....	357
Early stage data analysis – assessment of factor effects.....	359
Late stage data analysis.....	361
Summary and discussion.....	362
<b>References</b> .....	<b>363</b>
<b>Index</b> .....	<b>365</b>

# 01-Getting started

---

## Introduction

This help guide is broken up into a number of chapters. To get started it is recommended that you read through the first three chapters as they contain information on installing and starting MODDE, the experimental cycle, how MODDE works, and designing an experiment.

---

## Installation

You can install and run MODDE under Windows 7, Windows 8 and 10.

***Note:** You must have administrative privileges to be able to install the software.*

To install and activate MODDE after purchase, locate the delivery letter and follow the steps described below:

1. Download the installation file from the Sartorius Stedim Data Analytics web page [www.umetrics.com](http://www.umetrics.com) using the link in the delivery letter. (Without the link the downloading is more cumbersome and includes entering name, address etc.)
2. Open the file and enter personal information as well as product information found in the delivery letter.
3. If you want the Audit trail to be automatically turned on and locked, select the **Force using Audit trail to log investigation events** check box.

After completing the installation, MODDE needs to be activated. Activation is done either

- (a) over the internet automatically
- (b) by finding and downloading a license file following the directions in the message boxes
- (c) from an internal license server or
- (d) importing a license file from Sartorius Stedim Data Analytics.

Option (d) should only be used in situations where activation according to (a) and (b) is not possible. See the delivery letter, sent to the license administrator at your company, for instructions. For more information, refer to the Activate and manage license section in Chapter 4, File.

---

## Getting started with MODDE

Open MODDE by clicking its icon:



## Getting help

To read about the MODDE software look in the built-in **Help** (contains the same information as the user guide). To open the help window, click the question mark icon

 at the top right of the MODDE window or click **File | Help | View help**.

---

---

**Hint:** Press F1 to open the help window.

---

---

## Tutorials

To run tutorial examples, go to [www.umetrics.com](http://www.umetrics.com) (**Downloads**), select an example, open the investigation used in the tutorial, and follow the analysis steps.

## Starting a new investigation

To start a new investigation, on the **File** tab, click **New**, then click **Experimental design**.

---

## Experimental cycle

The experimental cycle consists of three phases:

1. The design phase where you define your factors and within which ranges they should be varied, your responses, objective, design and model.
2. The analysis phase where you explore your data, review the raw data and the fit, review diagnostics in plots and lists, refine and interpret the model.
3. The prediction phase where you use the model to predict the optimum area for operability.

---

## Design phase

On the **File** tab, click **New**, and then click **Experimental design** to open the **Design wizard**. The **Design wizard** will guide you through defining your factors, responses, objective, constraints, and other information. See Chapter 3 for more information about the **Design wizard** and the steps involved in the design phase of your investigation.

Once you have completed your experiments, fill in the response data in the worksheet and change the factor settings as needed.

---

## Analysis phase

After the response values have been entered in the worksheet you can review the raw data, fit the model, review the fitted model, interpret the model, and refine the model. The **Analysis wizard** on the **Home** tab can help guide you through this phase.

## Explore the data

To explore the unfitted data use the **Worksheet** tab. Plots and lists available are: the curvature diagnostic plot, scatter plot, histogram plot, descriptive statistics list, correlation plot and matrix, and the replicate plot.

For details on the **Worksheet** tab plots and lists, see Chapter 7, Worksheet.

## Evaluate the design

The condition number is used to evaluate the goodness of the design. As a rule of thumb the condition number for screening designs should not exceed 3. For RSM designs it should not exceed 10.

## Fit

When you are ready to fit a model to your design, click **Fit model** on the **Home** tab. MODDE automatically fits using MLR when the condition number is low and there are no mixture factors. The fit methods available are MLR, PLS and, if using mixture (formulation) factors, Scheffé MLR is also available.

## Review the fit using plots and lists

After fitting the model, the **Summary of Fit** plot is displayed summarizing the fit in four columns:

- **R<sup>2</sup>**: Percent of the variation of the response explained by the model. R<sup>2</sup> overestimates the goodness of fit.
- **Q<sup>2</sup>**: Percent of the variation of the response predicted by the model according to cross validation, and expressed in the same units as R<sup>2</sup>. Q<sup>2</sup> underestimates the goodness of fit.
- **Model Validity**: A Measure of the validity of the model. When the Model Validity column is larger than 0.25, there is no Lack of Fit of the model (the model error is in the same range as the pure error).
- **Reproducibility**: The variation of the response under the same conditions (pure error), often at the center points, compared to the total variation of the response.

## Diagnostics

MODDE has a number of diagnostic plots, for instance:

- Residual plots to find outliers, drifts, trends etc.
- Box-Cox plot to select the best transformation of Y.
- ANOVA, ANalysis Of VAriance, in particular review the lack of fit. The estimation of lack of fit is only available when there are replicated points as it compares the pure error and the model error.

## Interpret the model

To interpret the influence of terms on the model use the Coefficient and Effect plots and lists. The interaction plot is particularly useful if your model has strong interaction terms. To display the interaction plot, on the **Analysis** tab, in the **Model interpretation** group, click **Interactions**.

When PLS is used for regression, scores and loadings can be plotted. These plots provide an overview of the data. On the **Analysis** tab, click **PLS** to select the score or loading plot you want to display.

## Refine the model

If you discover bad outliers or want to remove or add a term to the model you can refine your model.

To *remove outliers* or insignificant model terms use the interactive Exclude tool.



Click **Exclude** and then click/mark the outlier/term in a plot. You can also exclude it in the worksheet, right-click the specific cell and click **Exclude values**. The model is automatically refitted.

---

*Note: When excluding an outlier or model term in a plot, the outlier or model term is only excluded for the displayed response.*

---

You can *add and remove terms* from/to the model that are insignificant/significant for all responses, on the **Home** tab, in the **Model** group, click **Edit model**. To edit a common model for all responses, set the **For response** box to **All responses**.

After refining your model you should once more review the fit and diagnostics.

---

*Note: You can have a different model for each response in the same investigation.*

---

## Prediction phase (using the model)

When you are content with the model (fit, predictivity, lack of fit etc.) you can use the model to make predictions and find the “best conditions” area. The plots and lists to use for this are found on the **Predict** tab. Here you can find the **Contour** plots, **Prediction** plots, **Sweet spot** plot, **Optimizer** and the **Setpoint validation** feature.

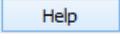
## Help

MODDE’s help is based on this user guide. The user guide documents are transferred to a compiled HTML file. To read the Help file Internet Explorer must be installed but does not need to be your default browser.

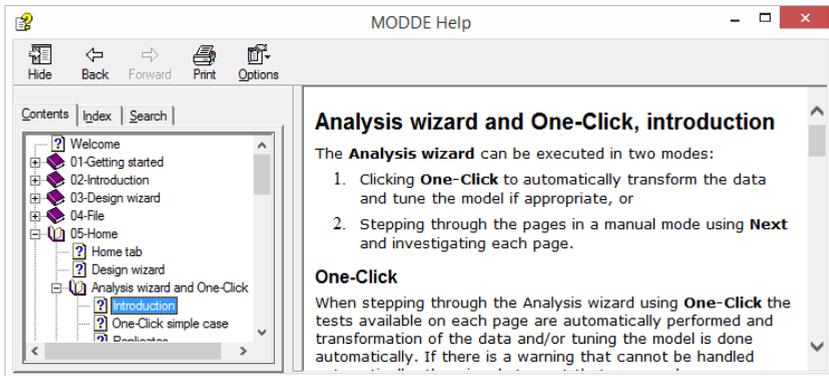
### MODDE Help

The help file is installed at the same time as MODDE and includes interactive help throughout the program.

To open MODDE's help:

- Click **Help**  in a dialog box or wizard.
- Click the help icon  in the top right corner of the MODDE window.
- Press **F1**.
- On the **File** tab, click **Help**, then click **View help**.

Use the **Contents**, **Index**, and **Search** tabs to find what you are looking for.



## Registration and activation

To register and activate, follow the installation instructions in the downloaded package.

To register later, on the **File** tab, click **Help**, then click **Manage license**.

## Sartorius Stedim Data Analytics

If you have an Internet connection, you can visit the web page of Sartorius Stedim Data Analytics ([www.umetrics.com](http://www.umetrics.com)) to get the latest news and other information.

On the **File** tab, click **Help**, then click **Sartorius Stedim Data Analytics** to visit the web page.

## About MODDE

To find license information and the version number of MODDE, on the **File** tab, click **Help**.

### About MODDE Pro

---

#### Version 12

Copyright © Sartorius Stedim Data Analytics AB

Umetrics MODDE, part of the Umetrics™ Suite of Data Analytics Solutions

MODDE® Design of Experiments Solution

### License

---

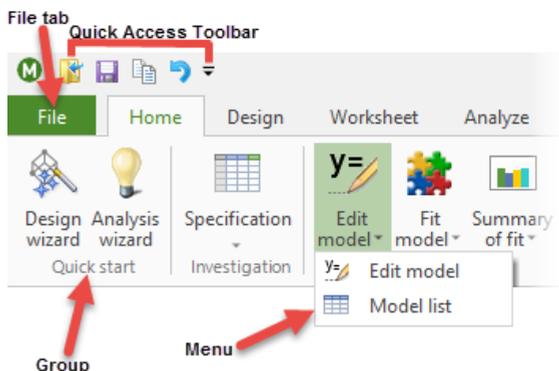
License:	Activated
License type:	
Product version:	MODDE Pro
Expiry date:	
Running on Windows Terminal Services:	
Running on virtual machine:	

HostID:

# Using MODDE

## MODDE ribbon description

MODDE's ribbon interface follows the standard guidelines that Microsoft recommends. The nomenclature explained here is used throughout the help guide for explaining where functions are located.

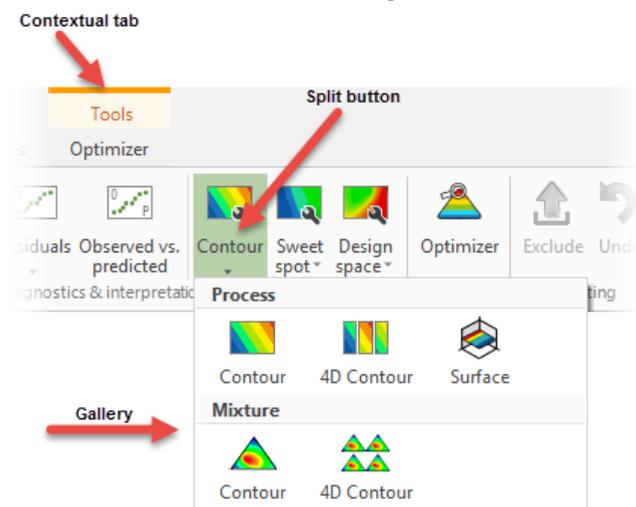


**File tab** - A menu of commands that involve the entire investigation or the active window, such as file-related commands.

**Quick Access Toolbar** - A collection of icons, located on MODDE's title bar, that provides shortcuts to commonly used commands. Users can add icons to this toolbar or remove them.

**Group** - A rectangular region on a tab that contains a set of related controls and commands. In the example above, the group's name is the **Quick start** group.

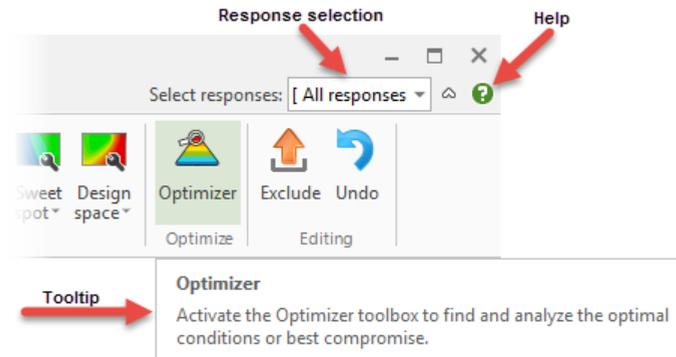
**Menu** - A list of functions that shows up when a button is clicked.



**Split button** - A split button is a button that completes two different actions depending on where the user clicks. Clicking the top half of the button with the image opens the most commonly used item in the gallery, or a wizard that can create any of the gallery items, while clicking the arrow under the button opens a gallery or menu offering more choices.

**Contextual tabs** - Tabs that are only displayed under certain circumstances. An example of this is the **Optimizer** tab that only appears after clicking **Optimizer** on the **Predict/Home** tab.

**Gallery** - A rectangular window that presents an array or grid of visual choices to a user.



**Select responses/Select factors** - In MODDE this box can be used to change the responses selected for the currently active plot. For some plots the **Select factors** box is available allowing selection of factor.

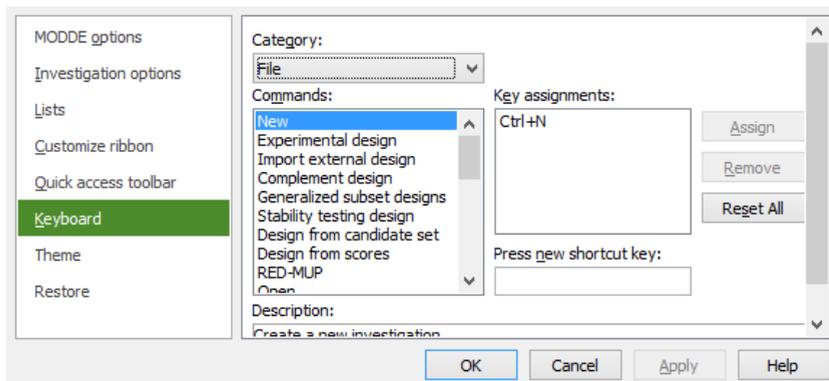
**Help** - Clicking the help icon opens MODDE's help.

**Tooltip** - A small window that displays descriptive text when a mouse pointer rests on a command or control.

## Keyboard shortcuts

MODDE includes many keyboard shortcuts in order to access commonly used functions. Many of these shortcuts will be familiar to Windows users, while some are MODDE specific.

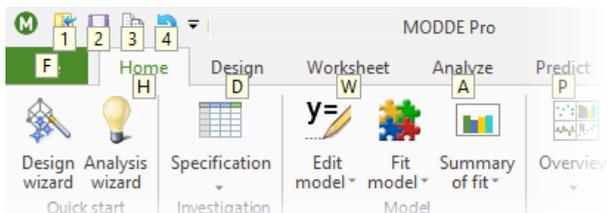
Key assignments can be modified as desired in the **Options** dialog, Keyboard page.



## KeyTips

In order to quickly navigate around MODDE using almost exclusively the keyboard instead of the mouse, MODDE implements the same system of KeyTips as Microsoft Office.

1. Press the Alt key to see the KeyTips.
2. Then press the indicated number or letter to run the associated command.



## General keyboard shortcuts

Copy: Ctrl+C, Ctrl+Insert

Paste: Ctrl+V, Shift+Insert

Cut: Ctrl+X, Shift+Delete

Print: Ctrl+P

Undo: Ctrl+Z, Alt+Backspace

Save: Ctrl+S

Delete: Delete

Select all: Ctrl+A

Insert: Insert

Full screen: F11

Start a new investigation: Ctrl+N

Open a different investigation: Ctrl+O

## List/spreadsheet specific

\* Ctrl+arrow up/down/left/right moves the cursor to corresponding edge of the worksheet.

\* Shift+arrow up/down/left/right - same as above but also selects cells.

\* Alt+PgUp/PgDn scrolls left/right by page.

\* Shift+Home/End scrolls the worksheet to the left/right marking cells.

\* Ctrl+Shift+Home/End marks from current cell to top left/bottom right.

\* In RED-MUP and Specification tabbed windows: Ctrl+PgDn/PgUp selects next/previous tab.

## MODDE specific keyboard shortcuts

Open the Design wizard: Ctrl+W

Open the Analysis wizard: Alt+W

Add to favorites: Ctrl+D

Add to report: Ctrl+R

Properties: Alt+Enter

## Mini toolbar

The mini toolbar appears when you click a plot or plot element and allows you to change the look of the plot or plot element. Some elements in plots and lists can be customized individually without using the **Format Plot** dialog box. This is useful if you would like one specific data point to stand out with a different shape, size, glow, or color. The settings icon on the mini toolbar allows you to turn off the mini toolbar completely.

To see if an element can be individually customized, click the element, then look at the mini toolbar to see the available options. Common options include:



Format plot - Opens the **Format Plot** dialog box.



Select - Swaps to the **Select** tool.



Zoom - Swaps to the **Zoom** tool.



Add plot element - Adds or removes a plot element such as a header, footer, timestamp, or regression line.



Symbol style - Choose between one of nine different styles for points in a plot.



Symbol size - Choose the size of the symbol for plot points.



Fill color - Changes the fill color.



Line width - Appears when clicking a line; changes the width of the line.



Line style - Appears when clicking a line; changes the style of the line.



Line color - Appears when clicking a line; changes the color of the line.



Font and font size - Change the font and font size of the text.



Grow font and Shrink font - Increase or decrease the size of the text.



Font color - Change the font color for the text clicked.



Bold and Italic - Change the text to bold or italic for the specified plot element.



Hide element - Hide the plot element.



Save as default style - Save a plot element as the default style for that element.



Settings - Settings related to the mini toolbar including the ability to turn it off.

## Quick access toolbar

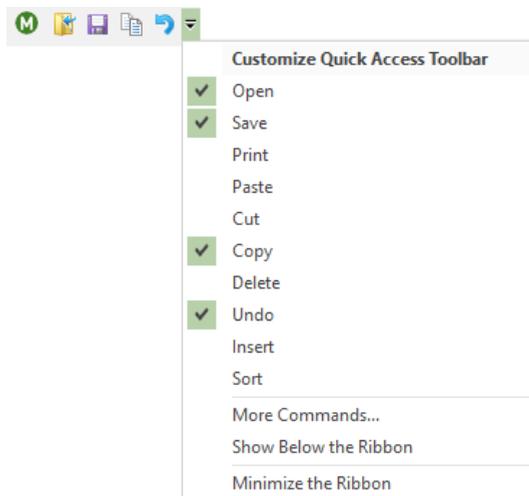
The quick access toolbar is the collection of icons at the top left of the MODDE window.

The default commands, from left to right, are:

- Open - Open investigation.

- Save - Save investigation.
- Copy - Copy the current plot or list.
- Undo - Undo last action.

The arrow to the right provides a menu with more options to add to the quick access toolbar. Click **More Commands...** in the menu to open the **Customize** window allowing you to add almost any command in MODDE to the quick access toolbar.



For more on the Copy to Clipboard dialog box that opens when copying a plot, see the Save and save as section in Chapter 4, File.

## Customizing the user interface

MODDE's user interface of ribbons, quick access toolbar, and keyboard shortcuts can be customized. You can add or remove functions from any of the tabs or the quick access toolbar, based on what you use most often. You can also add or change keyboard shortcuts so that your most commonly used commands are easy to access.

To customize MODDE's interface,

1. On the **File** tab, click **Options**.
2. Click the area you wish to change; **Customize ribbon**, **Quick access toolbar**, **Keyboard**, or **Theme**.
3. Make the desired changes and click **Close**.

All elements of the interface can be reset to their default state by clicking **Reset** on the appropriate screen. The default theme is *MODDE*, a customized theme.

## Investigation

Experimental plans in MODDE are organized into investigations. You can think of an investigation as a file folder containing all of the information related to a particular experiment. When you select or open a given investigation you can access, display and use all of its information. This information is organized in the following components: *factors, responses, constraints, inclusions, candidate set, model, design, worksheet, analysis, predictions, optimizer, audit trail, notes, and design space*.

### ***Managing investigations***

Investigations are binary files saved by MODDE with the extension \*.mip.

You can create new, open, and save investigations.

You can double-click a MODDE investigation (a \*.mip file), in Microsoft Explorer, to open that investigation.

MODDE does not save the fitted model. In order to review the results of the analysis and use make predictions, you need to fit the model by clicking **Fit model** or have **Automatic fit** turned on (default) in **File | Options, MODDE options**, when the investigation is opened. After the model has been fitted, you can open plots and lists to review the model and fit and create prediction plots and lists.

### ***Compatibility with older MODDE versions***

All investigations from MODDE 9 and upwards can be opened in this version of MODDE (the reverse is not true).



# 02-Introduction

---

## General description

MODDE - (**MO**deling and **DE**sign) is a Windows program for the generation and evaluation of statistical experimental designs.

Methods of statistical experimental designs have evolved since the pioneering work of Fisher in 1926. These methods, further refined by Box, Hunter, Scheffé, Tagushi and others, provide users with a powerful methodology for efficient experimentation.

---

## What is modeling and experimental design?

Experimental design is how to conduct and plan experiments in order to extract the maximum amount of information from the collected data in the presence of noise. The basic idea is to vary all relevant factors simultaneously, over a set of planned experiments, and then connect the results by means of a mathematical model. This model is then used for interpretation, predictions, optimization and identifying a design space.

---

## Objectives of modeling and experimental design

During an investigation you need answers to the following questions:

- Which factors have a real influence on the responses (results)?
- Which factors have significant interactions (synergies or antagonisms)?
- What are the best settings of the factors to achieve optimal conditions for best performance of a process, a system or a product?
- What are the predicted values of the responses (results) for given settings of the factors?

An experimental design can be set up to answer all of these questions.

---

## Screening models and designs

Screening is the first stage of an investigation where the goal is simply to identify the important factors. An important factor is a factor that causes substantial changes (effects) in the response when it varies.

In the screening stage simple models (linear or linear with interactions) are used, and experimental designs that allow the identification of the factors with the largest effects in the fewest possible number of experimental runs.

**MODDE supports:** Full factorial, Fractional factorial, General subset designs, L-designs, Plackett Burman, Rechtschaffner, Onion, D-Optimal designs, Reduced combinatorial designs, Definitive screening designs, Stability testing designs, and RED-MUP for screening experiments.

With mixture factors, MODDE supports the classical axial design when the region is a simplex.

---

*Note: The screening designs Generalized subset designs and Stability testing designs are available in the Specific application design section on the **File | New** menu.*

---

## Number of factors in screening designs

Process factors: MODDE supports up to 32 factors in problems involving process factors only.

Mixture factors: Up to 20 mixture factors are allowed in problems involving mixture factors only.

Process and mixture factors: In problems involving both mixture and process factors, up to 12 factors are supported.

## Number of factors with split objective

Split objective is available only when there are both process and mixture factors defined. Up to 12 factors are supported.

---

## Response surface modeling (RSM) designs

After screening, the goal of an investigation is usually to approximate the response by a quadratic polynomial (model) in order to:

- Understand in more detail HOW the factors influence the response; get a map of the system.
- Make predictions, optimize or find a region of operability.

**MODDE supports:** Three-level full factorial, central composite (CCC, CCO and CCF), Box Behnken, Rechtschaffner, Doehlert, Onion, and D-Optimal designs for RSM investigations.

With mixture factors, MODDE supports the modified simplex centroid when the experimental region is a simplex.

For investigations with only qualitative terms no square or cubic terms can be estimated. Creating RSM designs for such investigations is therefore impossible.

## Number of factors in RSM designs

Process factors: RSM designs are supported for up to 20 process factors

Mixture factors: Up to 15 mixture factors are allowed in problems involving mixture factors only.

Special Cubic Model is supported for up to 8 mixture factors, and the full cubic for up to 5 mixture factors.

Process and mixture factors: In problems involving both mixture and process factors, up to 12 factors are supported.

## Number of factors in split objective

Split objective is available only when there are both process and mixture factors defined. Up to 12 factors are supported.

---

## Fit methods

The data collected by the experimental design are used to estimate the coefficients of the model. The model represents the relationship between the response Y and the factors X1, X2, etc.

MODDE uses multiple linear regression (MLR) or Partial Least Squares (PLS) to estimate the coefficients of the terms in the model. MODDE recommends PLS when the investigation has a high condition number.

### Multiple Linear Regression (MLR)

With Multiple Linear Regression the coefficients of the model are computed to minimize the sum of squares of the residuals, i.e. the sum of squared deviations between the observed and fitted values of each response. The least squares regression method yields small variances for the coefficients and small prediction errors. It is important to note that MLR separately fits one response at a time and hence assumes them to be independent.

### Partial Least Squares (PLS)

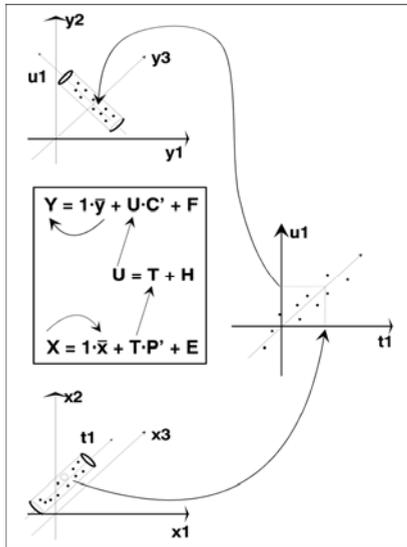
PLS deals with many responses simultaneously, taking their covariances into account. This provides you with an overview of how all the factors affect all the responses.

PLS has been extensively described in the literature so only a brief description is given here.

- PLS finds the relationship between a matrix Y (response variables) and a matrix X (model terms).
- PLS contains the multiple regression solution as a special case. With a single response or different models, and a given number of PLS dimensions, the PLS regression coefficients are identical to those obtained by multiple regression.
- The PLS model consists of a simultaneous projection of both the X and Y spaces on a low dimensional hyper plane with new coordinates T (summarizing X) and U (summarizing Y), and then relating U to T.

This analysis has the following two objectives:

1. To well approximate the X and Y.
2. To maximize the correlation between X and Y in the projected space (between u and t).



The dimensionality, number of significant PLS components, is determined by cross validation (CV), where PRESS (Predictive Residual Sum of Squares) is computed for each model dimension. MODDE selects automatically the number of PLS dimensions that give the smallest PRESS.

PRESS is then re expressed as  $Q^2 = (1 - \text{PRESS}/SS_Y)$ , where  $SS_Y$  is the sum of squares of  $Y$ .

## Results

Both MLR and PLS compute regression coefficients for each response. Thus  $Y$  is expressed as a function of the  $X$ 's according to the selected model (i.e. linear, interaction, or quadratic).

## Analysis phase

All results of model fitting, by MLR or PLS, are displayed in the same way, graphically and in lists.

### Review the model fit

Review the model fit by examining the following;

- Summary of the fit,  $R^2$ ,  $Q^2$ , Model validity, and Reproducibility for every response,
- Coefficients and their 95% confidence intervals,
- ANOVA table, and
- Effect plots for screening designs.

### Assess model adequacy

Assess the model adequacy further by reviewing the following plots;

- Normal probability plot of residuals,
- Plot of residuals against fitted values, run order or other factors, and
- Box-Cox plot to check for the optimal transformation of the response.

For PLS, summary of the fit by component and PLS score and loading plots are available.

---

## Prediction - using the fitted model

Use the fitted model to make predictions in the form of;

- Contour and rotatable 3D plots.
- Optimizer to find the “best conditions” for a desired profile of the responses. This helps in the interpretation of the results and to find a region of operability.
- Sweet Spot plot to draw a plot highlighting areas where the responses are within the specified ranges.
- Design Space to estimate the area of operability.
- Setpoint analysis and validation to investigate factor tolerance and robustness.

---

## Conventions

### Limitations in investigation names

The following characters cannot be used: = \ / : \* “ ? < > |.

### Limitations in factor and response names

The following characters cannot be used: ~ \* ? \ = [ ] and \$.

The length cannot be larger than 50 characters.

### Case sensitivity

MODDE is case insensitive. Lower or upper case in names will be displayed as entered, but for all comparisons lower or upper case are considered the same.

### Menu and tab reference syntax

In this user guide we use the following syntax when referring to the **File** tab commands:

- Click **Tab** | **Command**. An example: Click **File** | **Save**.
- On the **Tab** tab, click **Command**. An example: On the **File** tab, click **Save**.
- Click **Command** on the **Tab** tab. An example: Click **Save** on the **File** tab.

In this user guide we use the following syntax when referring to tab commands/buttons:

- On the **Tab** tab, in the **Group** group, click Command. An example: On the **Home** tab, in the **Model** group, click **Fit model**.
- Click **Command**, in the **Group** group, on the **Tab** tab. An example: Click **Fit model** in the **Model** group on the **Home** tab.
- Click **Tab | Command | Menu** item. An example: Click **Home | Specification | Worksheet**.

In this user guide we use the following syntax when referring to a tab/menu item or gallery item accessed by clicking a button:

- On the **Tab** tab, in the **Group** group, click **Button | Menu item**. An example: On the **Home** tab, in the **Diagnostics & interpretation** group, click **Contour | Surface**.
- On the **Tab** tab, in the **Group** group, click **Button**, and then click **Menu item/Gallery item**. An example for menu: On the **Home** tab, in the **Investigation** group, click **Specification**, and then click **Worksheet**. An example for gallery: On the **Home** tab, in the **Diagnostics & interpretation** group, click **Contour**, and then click **Surface**.
- Click **Tab | Button | Menu item**. An example for menu: Click **Home | Specification | Worksheet**. An example for gallery: Click **Home | Contour | Surface**.

## Select and mark

**Select** or **Mark** an item in plots, lists, or menus signifies clicking the item leaving it highlighted.

## Vector and matrix representation

R2 in MODDE is in this user guide denoted both R2 and R<sup>2</sup>.

Capital letters signify matrices, for ex. X, Y, P, T.

Letters typed in lower case signify vectors, for ex. x, y, p, t.

---

## Suggestions for further reading on experimental designs

1. Box, W.G. Hunter and J.S. Hunter, "Statistics for Experimenters", *John Wiley & Sons, Inc., New York (1978)*.
2. Box and N.R. Draper, "Empirical model-building and Response Surfaces", *John Wiley & Sons, Inc., New York (1987)*.
3. C.K. Bayne and I.B. Rubin, "Practical Experimental Designs and Optimization Methods for Chemists", *VCH Publishers, Inc., Deerfield Beach, Florida (1986)*.
4. Haaland, "Experimental Design in Biotechnology", *Marcel Dekker, Inc., New York (1989)*.
5. J.A. Cornell, "Experiments with Mixtures", *John Wiley & Sons, Inc., New York (1981)*.
6. D.C. Montgomery, "Design and Analysis of Experiments", *John Wiley & Sons, Inc., New York (1997)*.

# 03-Design wizard

---

## Introduction

The **Design wizard** provides guidance through setting up a new project or changing an existing one. The steps of the **Design wizard** are;

1. Define factors,
2. Define constraints that limit the design region,
3. Define responses,
4. Select objective, and
5. Select model and design.

### Accessing the Design wizard

To open the **Design wizard**, on the **Home** tab, in the **Quick start** group, click **Design wizard**.

---

---

**Hint:** Press Ctrl+W to open the **Design wizard**.

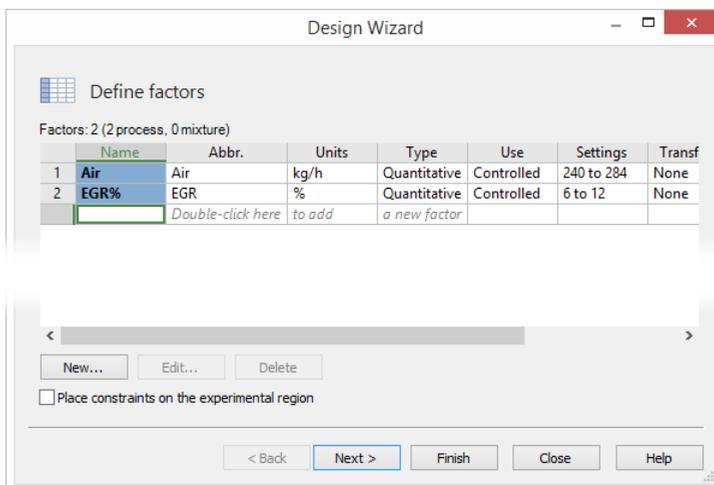
---

---

---

## Factors

The first page of the **Design wizard** is the **Define factors** page. This page contains the **Factors** spreadsheet so you can define (enter new factors), modify, and/or delete factors. Selecting the **Place constraints on the experimental region** box and clicking **Next** opens the **Constraints** page of the **Design wizard**. If there are no constraints placed on the experimental region, then clicking **Next** opens the **Define responses** page.

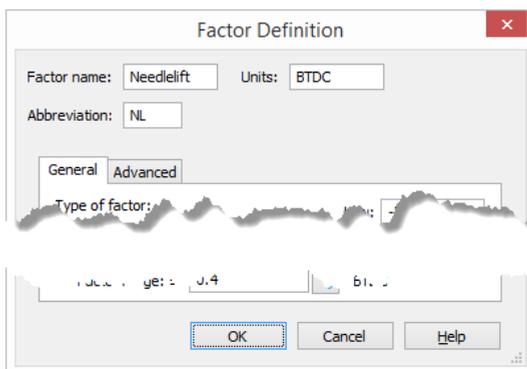


### Factor definition dialog box

The **Factor Definition** dialog box can be accessed either from the **Design wizard**, or from the **Factors** spreadsheet.

- On the **Design** tab, in the **Specification** group, click **Factors**. Double click the last row of the spreadsheet.
- From the **Design Wizard's** factor page, click **New...**

The **Factor Definition** dialog box is divided in an upper part and a lower part. The upper part displays **Factor name**, **Abbreviation**, and **Units** and is available independently of what is displayed in the lower part of the dialog box. The lower part has two tabs: **General**, which is the default tab when opening the dialog box, and **Advanced**.



#### Factor name

Enter the **Factor name** with up to 50 alphanumeric characters.

#### Abbreviation

The **Abbreviation**, used as the plot label in plots, is automatically filled with the first 3 characters of the factor name. You can change the abbreviation if desired, using up to 5 characters.

## Units

Enter the unit of the factor (optional). The units are displayed in the factor spreadsheet and can optionally be displayed in the worksheet, plots and lists. To display the units in the worksheet,

1. On the **File** tab, in **Options**, click **MODDE options**.
2. Set **Show units in lists** to *Yes*.
3. Click **OK**.

---

**Hint:** To set the units of a factor as °C (degrees Celsius), use Alt+0176 for the ASCII code. Hold down Alt, then press 0176 in succession, using the number pad on your keyboard.

---

## General tab

On the **General** tab of the **Factor Definition** dialog box, select the type of factor to be defined in **Type of factor**, the factor settings in **Low** and **High** or **Settings**, how it is used in **Use**, and the **Factor range**. When the type of factor is **Quantitative** or **Quantitative multilevel**, the **Advanced** tab has additional settings.

### Factor type and settings

Process factors are regular factors (i.e. temp, pH, etc.) that are not part of a mixture or formulation. They are expressed as amounts or levels, and can be varied independently of each other. **Quantitative**, **quantitative multilevel**, and **qualitative** factors are process factors.

Mixture factors are expressed as the fraction of the total amount of the formulation. Their experimental ranges lie between 0 and 1.

Mixture factors can be defined as **Formulation** or **Filler**.

For quantitative and formulation factors, the **Low** and **High** fields should be filled with the desired values.

For quantitative multilevel and qualitative factors, the levels planned to be used should be entered. For a qualitative factor each entry in **Settings** must be text, optionally including numbers. For quantitative multilevel each entry must be a number.

---

*Note: You can have both mixture factors and regular process factors defined as quantitative or qualitative in the same experiment.*

---

MODDE supports up to 12 factors when both process and formulation factors are defined in the same design. With only process factors, MODDE supports up to 32 factors for screening designs and up to 20 for RSM designs.

---

*Note: MODDE is limited in the range of the factor values. Factors ranging over a larger range (for example, low at 0.0001 and high at 10000) cannot be treated as factors.*

---

### **Quantitative (default)**

Quantitative factors are continuous factors defined at two levels, **Low** and **High**. To define more than two levels, click **Quantitative multilevel**.

### **Quantitative multilevel**

To specify more than two levels for a quantitative factor click **Quantitative multilevel**.

MODDE supports up to 24 levels for quantitative multilevel factors. Constraints are not allowed with this type of factor, and the available designs are **Generalized subset designs**, **Reduced combinatorial**, **Mixed full factorial**, **D-Optimal**, and three level designs when applicable.

The *Time* factor in the stability testing designs is a special multilevel factor, see details further down in this subsection.

### **Qualitative**

To specify a qualitative factor, click **Qualitative**.

**Qualitative** factors are discrete. For a qualitative factor, the levels should not stand in relation to each other. If the levels are a range, although discrete, then the factors should be defined as quantitative multilevel.

MODDE supports up to 24 levels for quantitative multilevel factors. Constraints are not allowed with this type of factor, and the available designs are **Generalized subset designs**, **Reduced combinatorial**, **Mixed full factorial**, **D-Optimal**, and three level designs when applicable.

RSM designs cannot be created with only qualitative terms. With one or more quantitative terms present up to 20 extended qualitative terms are allowed for RSM designs.

### **Formulation**

To specify a mixture factor, click **Formulation**.

Define a mixture factor as **Formulation**, when it is not an inert filler. Define the experimental range of the mixture formulation factor by entering its **Low** and **High** values.

MODDE supports up to 20 formulation factors in screening designs and up to 16 in RSM designs.

### **Filler**

Click **Filler** to specify a mixture factor as a filler when:

- It is always present in the mixture. That is, the sum of the **High** of the other mixture factors does not exceed 1.

- It accounts for a large percentage of the mixture.
- You are NOT interested in estimating the effect of the filler.

For a filler factor, the experimental range **Low** and **High** value fields are unavailable as it will be calculated as 1 minus the sum of the other mixture factors.

---

*Note: Only one mixture factor can be defined as **Filler**.*

---

A typical example of a filler factor is the solvent in a synthesis.

When you specify a filler factor, MODDE checks that the above conditions are met, and generates:

- A slack variable model (a model with the filler factor left out).
- Classical or D-Optimal process design.

If the conditions are not met, MODDE issues a message, and changes the filler factor to formulation.

### **Time**

When creating a stability testing design, the **Time** factor is predefined with the default settings 0, 3, 6, 9, 12, 18, 24, 36.

You can change these default settings, but the Time variable cannot be removed.

### **Use**

Under **Use** you can select how the factor will be used in the design. A factor can be **Controlled**, **Uncontrolled**, or **Constant**.

#### **Controlled (default)**

If the factor settings can be controlled, leave **Use** at the default setting (**Controlled**).

These factors can be regular process factors (i.e. pH, Temp, etc.) defined as quantitative, quantitative multilevel or qualitative, or mixture factors, defined as formulation or filler.

A **Filler** factor can only be defined as **Controlled**; other options are unavailable.

#### **Uncontrolled**

Define a factor as **Uncontrolled** (under **Use**) if it cannot be controlled, but you want to measure and record its value. Examples of such factors are ambient temperature or humidity.

Only regular quantitative factors can be uncontrolled, for all other factor types the option **Uncontrolled** is therefore unavailable.

#### **Constant**

Define a factor as **Constant** (under **Use**) when you want the worksheet to display a setting of a factor that is not changed.

Quantitative, qualitative, and formulation factors can be defined as **Constant** factors.

When mixture factors are constant, the mixture total **T** for the controlled mixture factors is equal to:  $T = 1 - \Sigma[\text{constant mixture factors}]$ .

Multilevel quantitative and filler factors cannot be defined as **Constant**.

### Precision in factor setting

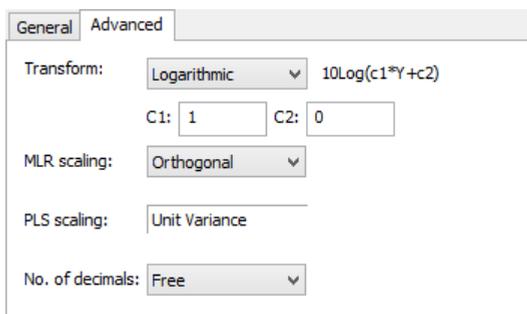
Specify the precision of the factor, on the **General** tab of the **Factor Definition** dialog box, under **Precision in factor setting**, enter a value in the **Factor range** field. The factor range is specified as  $\pm$  the given value.

To reset the precision to the default value (5% of factor range), click **Undo** .

The specified precision defines the limits of the 95% interval estimate. The precision is used in calculations of Probability of failure, Optimizer calculations, Setpoint analysis, Setpoint validation, and in the robust setpoint search.

### Advanced

For quantitative and quantitative multilevel factors, the **Advanced** tab of the **Factor Definition** dialog box provides the possibility to transform the factor and change the MLR scaling. For regular quantitative factors it is also possible to define the number of decimals to be used for the factor from this tab.



The screenshot shows the 'Advanced' tab of a dialog box with the following settings:

- Transform:** Logarithmic (dropdown), with the formula  $10\text{Log}(c_1 * Y + c_2)$  displayed to the right.
- C1:** 1 (input field)
- C2:** 0 (input field)
- MLR scaling:** Orthogonal (dropdown)
- PLS scaling:** Unit Variance (input field)
- No. of decimals:** Free (dropdown)

### Transform

To transform a factor, click the **Advanced** tab of the **Factor Definition** dialog box, and in the **Transform** box click the transformation of your choice.

When you transform a factor, the design is created in the transformed units, but the worksheet is expressed in original units. Hence transformation of a factor will change the center point and the star point values in the worksheet.

All transformed factors are displayed with a “~” (tilde) near the name in lists and plots.

The following transformations are available:

Transformation	Description
None	Default
Linear	$C_1 * Y + C_2$
Logarithmic	$\text{Log}_{10}(C_1 * Y + C_2)$
Negative Log	$-\text{Log}_{10}(C_2 - C_1 * Y)$
Exponential	$e^{(C_1 * Y + C_2)}$
Logit	$\text{Log}_{10}((Y - C_1)/(C_2 - Y))$
Power	$(C_1 * Y + C_2)^{C_3}$ where $C_3$ can be any value from -2 to 2.

When a transformation is selected (except **None**), the relevant constant fields are displayed.

The field **C3** is only displayed for the power transformation.

### ***MLR scaling***

When fitting with MLR, the factors can be scaled using orthogonal, mid-range, or unit variance scaling in the **Advanced** tab of the **Factor Definition** dialog box.

#### **Orthogonal (default)**

The factors are centered and scaled using the mid-range and Low and High values from the factor definition. This is the system default.

#### **Mid-range**

When factor is scaled using mid-range, it is centered only using the mid-range of the factor. Mid-range is calculated as  $(\text{High} + \text{Low}) / 2$  where High and Low are the values available in the worksheet.

#### **Unit variance**

When you select to scale a factor to unit variance the worksheet columns are scaled and centered to unit variance in the calculations.

#### **Mixture factors**

When fitting the model with MLR, the mixture factors are not scaled.

The model can be fit in pseudo components. This is recommended when the mixture region is regular.

For investigations containing both process and mixture factors, by default process factors are orthogonally scaled and the mixture factors are unscaled. The coefficients displayed as scaled and centered correspond to this default scaling of the variables i.e. mixture unscaled and process orthogonally scaled. If you select to display the unscaled coefficients, they correspond to all factors unscaled, including the process factors.

---



---

**Hint:** Select the same scaling for all the factors of the same type; the system default is recommended.

---



---

### ***PLS scaling***

When fitting the model with PLS, all factors including mixture factors are always scaled and centered to unit variance.

For mixture factors, when you select a fit method with pseudo components, the mixture factors are first transformed to pseudo components and then scaled to unit variance (pseudo components can be switched on/off).

### ***No. of decimals***

In the **Advanced** tab of the **Factor Definition** dialog box, you can select the number of decimals.

The values for number of decimals are: **Free, 0, 1, 2, 3, 4** and represent the number of digits displayed after the decimal point.

The **No. of decimals**-value should correspond to the precision with which the factor can be set in your equipment. It is important that it is not set too low since after setting this value all values for that factor will be rounded accordingly in the worksheet. If you do not know the precision of the instrument, leave **No. of decimals** set to **Free**. **Free** means that no rounding off of the results and values corresponding to this factor will take place.

For example, the settings of a factor in a CCC design, is its high value (in orthogonal scaled units) multiplied by the 4<sup>th</sup> root of the number of runs in the factorial part of the design. If the precision of this factor is set to 0 all decimal digits are removed from this factor setting in the worksheet and computation will take place using the values in the worksheet.

---

## Constraints

A common problem is that experimentation may not be possible in some region of the experimental space. For example it may not be possible to have high temperature and simultaneously low pH, and you want to cut-off the corner High temp, Low pH. In MODDE this is solved by adding a constraint.

A linear constraint is a function of the factors that specify a part of the experimental region to be included or excluded.

The resulting experimental region is an irregular polyhedron. The corners of this region are called the extreme vertices; they constitute part of the candidate set, i.e. a discrete set of potentially good runs.

D-Optimal designs are the only designs available when the experimental region is constrained to an irregular polyhedron.

Constraints can be defined for quantitative or formulation factors.

### Defining constraints

The constraints page of the **Design wizard** provides the opportunity to limit the possible region for the design. This page is only available if the **Place constraints on the experimental region** check box is selected on the factors page of the **Design wizard**.

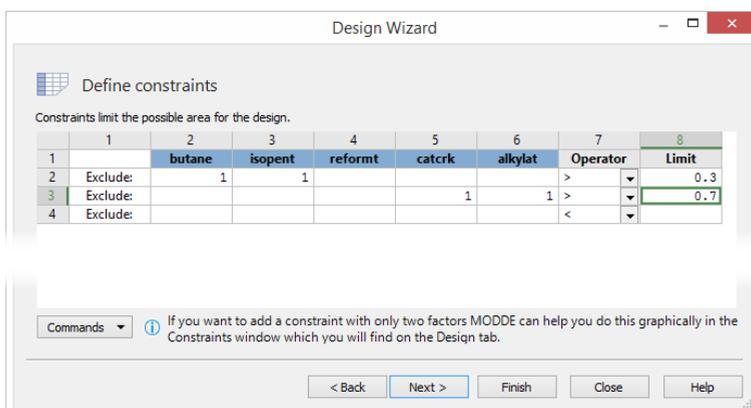
In the spreadsheet, each constraint (one per row) is defined as a mathematical relation.

---

***Note:** To add a constraint with only two factors, MODDE can help you do this graphically once the **Design wizard** is closed. On the **Design** tab, in the **Specification** group, click **Constraints**.*

---

Click **Next** to go to the response page of the **Design wizard**.



## Defining constraints in the spreadsheet

To define a constraint in the spreadsheet, enter the coefficients  $A_k$  of every factor in the constraint. Select “<” or “>” and enter the **Limit** of the constraint.

### *An example of entering a constraint in the spreadsheet*

For example, in an experiment with four mixture factors:

**X1** qualitative 3 levels: A, B, C

**X2** qualitative 2 levels: K, L

**X3** mixture

**X4** mixture

**X5** mixture

**X6** mixture

A set of constraints may be entered as follows:

	1	2	3	4	5	6	7
1		X3	X4	X5	X6	Operator	Limit
2	Exclude:	1		1	1	<	0.6
3	Exclude:		1		1	<	0.3

The first constraint specifies to exclude the experimental region where the sum of X3, X5, and X6 is  $< 0.6$ .

The second constraint specifies to exclude the experimental region where the sum of X4 and X6  $< 0.3$ .

## Constraints supported

MODDE supports linear constraints, specified as exclusions, for quantitative process factors or mixture factors  $X_k$  of the form

$$\sum A_k X_k < Limit$$

or

$$\sum A_k X_k > Limit$$

A constraint cannot be defined to include both quantitative and formulation factors.

Constraints cannot be defined in quantitative multilevel, qualitative, filler, uncontrolled, or constant factors.

MODDE supports up to 50 linear constraints.

## Constraints in qualitative or quantitative multilevel factors

Qualitative factors and quantitative multilevel do not appear in the constraint spreadsheet and cannot be used in constraints.

When you have a constraint between two qualitative factors, it is recommended that you convert them to one single qualitative factor.

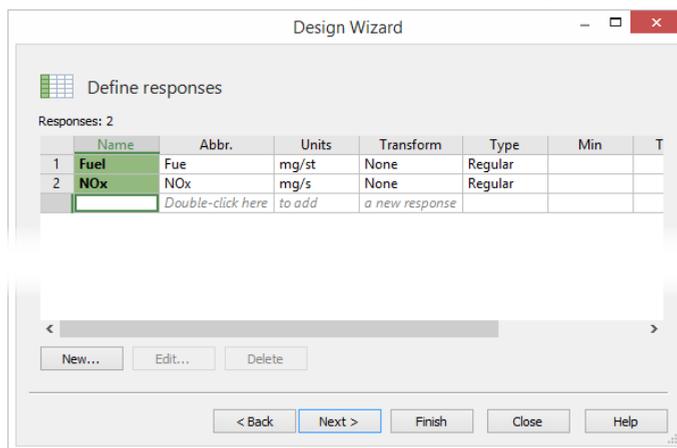
For instance, you have Granulators A, B, and C with Chopper or not but Granulator C does not have a Chopper. You should then, instead of specifying two qualitative factors 'Granulator' and 'Chopper', specify one factor 'GranulatorChopper' with the five settings 'GranulatorA\_ChopperYes', 'GranulatorA\_ChopperNo', 'GranulatorB\_ChopperYes', 'GranulatorB\_ChopperNo', and 'GranulatorC\_ChopperNo'.

Alternatively, delete the undesirable combination in the candidate set.

## Responses

### Defining responses

In the response page of the **Design wizard** you can define (enter new response), modify, or delete responses. Clicking **Next** opens the objective page.



### Response definition dialog box

The **Response Definition** dialog box can be opened by double-clicking the last row of the **Responses** spreadsheet or clicking **New...** on the response page of the **Design wizard**.

**Hint:** You can also just type in the last row of the **Responses** spreadsheet to add a response.

The image shows a 'Response Definition' dialog box with the following fields and values:

- Response name: Soot
- Units: mg/s
- Abbreviation: So
- Options tab:
  - Transform: Logarithmic (Formula:  $10\text{Log}(c1*Y+c2)$ )
  - C1: 1, C2: 0
  - MLR scaling: None
  - PLS scaling: Unit Variance
  - Min: (empty)
  - Target: (empty)
  - Max: 1
- Buttons: OK, Cancel, Help

### **Response name**

Enter the response name with up to 50 alphanumeric characters in the **Response name** field.

### **Abbreviation**

The **Abbreviation** is automatically filled with the first 3 characters of the response name. You can change the abbreviation if desired using up to 5 characters. The abbreviation is used as the plot label in plots.

### **Units**

Enter the unit of the response (optional). The units are displayed in the response spreadsheet and can optionally be displayed in the worksheet, plots and lists. To display the units in the worksheet,

1. On the **File** tab, in **Options**, click **MODDE options**.
2. Set **Show units in lists** to *Yes*.
3. Click **OK**.

### **Selecting type of response**

There are two types of responses: **Regular** and **Derived**. After defining the response and exiting the **Response Definition** dialog box it is not possible to change the type.

### **Limits**

Fill in the **Min**, **Target**, and **Max** fields if that information is available to you. These values are then automatically used in the **Design Space**, **Sweet Spot**, and **Optimizer** windows and displayed in plots where applicable.

## Regular responses

**Regular** responses are the standard responses measured, entered and fitted in the current investigation. Regular responses can be transformed and it is also possible to change the modifier for PLS scaling.

### Transformation

The following transformations are available:

Transformation	Formula	Formula to back transform
None	Default	
Lin	$C_1 * Y + C_2$	$(x - C_2) / C_1$
Log	$\text{Log}_{10}(C_1 * Y + C_2)$	$1/C_1 * (b^x - C_2)$
Negative Log	$-\text{Log}_{10}(C_2 - C_1 * Y)$	$1/C_1 * (C_2 - b^x)$
Exp	$e^{(C_1 * Y + C_2)}$	$1/C_1 * (\ln(x) - C_2)$
Logit	$\text{Log}_{10}((Y - C_1)/(C_2 - Y))$	$(10^x * C_2 + C_1) / (10x + 1)$
Power	$(C_1 * Y + C_2)^{C_3}$ where $C_3$ can be any value from -2 to 2.	$(x^{1/C_3} - C_2) / C_1$

When a transformation is selected (except **None**), the constants in the formula are entered in the fields displayed after selecting a transformation. The **C3** field is only displayed for the power transformation.

Specifying a transformation for a response is done to get the best mathematical fit of the estimated function.

---

*Note: You can specify or modify the current transformation by right-clicking the **Histogram plot** and clicking **Transform**.*

---

## MLR scaling

When fitting the model with MLR no scaling of responses is available.

## PLS scaling

When fitting the model with PLS it is possible to scale to unit variance with or without a modifier.

### Unit Variance (default)

With the default scaling option, the responses are centered and scaled to unit variance when fitting.

### Autoscale Modifier

In the **PLS scaling** box, select *Autoscale Modifier* to change the modifier. Leaving the modifier at the default (1) gives the same result as when selecting *Unit Variance*. Enter a different value of the modifier and the response will be scaled to unit variance multiplied by the value of the modifier.

---

*Note: To keep a response out of the analysis set its autoscale modifier to 0.*

---

## Derived responses

A derived response is a computed response as a function of the factors and/or fitted regular responses. When you add a derived response, you enter its formula. Derived responses can be edited and deleted.

Derived responses are displayed in the response spreadsheet. The values of the derived responses are entered automatically in the worksheet when the model is fitted.

---

*Note: Derived response values are only available after fitting the model. When responses are included in the formula, MODDE uses the fitted (predicted by the model) values of the responses in the computation.*

---

### Defining derived responses

To create derived responses open the **Response Definition** dialog box, enter the name, abbreviation and units of the response.

In **Type**, click **Derived**, and then click **Edit...**

This opens the first page of the **Derived Response Wizard**, which contains information about derived responses.

Clicking **Next** opens the page in which you enter the formula for the derived response.

Derived Response Wizard

Enter the formula for the derived response. Use the symbols 'v1', 'v2', etc for the available factors and responses shown in the list below.

Formula:

Available variables for use in the formula:

Sy...	Name	C..
v1	Air	
v2	EGR%	
v3	Needlelift	
v4	Fuel Pred	
v5	NOx Pred	
v6	Soot Pred	

< Back    Next >    Finish    Cancel    Help

When you click **Finish**, MODDE parses the formula for correctness, and only computes and displays the derived response in the worksheet, when you fit the model.

The derived response is added to the response spreadsheet and the worksheet.

### Modifying a derived response

To modify a derived response, you must edit its formula in the **Response Definition** dialog box. You cannot edit the values directly in the worksheet.

### Copying or deleting a derived response

Derived responses can be copied and deleted as regular responses.

---

*Note: Derived responses are deleted when regular factors are deleted, or changed and when responses that are part of the derived response are deleted.*

---

## Using sets of variables in derived responses

In MODDE you generate **one derived response at a time**.

Hence you can only use sets of variables with the operators **avg**, **stdev**, and **sum** that return one variable.

Examples using sets of variables

**Avg(v[1,3,4] + v[6,8,9])**

Results in the average of 3 variables,  $v_1+v_6$ ,  $v_3+v_8$ ,  $v_4+v_9$ .

**Avg(v5 + v[1:6])** is an illegal syntax, the two operands are not of the same size.

**Sum(v[1,3,4] + v9)**

Results in the sum of 3 variables,  $v_1+v_9$ ,  $v_3+v_9$ ,  $v_4+v_9$ .

**Stdev(v[3:5]\*v1)**

Results in the standard deviation of the 3 variables  $v_3*v_1$ ,  $v_4*v_1$  and  $v_5*v_1$ .

**Avg(v[1:5]\*v8/v7)**

Results in the average of 5 variables, i.e. the average of variables 1 to 5 each multiplied by the ratio  $v_8/v_7$ .

**Sum(v8/v7 \* v[1:5])** is an illegal syntax.

## Derived response syntax

MODDE recognizes the following syntax:

- Integer, Variables (factors or responses), List of Integers and Sets of Variables.
- Integer constant and floating points.
- Operator ‘:’ denotes a sequence, i.e., from: to, for example 6:8 means 6,7,8.
- List of integers, such as 1,3,5:8 is the same as 1,3,5,6,7,8.
- Variables (factors or responses) are denoted by *vint*, where *int* refers to the variable number in the worksheet (i.e. v5, v15, for variables 5 and 15).
- A set of variables (matrix) is denoted by v[int1,int2, int3:int4,int6] with **square brackets**. *Int* refers to the variable number in the worksheet. For example v[1,5,7:10] refers to the set of variables v1, v5, v7, v8, v9, and v10.

---

*Note: To denote a set of variables use the square brackets [ ], and not regular parenthesis ( ).*

---

## Operators and functions in derived responses

The operators and functions listed below are recognized and can be used with a single variable or a set of variables. Operators have the usual precedence, i.e.  $^ > * > / > + > -$ . Parenthesis can be used to group expressions in the usual way.

### Functions

The functions available are *Log10*, *Ln* (natural log), and *Exp* (exponential).

### Addition and subtraction

Addition/subtraction (+, -) can be applied to:

- A set of variables with a constant.
- A set of variables with a single variable.

- A set of variables with another set of variables of *the same size* (they are added pair wise).
- A single variable with a constant or a single variable.

### Power, multiplication, and division

Power, multiplication, and division (^, \*, /) can be applied to:

- A set of variables with a constant.
- A set of variables with a *single* variable.

---

*Note: Power, multiplication and division cannot be applied to a set of variables and another set of variables. The first operand can be a constant, a variable or a set of variables, but the second and following operands must be a single variable or a constant.*

---

### Additional operators

The following additional operators apply to variables or sets of variables:

<b>Avg(v[int1:int2])</b>	Average of variables vint1... to vint2
<b>Stdev(v[int1:int2])</b>	Standard deviation of variables vint1... to vint2
<b>Sum(v[int1:int2])</b>	Sum of variables vint1... to vint2

---

*Note: The parser is not case sensitive (t and T are treated identically by the parser).*

---

### Qualitative factors in derived responses

When you use a qualitative factor in the formula for a derived response, enter the values (weights), to be used when computing the derived response, for each qualitative level setting. If no settings are entered, '0' is used as value for all settings of the qualitative factor.

### Linked responses

A **Linked** response is a response available in one investigation but fitted in another. Linked responses are no longer available in MODDE.

In MODDE 10 and later investigations containing linked responses are converted to hold different models and/or worksheets. MODDE supports as many models in one investigation as there are responses.

---

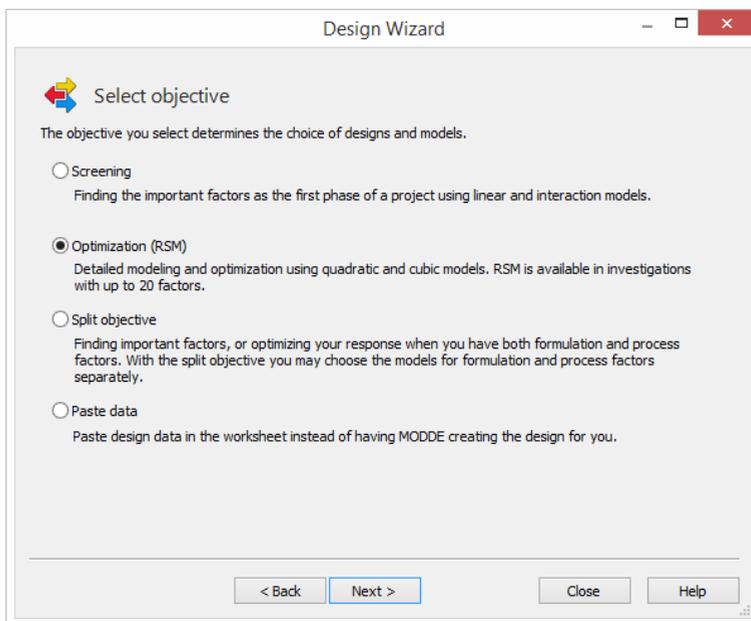
## Objective

### Select objective

On the **Select objective** page of the **Design wizard**, select the objective of the investigation.

The **Objective** is the purpose for creating the design. MODDE recognizes two objectives: **Screening** (first stage of an investigation when little is known) and **Optimization (RSM)** (optimization with the important factors.). The **Split objective** supports both screening and optimization, as does **Paste data**.

Click **Next** to open the model and design page of the **Design wizard**.



### **Screening**

Select the **Screening** objective when:

- You are starting an investigation and know little about the effects of the factors on the response, the behavior of the response in the experimental region, or the true size of that region.
- The goal is to reduce the number of factors to those with the largest effect on the response.

This objective is available for all types of factors and factor combinations.

### **Optimization (RSM)**

Select the **Optimization (RSM)** objective when:

- A lot is known about the investigation i.e. important factors, the size of the region etc.
- The goal is to approximate the response by a mathematical model for the purpose of prediction, optimization or finding a region of operability.

This objective is not available when all factors are qualitative.

### **Split objective**

Select the **Split objective** when:

- The investigation holds both process and mixture factors
- **AND** you want to specify **separate** models for each.

If you want to specify one model for both mixture and process factors, click **Screening** or **RSM** as objective.

The split objective is only available when there are both process and mixture factors available.

### Paste data

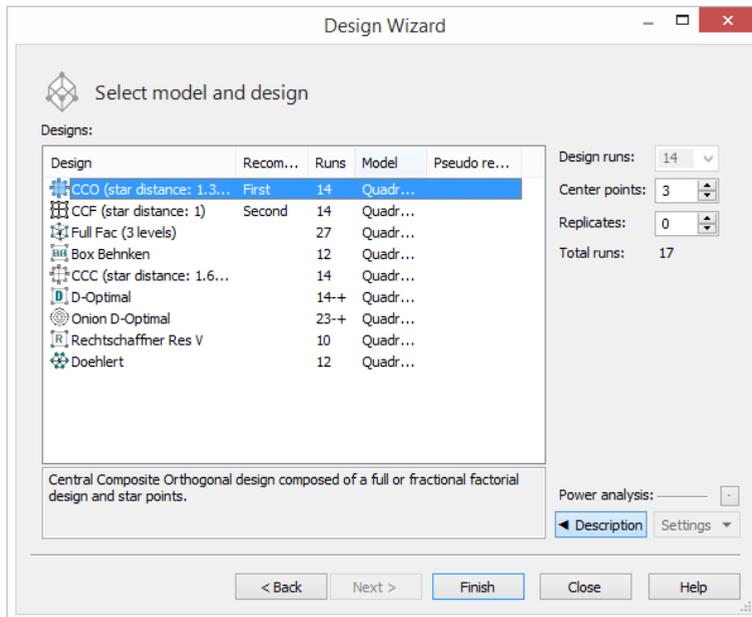
Click **Paste data** when you have the design and want to paste it instead of MODDE creating one for you.

After selecting **Paste data** and clicking **Finish**, the MODDE worksheet will expand dynamically to fit the size of the pasted data.

## Model and design

### Selecting model and design

The **Select model and design** page of the **Design wizard** allows for the choice of design and model.



**Hint:** Sort the list on a selected column by clicking its header.

If your design is classical, click **Finish** to generate the worksheet. If your design is D-Optimal or Reduced combinatorial, click **Next** to open the respective next pages of the wizard.

### Designs in MODDE

The design is the protocol for varying the factors in each experiment. Thus the design is a set of experimental runs spanning the experimental region.

Clicking **Description** reveals a short description of the currently marked design.

### ***Recommended designs***

According to the selected objective and number of factors, MODDE recommends two designs. Continue with the recommended design or select another one by pointing and clicking or pressing the **UP** and **DOWN** arrow keys on the keyboard. MODDE recommends classical designs whenever possible.

The recommendations are marked **First** and **Second** in the **Recommendation** column.

### ***Runs in design***

In the **Runs** column the number of runs in the design is displayed. When there is a '+' and/or a '-' sign after the number that means that the number of runs can be changed for that particular design.

### ***Model***

MODDE supports polynomial models, such as linear, interaction, and quadratic. Third order terms such as cubic or three factor interactions may be added to the model after design generation. On the **Home** tab, in the **Model** group, click **Edit model** to open the **Edit Model** dialog box.

### ***Screening models***

**Linear** and **Interaction** models are appropriate for the screening objective. When the model you select is:

- Linear – MODDE generates the linear model. You may edit the model and enter selected interactions.
- Interaction – MODDE generates the full interaction model with all the two factor interactions included.

### ***RSM models***

Quadratic models are used for the RSM objective. For classical mixture designs cubic designs are also available. When the model you select is:

- Quadratic – MODDE generates the full quadratic model holding all two-factor interactions and all the square terms of all the factors.
- Special cubic or cubic – MODDE generates models accordingly. Such models are only supported with mixture factors and include all two-factor interactions, all square terms, and some or all cubic terms.

### ***Split models***

When selecting the **Split objective**, the model for the process factors and the mixture factors can be specified independently of each other by clicking **Settings**.

### ***Pseudo resolution for blocked designs***

The pseudo resolution of the design is the resolution of the design when all the block effects (blocking factors and all their interactions) are treated as main effects under the assumption that there are no interactions between blocks and main effects, or blocks and two factor interactions.

The *Pseudo resolution* applies to designs when they are blocked. This means that a number higher than one has been selected in the **Blocks** box.

## Power

The *power* of a design is calculated from the guessed R2, desired alpha and selected number of runs. In this power calculation all runs and center points are used.

Conventionally, a power that is  $> 0.8$  is typically said to be good (just as alpha is typically set to 0.05). A higher power is of course better.

See also the Power - how high should it be? subsection later in this chapter and the Power of the design subsection in the Design appendix.

## Design runs

The **Design runs** box displays the number of runs for the selected design.

When there exist 2 or more fractional factorial designs of the same resolution, with different number of runs, the number of runs, in the **Runs** column, is marked with a “+”. MODDE defaults to the design with smallest number of runs. Use the **Design runs** arrow to select the larger design.

For example, with the screening objective for 7 or 8 factors there exist two resolution IV designs, one with 16 runs and the other with 32. By default MODDE selects the design with 16 runs. To select the design with 32 runs, click the **Design runs** arrow.

With D-Optimal designs the number of runs, in the **Runs** column, is marked with a ‘+’ and ‘-’ indicating that there exist smaller and larger designs.

When augmenting a design D-optimally, the number of runs includes the number of inclusions.

## Center points

The **Center points** box displays the number of center points. MODDE always recommends 3 center points. To change the number of center points from inside the **Design wizard**, on the **Select model and design** page enter the desired number of center points in the **Center points** box.

## Replicates

The **Replicates** box on the Select model and design page of the **Design wizard** displays the number of times to replicate the whole design including center points. The default is '0', meaning that the design is not replicated. Enter '1' here to replicate the design once. To change the number of replicates, enter the desired number in the **Replicates** box.

## Total runs

**Total runs** on the **Select model and design** page of the **Design wizard** is the total number of runs included in the worksheet is listed and includes:

- Design runs
- Center points
- Replicates.

## Blocks

MODDE supports Orthogonal blocking for the 2 levels Factorial, Fractional factorial, Plackett Burman, CCC, Box Behnken, and D-Optimal designs.

The maximum number of blocks supported by MODDE is 9, with a minimum block size of 2.

Select the number of blocks to include in your design from the **Blocks** box.

### ***Orthogonal blocking***

The method of dividing experiments into blocks, so that the block effect is uncorrelated with the main factor effects is called orthogonal blocking.

Orthogonal blocking is a way to deal with extraneous sources of variability that are not included in the model. For example if one is making 32 experiments and the batches of raw material are sufficient for 8 experiments, one would like to run the experiments in blocks of 8 such as the variation between batches of raw material does not affect the estimate of the main factor effects.

### ***Block interaction***

An interaction between a main effect and a block effect is called a **block interaction**.

When the design supports the interactions between the block effects and the main effects, the **Block interactions** check box, in the **Select model and design** page is active. You can select the check box if you want to add the block interactions to your model.

## **Settings**

For some designs special settings are available and with the design marked, **Settings** is active on the **Select model and design** page of the **Design wizard**.

Click **Settings** to:

- Edit the generators and/or model for fractional factorial designs of resolution III, V, and VI.
- Change the star distance for a CCC design.
- Specify the model when you have selected the split objective.

### ***Editing the model and generators for classical screening designs***

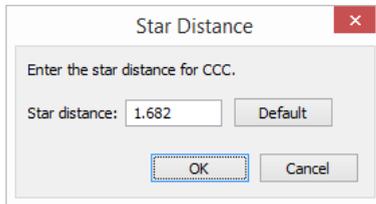
Before creating a fractional factorial design, the model and generators can be changed to better take into account user knowledge. That is, it is possible to unconfound certain model terms if desired by changing the generators.

Click **Settings** and then click **Model** to open the **Edit Model** dialog box and add interactions. Adding interactions makes the confounding pattern clear in **Generators**.

Click **Settings** and then click **Generators** to change the generators of the design.

### ***Star distance of CCC designs***

With CCC designs the star distance can be changed from the default by clicking **Settings** and then clicking **Star distance**. The default star distance is calculated as  $\sqrt{\sqrt{N}}$  where N is the number of runs of the factorial part.




---

**Note:** From 5 factors and upwards, the factorial part of the design is reduced and  $N$  of the factorial part is used in the calculation of the star distance. This means that the default star distance value is 2 for both 4 and 5 factors.

---

### **Specify the models with split objective**

When there are both process and mixture factors present the only designs available are the D-Optimal designs.

Select the split objective to specify separate models for the mixture factors and the process factors. Click **Settings**, click **Model** and the **Edit Model** dialog box is opened.

### **How high should the power be?**

High power gives us confidence that we would detect a significant model if there is one. Ultimately, the required power is determined for the study at hand, and depends on the consequences of a false negative.

Conventionally, a power that is  $1 - \beta \geq 0.8$  is typically said to be good (just as  $\alpha$  is typically set to 0.05), and is motivated by  $\beta = 4\alpha$  (Cohen, 1988). A higher power is of course better.

---

## **D-Optimal**

### **D-Optimal pages**

The D-Optimal pages are available when selecting a D-Optimal design on the **Select model and design** page in the **Design wizard**.

### **Design generation criteria section**

The **Design Generation Criteria** concerns the criteria on which the design is built.

#### **Design runs**

**Design runs** is the number of runs the D-Optimal algorithm will generate, not including the center points. You can change this number as desired. The smallest number of runs accepted is the number of terms currently included in the model.

#### **Model terms**

The number of terms currently in the model is listed after **Model terms**. This number is updated after changes in **Edit model**.

### ***Editing the model***

Click **Edit model** to edit the model. The **Edit Model** dialog opens and you can edit the model by adding or deleting terms in the specified model. This modifies the number of model terms.

---

*Note: When the investigation contains only mixture factors, **Edit model** is unavailable. The D-Optimal design is always generated from the full model specified in the design page.*

---

With investigations containing both mixture and process factors, you can only edit the process factor terms and the interactions between mixture and process factors.

### ***Potential terms***

By default MODDE includes a set of potential terms, i.e. additional terms not included in your model that might be important. The objective is to select a D-Optimal design rich enough to guard for the potential terms. If you want your design to be just sufficient for your specified model, clear the **Use potential terms** check box.

### ***Inclusions***

To use runs available from file as inclusions, click **Import** under *Inclusions*.

To edit the available inclusions, or paste/type runs to use as inclusions, click **Edit**.

If you have specified runs as inclusions in the **Design | Inclusions** window prior to entering the design wizard, the **Include in design** check box found under **Inclusions** is by default selected and the inclusions will automatically be part of the D-Optimal design. Clear this check box if you do not want the inclusions to be part of the D-Optimal design (but rather manually added at the end of the worksheet).

To add the inclusions after generating the worksheet, open the **Inclusions** on the **Design** tab, and then click **Add to worksheet**.

For more, see the Inclusions section in Chapter 6, Design.

### ***Degrees of freedom***

Number of **Degrees of freedom** of the residuals is calculated as:

Number of design runs – Model terms +1 (when you have center points)

The number of degrees of freedom recommended for D-Optimal designs in MODDE is at least five.

### ***Design alternatives section***

The **Design alternatives** section controls the number of designs generated by the D-Optimal algorithm.

### ***Design runs span***

The performance of a D-Optimal design, depends on the selected number of runs,  $N$ , and the number of terms in the model,  $p$ .

MODDE can generate several D-Optimal designs, varying the specified number of runs  $N$  and then evaluate them (G-efficiency, Condition number, Determinant) as functions of  $N$ .

In the **Design runs span** box, you can select the number of designs to generate with varying  $N$ .

If, for example, you select  $N \pm 3$  and 1 repetition, MODDE generate 7 designs ranging from  $N - 3$  to  $N + 3$ . The default is to generate 25 designs with  $N \pm 2$  and 5 repetitions.

### **Repetitions**

In the **Repetitions** box, select the number of designs you want to generate with the same number of runs,  $N$ . This will give a set of designs for each value of  $N$ .

### **Balancing the design**

When you have a qualitative factor, or when you have selected to block the design, you may want the design to have the same number of runs at each level of the qualitative factor. Thus, the design would be *balanced* with respect to the qualitative factor.

If you want a balanced design, select the qualitative factor in the **Balance on** box.

Balance on:  ▼

If you want MODDE to only select balanced designs, select the **Use balanced only** check box.

To be able to get a balanced design, the selected number of design runs must be a multiple of the number of levels of the qualitative factor. The number of design runs may be updated, if necessary, to be a multiple of the number of levels of the qualitative factor.

---

*Note: It is not always possible to generate balanced designs. When MODDE does not succeed in generating a balanced design, it issues a message. In this case, to generate a design, you must clear the **Use balanced only** check box.*

---

## **Candidate set section**

The **Candidate set** section concerns the set of design runs to select the D-Optimal design from.

### **Generating a new candidate set**

**Generate new** is by default selected the first time the D-Optimal page is opened. Click **Generate new** when you have changed the model or selected the **Use potential terms** check box and you want to generate a new set of candidate runs.

### **Using the current candidate set**

**Use the current candidate set** is available after importing or creating a candidate set.

Once the candidate set has been generated, the **Use the current candidate set** option is by default selected.

### **Editing the candidate set**

Click **Edit** to edit the candidate set. This can be done both with **Generate new** and **Use the current candidate set** selected. A spreadsheet opens with the candidate set. Make your changes and click OK to Save.

The candidate set can also be opened for editing by clicking **Candidate set** on the **Design** tab.

### **Importing a candidate set**

You can import a candidate set from a number of file types.

To import the candidate set:

1. Click **Import** found in the **Candidate set** section.
2. Find the file holding the candidate set and click **Open**.
3. The **Import Candidate Set** window opens allowing you to specify the row containing the factor names, and optionally the column holding the experiment names. Here you can exclude and include rows and columns too. The row defined as **Factor Name** in the candidate set-file must contain the factor names and they must be identical to those defined in the MODDE investigation. Including uncontrolled, filler, and constant factors is optional.

### Size of candidate set

The size of the candidate set in MODDE is by default limited to 512 000 rows when MODDE creates the candidate set for you.

You can change this limit in the **Max candidate set size** field, in the MODDE options page, opened by clicking **File | Options**.

The maximum size of the candidate set that you can create and generate a design from is limited by the amount of RAM in your computer.

### D-Optimal results

When you click **Next** on the **Change D-Optimal settings** page, MODDE generates the D-Optimal designs and displays them in the **D-Optimal results** page.

D-Optimal results. Select which design to use as your Worksheet from the generated designs.

Auto-select design by: G-efficiency

Generated designs:

No.	Runs	DF			Norm. log(D...	Condition n...
1	15	6	46.7639	8.77113	-0.298978	8.1806
2	15	6	44.4576	8.93914	-0.282178	7.93983
3	15	6	33.4403	8.70926	-0.305165	6.87079
4	15	6	49.5955	8.86226	-0.289865	6.74737
5	15	6	61.1472	8.84847	-0.291244	6.15409
6	16	7	58.7702	9.18757	-0.285363	8.39783
7	16	7	59.4399	9.24721	-0.279399	7.79584
8	16	7	54.6616	9.06635	-0.297485	8.36143
9	16	7	81.5553	9.32959	-0.271161	7.48545
1.	16	7	54.4774	9.12928	-0.291192	8.17326
1.	17	8	47.8671	9.42812	-0.287637	8.5854
1.	17	8	52.0436	9.44281	-0.286168	7.75378

Display as plot

< Back    Next >    Finish    Close    Help

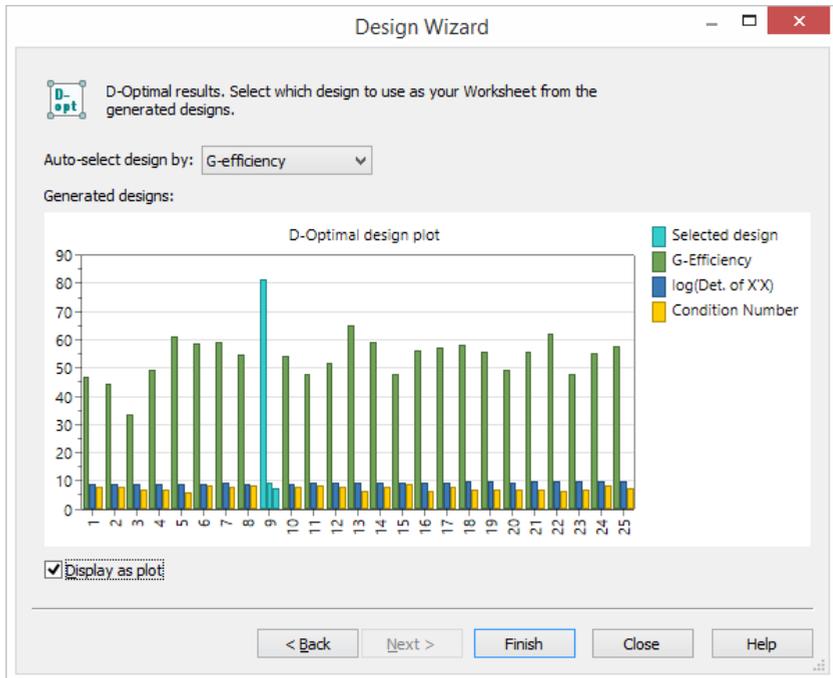
By default, the best design according to **G-efficiency** is selected. Use the **Auto-select design by** box to instead select the best design according to **Determinant** or **Condition number**. Or select another design manually by marking the design.

---

**Hint:** Sort the list according to any column by clicking its header.

---

To see the D-Optimal results as a plot, select the **Display as plot** check box.



The currently selected design is colored according to the legend.

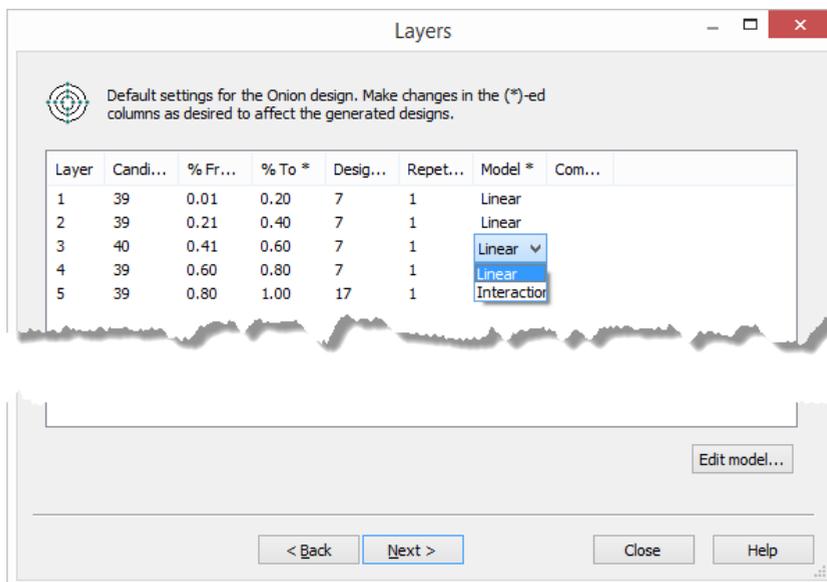
Click **Finish** to generate the worksheet. Any already existing design and worksheet will be deleted.

To regenerate a D-Optimal design, using one of the calculated ones, on the **Design** tab, click **New design | Select from already generated D-Optimal designs**.

## D-Optimal onion pages

### *Generating the design*

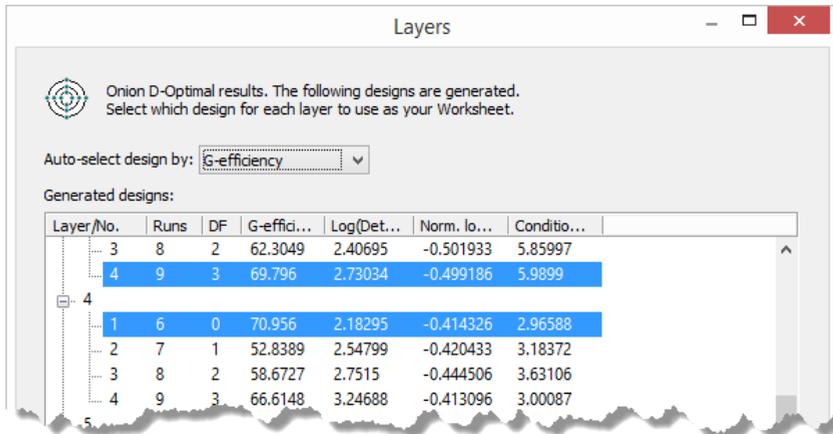
After selecting an Onion design, and clicking **Next**, the **Layers** page is displayed. This page is organized by layer – from inner (first layer) to outer (last layer).



The following is listed:

1. The number of the **Layer** (starting from inside).
2. The number of **Candidate Runs** in the layer for imported candidate sets.
3. The span of the layer defined by its **% From** (Percentile) and **% To** distance to the center of the multivariate space. You can change the span of a layer as long as the number of runs in the candidate set remains one and half times larger than the number of design runs in that layer. Overlapping span between layers is not allowed. When you change the span of the layers, MODDE updates the number of candidate set runs in each layer. If the span of the layers overlap, or the number of runs in the candidate set is not large enough, the layer is colored in red and a message indicating the problem is displayed. You must fix the problem before clicking **Next**.
4. The number of **Design Runs** in each layer. You can change the number of design runs. The number of desired runs must be at least equal to the number of terms in the model. The recommended number of runs includes 3 degrees of freedom for the outer layer and 1 degree of freedom for the rest of the layers.
5. Select the number of D-Optimal designs, in each layer, you want to generate with the same number of runs. The default number for **Repetition** is 1.
6. The **Model** for each layer. You can change the model in each layer. Click the model you want to change and select another model or mark the layer and click **Edit model** to customize the model. After editing a model, MODDE writes **Model edited** in the **Comment** column. MODDE updates the design runs according to the new selected model; default *Linear* for the inner layers and *Interaction* for the outer layer.

When you click **Finish**, MODDE generates several D-Optimal designs in each layer varying the number of runs by plus and minus 2, and displays the Onion D-Optimal results page.



The table on this page displays, for every layer, the generated designs statistics. By default, in this table, the designs with the highest **G-efficiency** are selected.

You can select a different design in a given layer by marking it in the list or using the **Auto-select design by** box and selecting a different criterion.

## Onion on the Design tab

### Generate

With Onion D-Optimal designs, you *cannot* generate a new set of D-Optimal designs by clicking **New design** in the **D-Optimal** group on the **Design** tab.

If you want to generate a new set of Onion D-Optimal designs, click **Design | Objective**. The worksheet will be deleted and you can follow the wizard to generate a new design and worksheet.

### Candidate set

To display the candidate set, on the **Design** tab click **Candidate set**.

The columns in the spreadsheet are as follows when the candidate set was imported:

1. **Design Run** number. Corresponds with the experiment number in the worksheet.
2. The **Exp Name** (experiment name) when available.
3. The **Layer** number (the innermost layer = 1).
4. The distance to the center of the multivariate space in **Percent**.
5. The design variables.

Candidate Set						
1	2	3	4	5	6	7
Design Run	Exp Name	Layer	Percent	M1.11	M1.12	M1.13
center	TETRACAINE	-	1.00	-1.33018	-0.251272	0.67179
1	4-AMINO-N-[4,5-DIMETHYL-2-OXAZOLYL]BENZENESULFONAMIDE	1	1.00	1.31127	-0.557021	-0.51905
2	chlorpropamide	1	2.00	0.129045	-0.61343	-1.3065
3	ALBENDAZOLE SULFONE	1	2.00	1.64849	-0.263542	0.2555
4	Procaine (HCl)	1	3.00	0.214817	-1.15308	1.0716
5	6-Benzylaminopurine	1	3.00	-0.140418	-1.02281	1.0315
6	SULFAMETHOXAZOLE	1	4.00	1.77938	-1.04987	-0.6955
7	lormetazepam	1	4.00	-1.00479	0.366163	0.18931
8	Nabumetone	1	5.00	-2.16221	-1.17493	-0.04242
9	1 FENBENDAZOLE SULFONE	1	5.00	1.04033	0.526712	0.33211

When the candidate set is generated by MODDE the additional columns **Layer** and **Percent** are unavailable. The layer belonging is displayed in the onion plots.

## Reduced combinatorial

### Design generation options

The **Design generation criteria** section concerns the criteria on which the design is built while the **Design alternatives** section controls the number of designs generated what to balance on.

#### *Design runs*

**Design runs** is the number of runs the algorithm will generate, not including the center points. You can change this number as desired. The smallest number of runs accepted is the number of terms currently included in the model.

#### *Replicated runs*

The number of runs that should be replicated. The replicated runs are used to get an estimation of the replicate error when lacking center points.

#### *Model terms*

The number of terms currently in the model is listed after **Model terms**. This number is updated after changes in **Edit model**.

#### *Editing the model*

Click the **Edit model** button to edit the model. The edit model dialog opens and you can edit the model by adding or deleting terms in the specified model. This modifies the number of model terms.

#### *Degrees of freedom*

Number of **Degrees of freedom** of the residuals is calculated as:

Number of design runs – Model terms +1 (when you have center points)

The number of degrees of freedom recommended for D-Optimal and reduced combinatorial designs in MODDE is at least five.

#### *Design runs span*

The performance of the design depends on the selected number of runs,  $N$ , and the number of terms in the model,  $p$ .

MODDE can generate several designs, varying the specified number of runs  $N$  and then evaluate them (G-efficiency, Condition number, Determinant) as functions of  $N$ .

In the **Design runs span** box, you can select the number of designs to generate with varying  $N$ .

If, for example, you select  $N \pm 3$ , MODDE generate 7 designs ranging from  $N - 3$  to  $N + 3$ . The default is to generate 5 designs with  $N \pm 2$ .

### Balancing the design

When you have more than one qualitative or multilevel factor, can select for the design to have the same number of runs at each level of one of the qualitative/multilevel factors. Thus, the design would be *balanced* with respect to the qualitative or multilevel factor.

If you want to balance the design on another factor than the default, select the desired factor in the **Balance on** box.

Balance on:

To be able to get a balanced design, the selected number of design runs must be a multiple of the number of levels of the qualitative/multilevel factor. The number of design runs may be updated, if necessary, to be a multiple of the number of levels of the factor.

### Reduced combinatorial design selection page

When you click **Next** on the **Change reduced combinatorial settings** page, MODDE generates the designs and displays them allowing you to select which design to go ahead and generate.

Design Wizard

Select which design to use as your Worksheet from the generated designs  
Balanced designs have a "B" after the number of runs.

Auto-select design by:

Generated designs:

No.	Runs	DF	G-effici...	Log(Det. of ...	Norm. log(D...	Condition ...
1	10	4	47.014	8.29186	-0.192626	3.53032
2	11	5	40.6002	8.63812	-0.186336	3.16764
3	12:B	6	52.6959	9.01711	-0.17419	3.17717
4	13	7	47.6639	9.28203	-0.172784	3.33085
5	14	8	54.2414	9.58816	-0.165098	3.02798

By default, the best design according to **G-efficiency** is selected. Use the **Auto-select design by** box to instead select the best design according to **Determinant** or **Condition number**. Or select another design manually by marking the design.

---

**Hint:** Sort the list according to any column by clicking its header.

---

To see the results as a plot, select the **Display as plot** check box, see the D-Optimal results subsection earlier in this chapter.

Click **Finish** to generate the worksheet. Any already existing design and worksheet will be deleted.



# 04-File

---

## Introduction

The **File** tab includes functions that deal with your entire MODDE investigation as opposed to the data itself. Here you can save, print, and send your investigation as well as change settings and options.

The functions available are grouped into 10 groups

- **Info** - Provides access to Protect investigation, the report, and displays investigation properties
  - **Protect investigation** - Controls access and changes that can be done to this investigation.
  - **Report** - Generate a report for this investigation using the default template or any custom template.
- **New** - provides access to options for creating a new design or importing an existing design.
  - **Experimental design** - Start the classical experimental design setup from here.
  - **Import external design** - Import a design saved in another file format.
  - **Complement design** - Add new experiments to resolve interactions or non-linearities in the current design.
  - **Generalized subset designs** - Create a sequence of complementary subset design sets.
  - **Stability testing design** - Create a design that spans over time to investigate the stability of the product.
  - **Design from candidate set** - Import a candidate set and create a design from a D-Optimal selection.
  - **Design from scores** - Import scores (factors) from a SIMCA usp file and create a design from a D-Optimal selection.
  - **RED-MUP** - Special designs created for 96, 384, or 1536 well plates.
- **Open** - Open an existing investigation.
- **Save** - Save the current investigation.

- **Save as** - Save the current investigation with a new name and/or at a new location.
- **Print** - Print the active window.
- **Share** - Share the data in the current investigation.
  - Send a copy of the investigation by email as an attachment.
  - Export the worksheet to SIMCA.
- **Close** - Close the current investigation.
- **Help** - Get help with using and activating MODDE.
  - **Manage license** - Activate MODDE over the internet and other options.
  - **View help** - Get help using MODDE.
  - **Sartorius Stedim Data Analytics** - Open Sartorius Stedim Data Analytics' website in a browser.
  - **Knowledge base** - Open Sartorius Stedim Data Analytics' website in a browser, with the Knowledge base page active.
  - **About us**
  - **Webshop**
- **Options** - Change options and customize MODDE.

---

## Info

The **Info** section on the **File** tab provides access to two main features, **Protect investigation** and **Report** as well as giving a quick summary of the investigation **Properties** and displaying the path to the current investigation.

### **Properties**

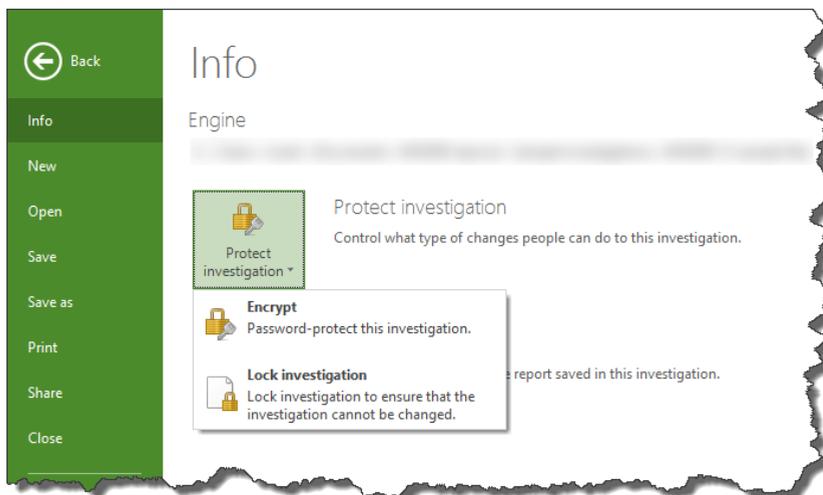
Under **Properties** information about the current investigation is listed;

- Size
- Creation date
- Modified date
- Objective
- Process model
- Mixture model
- Design
- Runs in design
- Center points
- Replicates
- Factors
- Responses

- Number of runs.

## Protect investigation

To protect an investigation, you can either lock it so that changes cannot be made, or encrypt it with a password.



## Encrypt investigation

Enter a password and the investigation is encrypted and password protected. This investigation can now be opened only with the selected password.

To protect your investigation,

1. On the **File** tab, click **Info**,
2. Click **Protect investigation**,
3. Click **Encrypt**,
4. In the **Encrypt Investigation** dialog box, enter a password to use for the encryption and confirm it,
5. Click **OK**.



### Remove encryption

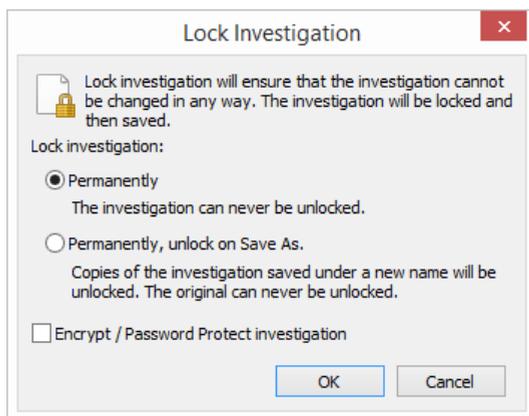
If the investigation is already encrypted, the encrypt option is replaced by **Remove encryption**.

To remove the encryption,

1. On the **File** tab, click **Info**,
2. Click **Protect investigation**,
3. Click **Remove encryption**,
4. Enter the password and click **OK**.

### Lock investigation

Locked investigations are automatically fitted when opened. Any plot or list can be displayed, but changes cannot be made to the investigation. The investigation becomes 'Read only' with the exception of the prediction spreadsheet. If you click **Permanently, unlock on Save As**, a copy of the investigation can be unlocked by clicking **Save as** on the **File** tab. When **Permanently** is selected the investigation cannot be unlocked.



It is possible to also encrypt and password protect the investigation by selecting the **Encrypt / Password Protect investigation** check box. The investigation is then encrypted. This check box is optional; select it only if you want the investigation to be password protected.

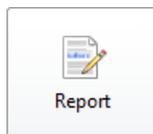
---

*Note: The lock cannot be removed.*

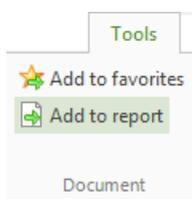
---

### Report

MODDE has an automatic report generator. To create a new report, on the **File** tab, click **Info**, then click **Report**.



You can also right-click a plot or list and then click **Add to report**. To add the currently active plot or list to the report, click **Add to report** on the **Tools** tab or press Ctrl+R.




---

*Note: When a report was saved with the investigation, MODDE automatically opens that report.*

---

The report can use the MODDE default template or any template previously saved. All formatting functionality is available for writing the text. Plots and lists can be added to the report at any time, as placeholders or actual plots and lists.

A placeholder tells MODDE the desired item to fill from the current investigation when you click **Update report** on the **Home** tab of the Report. If you add plots and lists as placeholders and save the template, you can generate a report in the desired format, for any investigation, by selecting the saved template and clicking **Update report**.



For details, see Chapter 14, Report.

---

## New

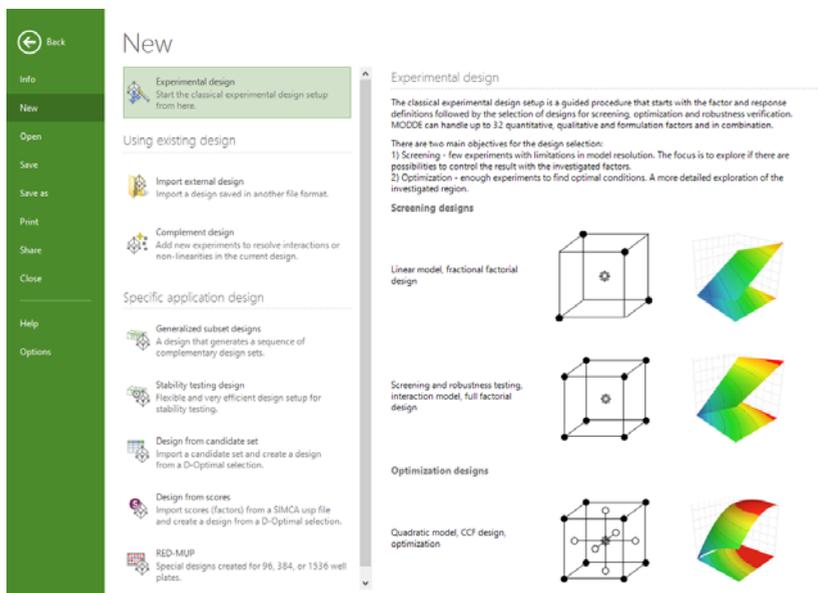
To create a new investigation on the **File** tab, click **New**.

This provides access to options for creating a new design or importing an existing design.

### Experimental design

On the **File** tab, click **New**, then click **Experimental design**.

Start the classical experimental design setup from here. A wizard will guide you through all steps for setting up an experimental design - traditional designs as well as super saturated and D-Optimal designs. The wizard will always make a design recommendation.



## Design wizard

The **Design wizard** opens when you start a new experimental design. This wizard guides you from the start of the investigation to the generation of the worksheet. The **Design wizard** can also be opened on the **Home** tab, in the **Quick start** group, click **Design wizard**. Exit the wizard at any time by clicking **Finish**.

## Using existing design

To create a new investigation from an existing design, on the **File** tab, click **New**, then click either **Import external design** or **Complement design**.

### Import external design

Import a design saved in another file format.



External data can easily be imported to MODDE for analysis and optimization, as follows:

1. Click **File** | **New**.
2. Click **Import external design**.
3. Browse and select the file to import.
4. Click **Next** and specify what is what in the **Data Specification** dialog.

### Complement design

The current investigation can be complemented with new experiments to resolve interactions or non-linearities.

To add new experiments to the current design, on the **File** tab, click **New**, then click **Complement design**.

A wizard will guide you through classical fold over complementation or specific requests.



Use **Complement design** when you want to:

- Estimate separately a set of terms (interactions, or main terms and interactions) that were confounded in a Resolution III or IV fractional factorial design.
- Complement a screening design to an RSM design supporting the full quadratic model.
- Complement a screening design to estimate selected curvature effects.
- Add additional experimental runs to improve the quality (i.e. the condition number or G-efficiency) of an existing set of experiments.
- Use already performed experiments in a Doehlert design to set up a new Doehlert; moving the center of the design or adding a factor.
- Use already performed experiments in a Super-Saturated Plackett Burman design to add experiments resulting in a regular Plackett Burman design.

### Available complementing methods

You can complement your design using the following complementing methods:

**Fold over:** With screening fractional factorial designs of resolution III and N (number of runs) equal to 8 or 16, it is usually recommended to complement by fold over. For more, see the Fold over subsection later in this chapter.

**Estimate squares of selected factors in factorial designs:** If the screening design indicates the presence of curvature, you may want to estimate the square terms in selected factors or you may want to upgrade your design to a full RSM in these selected factors. For more, see the Estimate squares terms subsection later in this chapter.

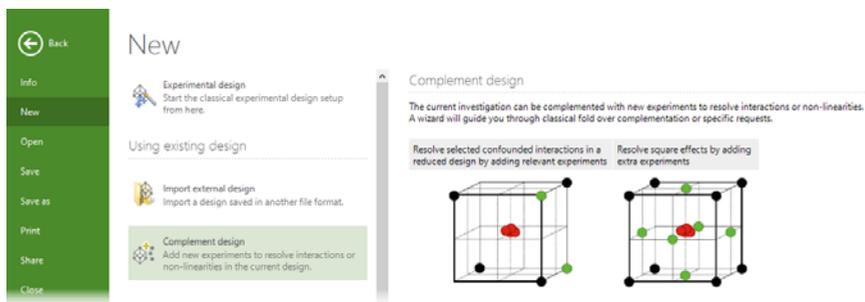
**D-Optimal:** This option should only be selected if it is not possible to use a different complementing method. If you want to upgrade the design to support a customized model, or if your investigation contains mixture factors, select D-Optimal. For more, see the Complementing a design with D-Optimal subsection later in this chapter.

### Design specific complementing methods

**Doehlert:** When you have performed a Doehlert design and want to move the design center or add a factor. For more see the Complementing Doehlert designs subsection later in this chapter.

**Screening to RSM Rechtschaffner:** When you have performed a screening Rechtschaffner design. Complemented by adding the star points resulting in a 3-level Rechtschaffner.

**Plackett Burman Super-Saturated to Plackett Burman:** When you have performed a PBSS. Complemented with runs adding up to a regular Plackett Burman design.



### ***Fold over***

When you choose to complement your design with fold over MODDE makes a new investigation consisting of the design of the current investigation plus its fold over (complement). The fold over design has as many experimental runs as the original design.

Fold over designs are available for fractional factorial design of resolution III or IV and Plackett Burman designs.

With the complete design (original + fold over), all main effects are clear from 2 factors interactions. With resolution III and IV designs MODDE automatically adds a block factor. You may remove the block factor from the model in **Edit model**.

To fold over your design:

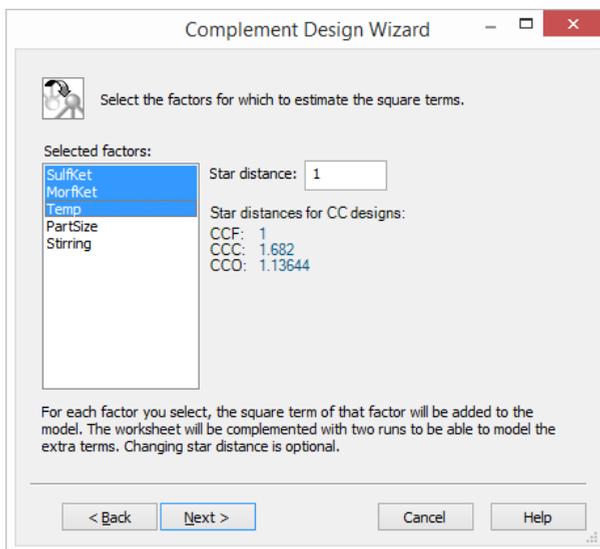
1. Click **File | New | Complement design**.
2. Click **Fold over** and then click **Next**.
3. Enter the name and location of the new investigation. It is recommended to add an additional center point to detect a shift in the mean.
4. Click **Finish** and the new investigation opens.

### ***Estimate squares terms in a screening design***

For this complementing method to work well the original design should be of resolution V or the collapsed design in the selected factors has to be a full factorial.

On the **File** tab, click **New**, and then click **Complement design**:

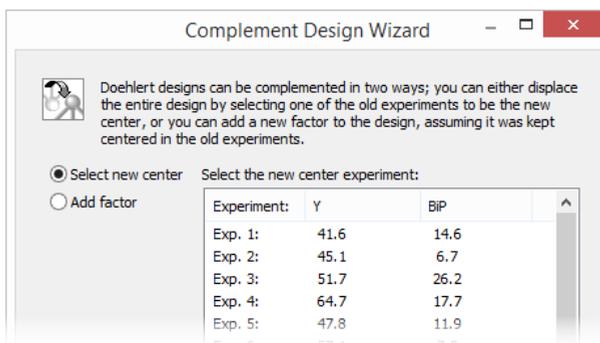
1. Click **Estimate square terms in a screening design** and click **Next**.
2. Select the factors for which to estimate the square terms. Note that the calculated star distances for CCC and CCO are displayed.
3. Optionally change the star distance. If the star distance is not changed, the design is complemented to a CCF design by adding the face center runs. To change, click the desired star distance to automatically enter it in the **Star distance** field. Click **Next**.
4. Enter the name, location of the new investigation and number of additional center points.
5. The model has been updated with the squares of the selected factors. The unselected factors are set on their averages in the worksheet.



### **Complement Doehlert**

Doehlert designs can be complemented by expanding the design region or adding a factor.

To complement a Doehlert design, click **Doehlert** in the **Complement Design Wizard** and click **Next**.



As the dialog suggests you can complement by:

- Leaving the default **Select new center** selected and selecting one of the experiments of the design as the center of a new Doehlert design.
- Selecting **Add factor** and typing the name of the new factor and typing the **Center** and **Range** in the respective boxes. The value you type in **Center** is the value that will be entered in the design for the already performed experiments, for the new factor.



Select how to complement, click **Next**, enter the name, location of the new investigation and number of additional center points, and then click **Finish** to generate the new investigation.

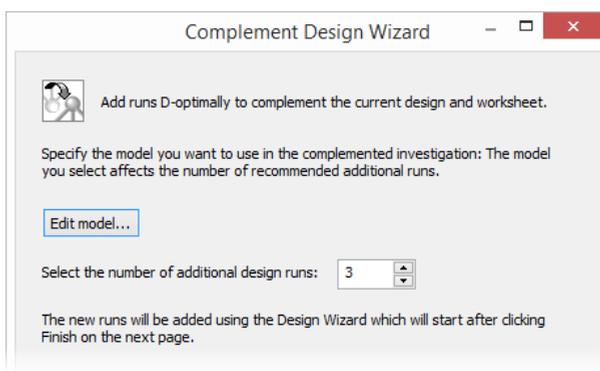
### ***D-Optimal***

Complementing a design D-optimally is the most flexible way of complementing a design.

#### **Process factors only**

On the **File** tab, click **New**, and then click **Complement design**:

1. With **D-Optimal** marked, click **Next**.
2. Click **Edit model...** to open the **Edit Model** dialog box and add model terms. Click **OK** to return to the Complement Design Wizard.
3. MODDE recommends the number of additional new design runs based on the specified model, to ensure the proper degrees of freedom.
4. Click **Next**.
5. Enter the name and location of the new investigation and click **Finish**. The D-Optimal pages in the Design wizard guide you in generating the new investigation. Here the original design runs in the selected factors are used as inclusions and the displayed **Design runs** in the D-Optimal page includes the inclusions. The additional new runs are selected D-optimally to support the selected model.



### Mixture factors

Complementing the design D-optimally with mixture factors, only allows changing the model in **Settings | Model** available after entering the name and location of the new investigation.

The displayed number of runs includes the original design runs. Click **Next**, the wizard guides you in generating the new investigation.

### Mixture and process factors

Complementing a design D-optimally with mixture and process factors, only allows changing the model in **Settings | Model** available after entering the name and location of the new investigation. The objective is set to **Split objective**.

The displayed number of runs includes the original design runs as inclusions. Click **Next** and the wizard guides you in generating the new investigation.

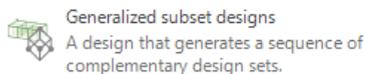
## Specific application design

MODDE provides options to create specific application designs such as Generalized subset designs, **Stability testing design**, **Design from candidate set**, **Design from scores**, and **RED-MUP**.

To create a specific application design, on the **File** tab, click **New**, then click the desired option.

### Generalized subset designs

In Generalized Subset Designs, GSD, a sequence of combinatorial design sets are used to evaluate the effect of the factors and settings. The purpose is to use one or several subsets of all possible combinations in a sequence and stop when a sufficient amount of information has been obtained.



In Generalized subset designs all combinations are covered using generalized fractional factorial subset designs. This design strategy is called complementary combinatorial design.

For example, 3 factors ( $n$ ) with 2+3+4 levels = 24 possible combinations/experiments can be distributed over 2 subsets as 12+12 experiments or over 3 subsets as 8+8+8 experiments.

The number of possible combinations  $k$  can be divided into  $p$  subsets,  $k/p$ , where  $p$  is an integer. Each subset  $m_i$  will contain a unique set of combinations. One constraint is that each subset  $m_i$  shall be as balanced as possible and the best possible representation of  $k$ .

For more information of the practical handling in MODDE, see the Generalized subset designs appendix.

### Algorithm short version;

1. Choose an  $n$ -dimensional design region with  $k$  runs in its candidate set and a design with  $m$  runs:  $m = k/p$ .
2. Decompose factors into  $p$  sets.
3. Generate the matrix  $M = 2 - (p, n, p^{n-3})$  OA (orthogonal array) from Latin Squares.

4. Reduce  $M$  to all rows mapping to non-empty sets.
5. Map each row of  $M$  to the decomposed factor sets and make a full factorization of the elements in the active sets:  $M_i$ .
6. The final design  $D$  is given by concatenating the mappings  $M_i$  from step 5.

### **Stability testing design**

In stability testing Generalized subset designs are used to evaluate all combinations of factors and settings over time. The purpose is to ensure that none of the quality attributes for the stability testing will exceed the specification during the products specified shelf life.



In a typical stability test all combinations are tested over several time points (*i.e.* 0, 3, 6, 9, 12, ... months). A way to reduce the total number of tests is to distribute subsets of all possible combinations to specific time points stipulating that all combinations are tested over a specified period of time. This design strategy is called Complementary combinatorial design.

For example, 3 factors ( $n$ ) with 2+3+4 levels = 24 possible combinations/experiments can be distributed over 2 time points as 12+12 experiments or over 3 time points as 8+8+8 experiments.

The number of possible combinations  $k$  can be divided into  $p$  subsets,  $k/p$ , where  $p$  is an integer. Each subset  $m_i$  will contain a unique set of combinations. One constraint is that each subset  $m_i$  shall be as balanced as possible and the best possible representation of  $k$ .

For more information of the practical handling in MODDE, see the Stability testing design section in the Generalized subset designs appendix.

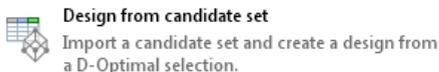
### **Algorithm short version;**

1. Choose an  $n$ -dimensional design region with  $k$  runs in its candidate set and a design with  $m$  runs:  $m = k/p$ .
2. Decompose factors into  $p$  sets.
3. Generate the matrix  $M = 2 - (p, n, p^{n-3})$  OA (orthogonal array) from Latin Squares.
4. Reduce  $M$  to all rows mapping to non-empty sets.
5. Map each row of  $M$  to the decomposed factor sets and make a full factorization of the elements in the active sets:  $M_i$ .
6. The final design  $D$  is given by concatenating the mappings  $M_i$  from step 5.

### **Design from candidate set**

Import a candidate set and create a designed selection of observations. The selection can be done using layers (Onion) or D-optimally.

To create a design based on a candidate set, on the **File** tab, click **New**, then click **Design from candidate set**.



### Onion design in regular factors with imported candidate set

When you want create an onion design from an imported candidate set other than scores from SIMCA:

1. On the **File** tab, click **New**, then click **Design from candidate set**.
2. Click **Browse** to find the file holding the candidate set. Many file types are supported.
3. Click **Next** to open the **Import candidate set** dialog box specifying factor names, experiment names, and data.
4. Click **OK** to open the factors page of the **Design wizard**. When the candidate set has been imported in this way you cannot modify or add factors using the **Design wizard**.
5. To place constraints on the factors, click the **Place constraints on the experimental region** check box and click **Next** to open the constraints page of the **Design wizard**.

---

*Note: Only regular factors are imported here. If your candidate set contains qualitative or formulation factors you have to enter the factors and settings in the **Factor Definition** dialog box and import the candidate set from the **D-Optimal** page instead.*

---

The designs available here are **Onion** and **D-Optimal** designs. Onion is only available when there are enough experiments in comparison with the number of factors.

### Design from scores

This is a typical QSAR application where a subset of molecules is selected by design. Molecules described by many variables are compressed to score vectors in SIMCA. These score vectors can then be imported directly from the SIMCA usp file.

To create a design based on imported scores from a SIMCA file, on the **File** tab, click **New**, then click **Design from scores**.



### Multivariate designs

To be able to import scores from a SIMCA project, SIMCA 13 or later needs to be installed. The only designs available are *D-Optimal* and *Onion*.

When you want to create a design using the scores from a SIMCA project as factors:

1. On the **File** tab, click **New**, then click **Design from scores**.
2. Click **Browse** to select the SIMCA project.
3. Select the model from the **Model** box and click **Next** to import the factors (scores) from SIMCA. The score variables with all observations (rows) are then automatically loaded from the SIMCA file and define the candidate set.
4. Click **Next** to continue through the design wizard and the creation of the D-Optimal or onion design.

---

*Note: SIMCA opens automatically and remains open in the background while focus returns to MODDE.*

---

### **RED-MUP**

Rectangular experimental design for multi unit platforms - an efficient and time-saving approach for application of DOE to 96, 384 and 1536 well plates.



#### **RED-MUP**

Special designs created for 96, 384, or 1536 well plates.

RED-MUPs are designs available for 96 (8x12), 384 (16x24), and 1536 (32x48) runs. The designs are built from sub-designs.

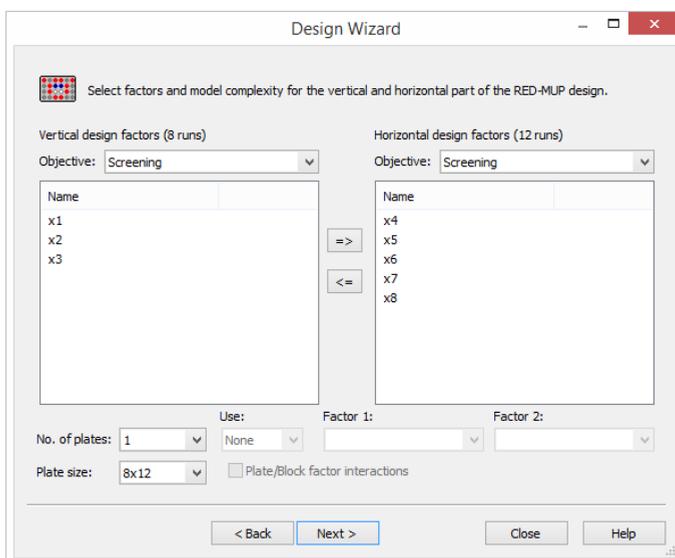
To create a **RED-MUP**:

1. On the **File** tab, click **New**, then click **RED-MUP**.
2. Define all factors for the two sub-designs and then click **Next**.
3. Define the responses and then click **Next**.
4. Select the objective for both the vertical and the horizontal designs: **Screening** or **Optimization (RSM)**.
5. Leave the factors that should be included in the vertical design with fewer runs to the left.
6. Move the factors that should be included in the horizontal design to the right.
7. On this page you can also select the number of plates used, and if applicable the plate factors that contain plate information.
8. Select the desired plate size in the **Plate size** box and optionally select the **Plate/Block factor interactions** check box.
9. Click **Next** to select the vertical design then click **Next** again to select the horizontal design. MODDE adds center points when the selected design does not fill up the plate size.

---

*Note: Some special RED-MUP designs, that aim to make better use of the plate, are available for the 96 well plates (8 x 12).*

---



After stepping through the Design Wizard the special RED-MUP worksheet is created.

### **RED-MUP Worksheet**

The RED-MUP Worksheet has tabs to display;

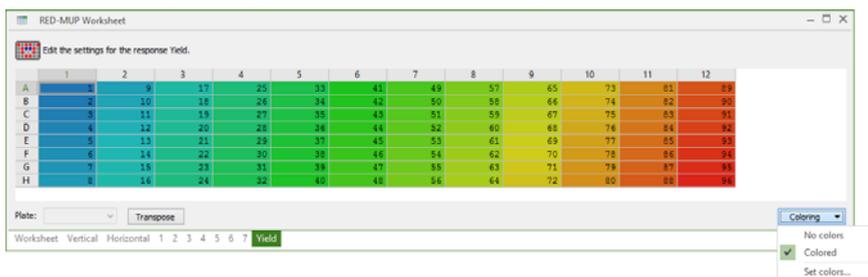
- **Vertical** spreadsheet displaying the vertical design.  
The spreadsheet has as many rows as specified by the first number of the plate size and the rows are indexed alphabetically A, B, C. For the 8x12 plate this means A-H, for the 16x24 plate A-P, etc. The design index is a runorder type column that can be changed and sorted to manipulate the setup on the factor and response tabs.
- **Horizontal** spreadsheet displaying the horizontal design.  
The spreadsheet has as many rows as specified by the second number of the plate size and the rows are indexed numerically 1, 2, 3. For the 8x12 plate this means 1-12, for the 16x24 plate 1-24, etc. The design index is a runorder type column that can be changed and sorted to manipulate the setup on the factor and response tabs.
- Factor spreadsheets displaying the individual factor settings for all combinations of vertical and horizontal settings.
- Response spreadsheets displaying the individual response values (once filled in) for all combinations of vertical and horizontal settings. This is where you fill in your response values.
- **Worksheet** which is the compilation of all factor and response spreadsheets.

If there are settings in the vertical or horizontal setups that should not, or cannot, be performed, you can delete in the Vertical or Horizontal spreadsheets and all the other RED-MUP worksheets are automatically updated to reflect the change. Unused cells are grayed out and also excluded from the main worksheet.

### **Coloring**

The RED-MUP response spreadsheets are by default colored according to the value range for each response.

- To remove the coloring, click **Coloring | No colors**.
- To change the range, click **Coloring | Set colors** and select other colors by clicking the Begin color and/or End color maps.



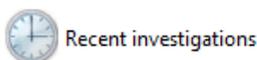
## Open

To open a MODDE investigation, on the **File** tab, click **Open**.

You can also open an existing investigation by clicking **Open**  on the Quick Access Toolbar or using the Windows keyboard shortcut Ctrl+O.

## Recent investigations

The **Recent investigations** list shows recently opened investigations.



Click the pin icon  to the right of the investigation name to pin the investigation to the top of the list. Click the name of the investigation in the list to open any recent investigation.

## Recent folders

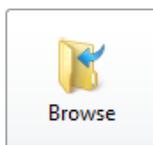
The **Recent folders** list shows folders where you have recently opened or saved an investigation.



Click the pin icon  to the right of the folder's name to pin a folder to the top of the **Recent folders** list. Click the folder name in the list to open any recently opened folder.

## Browse

To browse for existing MODDE investigations, click **Browse**.



---

## Save

To save the current investigation, on the **File** tab, click **Save**. You can also use the Windows keyboard shortcut Ctrl+S or click the save icon  on the Quick Access Toolbar at the top of the MODDE window.

---

## Save as

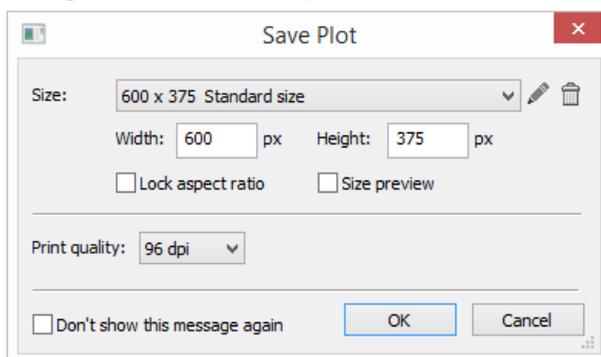
To save the current investigation in a different location or with a different name, click **Save as** on the **File** tab, then click one of the folders in the **Recent folders** list or **Browse**. In the **Save As** dialog box that opens, enter the new name and location of your investigation. MODDE switches to the "new" investigation.

## Save plot as

An individual plot can be saved in a variety of formats.

To save a plot, right-click the plot, then click **Save as....**

This opens the **Save Plot** dialog box.



For details about the **Save Plot** dialog, see the Save or Copy a plot section later in this chapter.

## Save list as

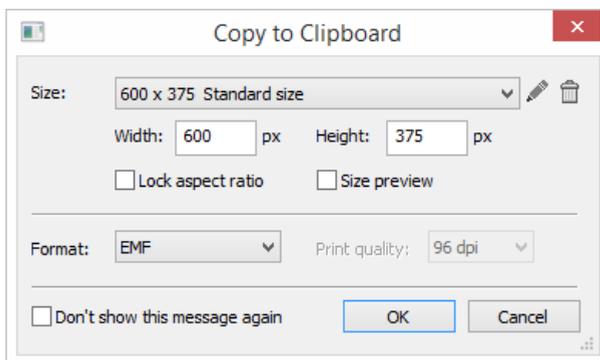
To save a list,

1. Right-click the list.
2. Click **Save as....**
3. Select location, file name, and file type.
4. Click **Save**.

Lists can be saved as either webpages (htm/html) or as text files (txt).

## Save or Copy a plot

The **Save Plot** and **Copy to Clipboard** dialog boxes are very similar, merely the presence/absence of the **Format** drop down differentiates them. When copying, the plot format can be selected in this dialog box while, for save, it can be selected in the **Save As** dialog which opens after clicking **OK**.




---

*Note: For plots with constants, the constants are copied, saved and printed if the property pane is visible.*

---

### **Size**

The size of the plot is defined by the values in the **Width** and **Height** fields. These fields are automatically updated when switching between the predefined sizes in the **Size** box. Likewise, the content of the **Size** box is updated when changing the width and height.

---

*Note: To keep the aspect ratio of the plot while customizing the size, select the **Lock aspect ratio** check box before changing.*

---

The available options in the **Size** drop down box are,

- *Original size* - the current plot size

Suitable for documentation

- *600 x 375 Standard size*
- *300 x 300 Square, fits two side by side*
- *600 x 600 Square*

Suitable for presentations

- *755 x 465 Fits one plot per slide*
- *755 x 270 Fits two above and below*
- *370 x 465 Fits two side by side*

Custom sizes

- *Add custom size* - opens the **Add Custom Size dialog** box with the current plot size predefined. Clicking **OK** adds the specified size under the **Custom sizes** header.
- *Delete customizations* - restores the list to its original content.

### **Edit and delete**

To edit the current size, click the **Edit current size** pencil to the right of the **Size** drop down.

To delete the current size, click the **Delete current size** garbage bin.

## Size preview

Selecting the **Size preview** check box displays the plot exactly as it will be saved or copied. This feature allows you to verify that the layout and text formats harmonize with the selected save/copy size before inserting/pasting it.

## Print quality and plot format

The default print quality is 96 dpi. This is sufficient for the web and presentations, but not for high quality print where 300 dpi is recommended.

The plot formats available are Bitmap (BMP), EMF (only 2D), PNG, JPG, Encapsulated PostScript (EPS), and SVG (only 3D and only when saving).

---

*Note: Using EPS with high dpi creates a very large plot file.*

---

## Restore to original list of sizes

After making changes to the **Size** list by adding custom sizes or deleting original sizes the list can be restored to its original content by, at the bottom of the **Size** list, clicking **Delete customizations**.

---

## Print

To print, preview or change the print setup click **File | Print**.

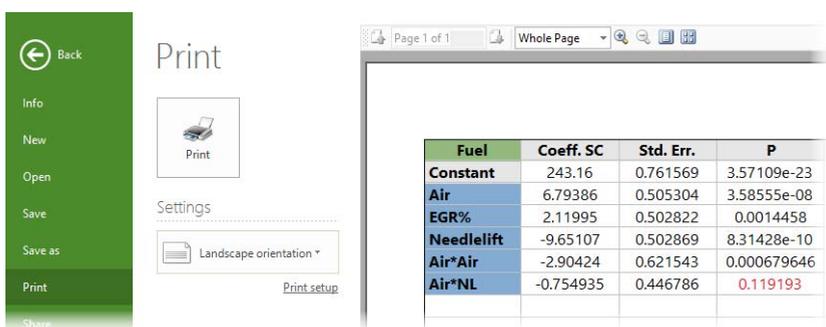
When a list is the active window, the **Print in color** check box is available here. To print your list without background coloring (whether in color or in gray scale), clear this check box.

For plots with constants, such as the Contour plot, the constants are copied, saved and printed if the property pane is visible.

---

*Note: Plots are printed as viewed on the screen except when printing to a pdf writer.*

---



Print is available for the active plot or list. MODDE provides a variety of ways to print,

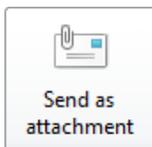
- Ctrl+P, the Windows keyboard shortcut for print,
- Right-click the desired plot or list and click **Print**,
- On the **File** tab, click **Print**.

---

## Share

### Send as attachment

On the **File** tab, click **Share**, then click **Send as attachment** to send a copy of the investigation by email as an attachment.



### Export to SIMCA

To create a SIMCA project using the current worksheet, click **File | Share** and then click **Export to SIMCA** to open SIMCA with the newly created project active. The worksheet is exported and SIMCA recognizes factors as x-variables and responses as y-variables. Models, transformations, cross-validation groups are not exported. Reference mixture is not exported nor supported.



---

*Note: This feature requires an active SIMCA 13 or later.*

---

---

## Close

On the **File** tab, click **Close** to close the currently open investigation.

---

## Help

The **Help** page on the **File** tab provides various resources for support using MODDE, handling licenses, and about the program itself.

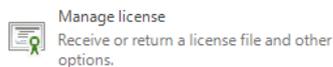
The screenshot shows the MODDE software interface. On the left is a green sidebar with a 'Back' button and menu items: Info, New, Open, Save, Save as, Print, Share, Close, Help, and Options. The main window is titled 'Activate' and contains several sections: 'Manage license' (Receive or return a license file and other options), 'Support' (View help, Sartorius Stedim Data Analytics, Knowledge base, About us), and 'Shop' (Webshop). On the right, there is a 'MODDE Pro' header, 'About MODDE Pro' section (Version 12, Copyright © Sartorius Stedim Data Analytics AB), a 'License' section with fields for License type (MODDE Pro), Product version, Expiry date, Running on Windows Terminal Services, Running on virtual machine, and HostID. At the bottom, there is a disclaimer about software licenses and a warning about unauthorized reproduction.

## Activate and manage license

To open options related to activating MODDE and managing your license,

1. Click the **File** tab.
2. Click **Help**.
3. Click **Manage license**.

### Activate



Follow the instructions in the dialog boxes.

Note that deactivation can only be done for a license activated using an activation key.

## View help

Get help using MODDE. Opens the **MODDE Help** window. MODDE's help is based on this user guide.

## Sartorius Stedim Data Analytics

Opens Sartorius Stedim Data Analytics' website in a browser.

## Knowledge base

Opens Sartorius Stedim Data Analytics' website in a browser, with the Knowledge base page active.

## About us

Sartorius Stedim Data Analytics is a world leader of Multivariate Technology. With our Umetrics Suite we provide software for Design of Experiments and Multivariate Data Analysis. Tools that transform your data into information, guiding you to confident decisions and helping you understand complex products and processes.

We offer complete solutions for off-line as well as on-line applications for continuous and batch processes. All supported by training and consulting services to ensure you get the most of your data.

## Webshop

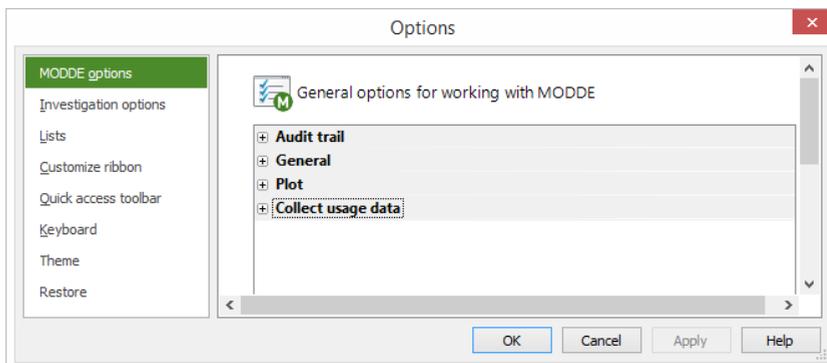
Opens Sartorius Stedim Data Analytics' website in a browser, displaying the items available in the Webshop.

---

## Options

On the **File** tab, click **Options** to access the options for a number of sections of MODDE.

The **Options** dialog box comprises all general options as well as investigation specific options and other customizations, such as Theme (not described further here).



## MODDE options

The **MODDE options** page is divided into the parts **Audit trail**, **General**, and **Plot**.

 General options for working with MODDE

[-] <b>Audit trail</b>	
Enable the audit trail for new investigations	No
[-] <b>General</b>	
Automatically display output window	Yes
Automatic fit	Yes
Close open investigation when opening a new investigation	Yes
Factor axes in back transformed units	Yes
Max number of layers in onion designs	10
Max candidate set size	512000
Probability of failure presentation	%
Replicate tolerance	0.1
Save Design Space Explorer cache	Yes
Show expanded design factors	Yes
Show units in lists	No
Stop One-Click for each response	Yes
[-] <b>Plot</b>	
Header format	Default
Header format 1	\$(PlotTitle) - \$(InvestigationTitle) \$(FitMethod;"\$(FitMethod)") \$(ColorBy;"Colored by \$(ColorBy)"). \$(Labels;"\$(Labels) labels")
Header format 2	\$(PlotTitle) \$(FitMethod;"\$(FitMethod)") \$(InvestigationTitle) \$(ColorBy;"Colored by \$(ColorBy)"). \$(Labels;"\$(Labels) labels")
Plot engine	Direct2D
[-] <b>Collect usage data</b>	
Collect anonymous data	Yes

Show details Reset

### Audit trail

Set Audit trail options.

- **Enable the audit trail for new investigations** - Automatically enables the audit trail for all new investigations.

For more, see the Audit trail section in Chapter 10, View.

### General

Under **General** you can change the following options,

- **Automatically display output window** - Automatically open or activate the output window when MODDE encounters warnings and errors. This window displays details about the warnings and errors.
- **Automatic fit** - Automatically fit the investigation when it is opened.
- **Close open investigation when opening a new investigation** - Close already open investigation when opening or creating a new investigation.
- **Factor axes in back transformed units** - Display factors on plot axes in back transformed units.
- **Max number of layers in onion designs** - Maximum number of possible layers for an Onion design. MODDE's default is 10.

- **Max candidate set size** - The maximum number of rows in the candidate set. This number will prevent MODDE from creating candidate sets that are too large. The maximum size of the candidate set that you can create and generate a design from is limited by the amount of RAM in your computer.
- **Probability of failure presentation** - The unit to display the probability of failure in, % (default) or DPMO (defects per million opportunities).
- **Replicate tolerance** - The replicate tolerance [0 .. 0.20] is the number used when MODDE decides whether experiments can be considered to be replicates or not. Default is 0.1, that is, 10% of half the range for each factor.
- **Save Design Space Explorer calculations** - Include the Design Space Explorer calculated results with investigation (.mip) when saving it. If the calculations are saved with the investigation, no recalculations are necessary when redisplaying this plot. *WARNING!* When selected, the .mip may become large.
- **Show expanded design factors in RED-MUP worksheet** - Show each design factor expanded.
- **Show units in lists** - Show units after the factor and response names in lists and spreadsheets. By default this option is set to *No*.
- **Stop One-Click for each response** - By default One-Click in the Analysis Wizard stops at the Replicate plot page for each response, whether there is a warning or not. Setting this option to *No* will result in that One-Click only stops when there is a warning or if there are no warnings at the Summary-page last in the wizard.

### **Plot**

- **Header format** - Select which of the two header formats to use. Leaving *Default* results in 'Header format 1' for investigations with short names and 'Header format 2' for investigations with long names.
- **Header format 1** - When selected, header format 1 by default displays plot type, investigation name and fit method on the first row and coloring and label information on the second.
- **Header format 2** - When selected, header format 2 by default displays plot type and fit method on the first row, investigation name on the second and coloring and label information on the third.
- **Plot engine** - Select which plot engine to use for 2D plots. Available options are Direct2D or GDI+. 3D plots are always displayed using GDI+.

---

*Note: The order and what is displayed in the headers can be changed as desired using the displayed terminology. Maximum number of header rows is four.*

---

### **Collect usage data**

- **Collect anonymous data** - Collects anonymous data such as number of clicks on **One-Click** in the Analysis Wizard.

## Investigation options

On the **File** tab, click **Options**, then click **Investigation options** to view the settings valid for the current investigation, grouped under the headers **Audit trail**, **General**, **Statistical options**, and **Design space and optimizer options**.

 Options for the current investigation

<b>Audit trail</b>	
Enable the audit trail	No
<b>General</b>	
Design matrix	Current worksheet scaled and centered
Factor representation	Factor name
Factor presentation [Qualitative]	Regular
Factor presentation [Qualitative, Coefficients plot]	Regular
Number format	Auto
Plot label type	Experiment number
Response representation	Response name
<b>Statistical options</b>	
Alpha level	5%
Coefficients	Scaled and centered
<b>Interval type</b>	
Confidence level	95%
Tolerance proportion	95%
Correlation in probability of failure	No
Prediction block effect	Fixed
R2	R2 [explained variation]
Residuals	Deleted studentized
<b>Predict tab and optimizer options</b>	
Acceptance limit (%)	1
<b>Interval type</b>	
Confidence level	95%
Tolerance proportion	95%

Available settings include,

### **Audit trail**

When **Enable the audit trail** is *Yes*, the **Audit Trail** logs every change that affects the investigation. Your administrator may have locked the settings on this tab. By default the audit trail is disabled. For details, see the Audit trail section in Chapter 10, View.

### **General**

#### **Design matrix**

Select the data to display in the design matrix. Available options are *The design as generated by MODDE* and *Current Worksheet scaled and centered* (default).

#### **Factor representation**

Select whether to show the *Factor name* (default) or the *Factor abbreviation* in lists and spreadsheets.

#### **Factor presentation [Qualitative]**

Select whether to display the qualitative factors in the *Regular* (default) format, showing the orthogonal settings, or *Extended*, showing all settings. This option is valid for all plots and lists displaying factors except the coefficient plot.

**Factor presentation [Qualitative, Coefficient plot]**

Select whether to display the qualitative factors in the *Regular* format, showing the orthogonal settings, or *Extended* (default), showing all settings. This option is valid for the coefficient plot only.

**Number format**

Select the number of decimals to display in lists and tables. Available options are *Auto* (default), *.00*, *.000*, *.0000*, *.00000*, and *Scientific* (e.g. 1.234e-10).

When exporting unscaled coefficients to use in other applications, use the scientific format to get maximum precision in the coefficients.

**Plot label type**

Select *None*, *Experiment number* (default), *Experiment name*, *Run order*, or *Combination ID* (available for Generalized subset designs and Stability testing designs only).

For Stability testing designs the default *Experiment name* is the *Combination ID* and the *Time*, e.g. C1:T0 for the first row. The Combination ID denotes a specific combination of settings for all factors but the Time factor.

For Generalized subset designs the default *Experiment name* is the *Combination ID*, e.g. c1 for the first row. The Combination ID denotes a specific combination of settings for all factors.

**Response representation**

Select whether to show the *Response name* (default) or the *Response abbreviation* in lists and spreadsheets.

**Statistical options****Alpha level**

Select the **Alpha level** for the lack of fit plot. Available options are *1%*, *5%* (default), and *10%*.

**Coefficients**

Select the type of coefficients. Available options are *Scaled and centered* (default for the regular coefficient plot), *Unscaled*, *PLS orthogonal*, and *Normalized* (default for the coefficient overview plot).

To select presentation format for qualitative factors in the Coefficient plots, use the **Factor presentation [Qualitative, Coefficient plot]** option earlier in this page. By default the factors are displayed in the *Extended* form in this plot, showing all settings.

For more, see the Coefficients section in the Statistical appendix.

**Interval type and probability levels**

Select the type of interval to use to estimate the uncertainties. Available interval types are *Confidence* (aka *Average*), *Prediction* (aka *Individual*), and *Tolerance*.

An interval estimation is used in statistics as an uncertainty measure of a population parameter computed from sample data.

The simulation will generate points with variance based on the selected interval estimate. This will in turn generate the required safety margins that correspond to the defined probability limit of the ideal average prediction. The safety margin increases from *Confidence* to *Prediction* (and in general) to *Tolerance*.

We use interval estimates for two purposes in MODDE:

- to assess if model parameters are significantly different from zero (null-hypothesis significance tests) and
- to state an interval within which we are confident that we find future predictions.

For more, see the Interval type and probability levels subsection later this section and the Interval estimates subsection in the Statistical appendix.

### **Correlation in probability of failure**

Option to turn on taking response correlation into account. Set as *Yes* if there is an underlying unmodeled factor that the responses depend on, or if there is causality between the responses.

For more, see the Correlation in probability of failure subsection in the Design space appendix

### **Prediction block effect**

Select to use fixed or random block effect in predictive functions.

Select the block factor as *Fixed* effect when the external variability can be *set at will* and the primary objective for blocking is to eliminate that source of variability.

Select the block factor as *Random* effect when the external variability *cannot be controlled* and set at will, and the primary objective is to make prediction without specifying the block level, and taking into account the external variability.

For more, see the Random versus fixed block factor section in the Statistical appendix.

### **R2**

Select the type of R2 to display in the summary plots. Available options are *R2 [explained variation]* (default) and *R2 adjusted [explained variance]*.

### **Residuals**

Select the type of residuals to display in the residual plots. Available options are *Raw* in the original units, *Standardized*, and *Deleted studentized*. The default depends on fit method and number of degrees of freedom. For more, see the Residuals section in the Statistical appendix.

### **Predict tab and optimizer options**

#### **Acceptance limit**

Enter the default **Acceptance limit** for the investigation in percentage or DPMO.

This default Acceptance limit is used in all Design space plots and Setpoint features. The default value is *1%* outside (*10 000* DPMO).

DPMO = Defects Per Million Opportunities.

#### **Interval type and probability levels**

Select the type of interval to use to estimate the uncertainties. Available interval types are *Confidence* (aka *Average*), *Prediction* (aka *Individual*), and *Tolerance*.

An interval estimation is used in statistics as an uncertainty measure of a population parameter computed from sample data.

The simulation will generate points with variance based on the selected interval estimate. This will in turn generate the required safety margins that correspond to the defined probability limit of the ideal average prediction. The safety margin increases from *Confidence* to *Prediction* (and in general) to *Tolerance*.

We use interval estimates for two purposes in MODDE:

- to assess if model parameters are significantly different from zero (null-hypothesis significance tests) and
- to state an interval within which we are confident that we find future predictions.

For more, see the Interval type and probability levels subsection later this section and the Interval estimates subsection in the Statistical appendix.

## Audit trail

When the **Audit Trail** is turned on each investigation in MODDE has a separate audit trail. Each audit trail consists of one or more sessions that in turn consist of events. A new session is started and appended to the audit trail when an investigation is opened, and ends when the investigation is saved.

In addition to logging events, MODDE logs information about the user, and date and time of the events.

To view the audit trail, click the **Audit Trail** tab in the **Output / Notes / Audit trail** pane. If this tab is not shown, display it by selecting the **Audit Trail** check box on the **View** tab.

### ***Enable and disable the Audit Trail***

By default the audit trail is disabled.

To turn it on for the current investigation, in **File | Options**, on the **Investigation options** page, under Audit trail, select *Yes* in **Enable the audit trail**.

To turn on the audit trail for new investigations, in **File | Options**, on the **MODDE options** page, under Audit trail, select *Yes* in **Enable the audit trail for new investigations**.

Administrators can disable the turning on and off of the audit trail, i.e., always have it on or off. For instructions on how to disable the audit trail options, see the knowledge base at [www.umetrics.com](http://www.umetrics.com).

### ***Save or clear the Audit Trail***

To empty the audit trail, in the Audit Trail pane, right-click and click **Clear Audit Trail**.

To save the current version of the audit trail, separate from the investigation, in XML format, in the Audit Trail pane, right-click and click **Save as**.

### ***Audit trail and internet explorer***

MODDE uses Internet Explorer functionality to display the audit trail.

### ***Logged in the audit trail***

Specific actions the MODDE audit trail logs are:

- Factors (adding, modifying, deleting), displaying all details about the factor after the change
- Responses (adding, modifying, deleting), displaying all details about the response after the change
- Constraints (modifying, deleting)

- Candidate set
- Inclusions
- Reference mixture
- Objective
- Generators
- Design
- Complement design
- Model
- Worksheet, every change of every cell
- Fit method
- Activation and deactivation of the Audit Trail.
- The audit trail also registers when a digital signature in the Audit trail is incorrect.

---

## Coefficients

### Scaled and centered coefficients

The regression coefficients that are displayed in MODDE are computed for centered and scaled data. It is also possible to select to display "unscaled and uncentered" coefficients.

The scaled and centered coefficients are the coefficients of the fitted model, for which the factors were centered and scaled. The default scaling in MLR is *orthogonal* scaling. With PLS, the factors are centered and scaled to *unit variance*.

### Orthogonal coefficients in PLS

The "centered and scaled" coefficients in PLS regression models are computed from factor values scaled to unit variance.

The *orthogonal coefficients* in PLS regression re-express the coefficients such that they correspond to factors that are *centered* and *orthogonally scaled*, i.e., by using the mid-range and low and high values in the factor definition (coded as -1 and 1), see the Scaling subsection for more details. Orthogonal coefficients in PLS regression models are not available when there are only formulation factors in the investigation.

With process and mixture factors, the PLS orthogonal coefficients refer to process factors scaled orthogonally, and mixture factors unscaled (original units).

The orthogonal coefficients in PLS regression models are only available when the model is fit with PLS regression.

---

**Note:** *The orthogonal coefficients in PLS regression models are only meant for comparison with the corresponding MLR coefficients. They are incorrect unless the design is balanced and the mean is equal to the mid-range.*

---

## Unscaled

The unscaled coefficients are the coefficients corresponding to unscaled, uncentered data. When exporting unscaled coefficients to use in other applications, be sure to use the *E-format* in order to obtain maximum precision in the coefficients.

## Normalized coefficients

To make the coefficients comparable between responses when the responses have different ranges, the "centered and scaled" coefficients are normalized with respect to the variation in  $Y$ . That is, they are divided by the standard deviation of their respective response, *i.e.*, by the standard deviation in the corresponding  $Y_i$ , for  $i = 1, 2, \dots, m$ .

## Confidence interval

Intervals for coefficients (such as *e.g.*, a confidence interval) and predictions are computed using the total number of observations, regardless of missing values when the regression model is a PLS regression model and all polynomial models (*i.e.*, the polynomial model specified by the factors) for all responses are the same. For MLR and PLS regression models with different polynomial models for different responses, the total number of observations is the number of elements in the response without missing values. This total number of observations is displayed as  $N$  at the bottom of plots and lists.

## Extended or compact format

For qualitative factors at  $q$  levels, with  $q > 2$ , MODDE generates  $q - 1$  dummy variables, indexed from 2 to  $q$ .

By default the Coefficient plot displays all  $q$  settings using the **Extended** format. See the Investigation options subsection in Chapter 4, File, for details.

## Interval type and probability levels

Select the type of interval to use to estimate the uncertainties. Available interval types are *Confidence* (aka *Average*), *Prediction* (aka *Individual*), and *Tolerance*.

An interval estimation is used in statistics as an uncertainty measure of a population parameter computed from sample data.

The simulation will generate points with variance based on the selected interval estimate. This will in turn generate the required safety margins that correspond to the defined probability limit of the ideal average prediction. The safety margin increases from *Confidence* to *Prediction* (and in general) to *Tolerance*.

We use interval estimates for two purposes in MODDE:

- to assess if model parameters are significantly different from zero (null-hypothesis significance tests) and
- to state an interval within which we are confident that we find future predictions.

There are several types of interval estimates that can be used in MODDE. A common statement is to say that with a confidence of *e.g.* 95 % we will find a future sample in this region enclosed by the interval, *i.e.* average  $\pm$  interval.

The types of intervals available in MODDE are;

### **Confidence interval**

This interval encloses average of the true population, with some confidence, and is mainly used to illustrate the variance of the model coefficients.

### **Prediction interval**

This interval encloses a region within which we are confident that the next observation will fall.

### **Tolerance interval**

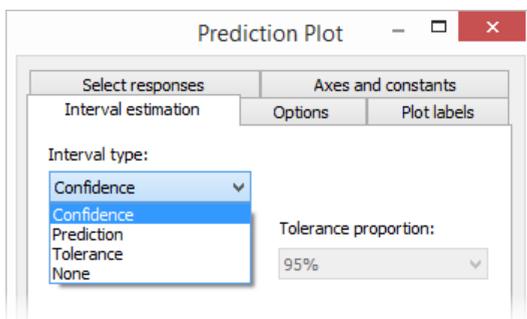
This interval encloses a region within which we are confident that some proportion of future samples will fall.

### **Specifying the probability level**

The *Confidence interval* and *Prediction interval* require an acceptance level, *i.e.*, roughly speaking a probability. It is usually expressed as the *Confidence level* (90%, 95% or 99%). The *Tolerance interval* requires an acceptance level, but also requires a parameter for the fraction of future samples that fall within the interval, called the *Tolerance proportion*. The default setting In MODDE for evaluation of model parameters is *Confidence interval* at 95 %. The default setting for predictions is *Prediction interval* at 99%.

### **Plot properties**

The interval settings can be adjusted by right-clicking the plots, selecting **Properties** and in the dialog that appears, clicking the **Interval estimation** tab. For Coefficients and other model plots with interval estimates only the Confidence intervals are available, thus the tab is named **Confidence interval**.



In the design space simulations, the Confidence level is controlled by the simulation procedure; therefore, the *Confidence level* box is unavailable. If the interval type is *Tolerance interval*, the *Tolerance proportion* is controlled by the simulation procedure and therefore the *Tolerance proportion* box is unavailable.

In the **Design space options** page, a **Probability of failure limit** of 1% corresponds to 99% in **Confidence level** for the *Confidence* or *Prediction interval* types. When the interval type is *Tolerance*, the **Probability of failure limit** of 1% corresponds to a **Tolerance proportion** of 99%.

---

**Note:** If the *Use model error* check box is available but cleared, settings on the *Interval estimation* page have no effect.

---

For more, see the Interval estimates section in the Statistical appendix.

---

## Residuals

### Raw residuals

The *raw residual* is the difference between the observed and the predicted values, *i.e.*

$$e_i = y_i - \hat{y}_i.$$

### Standardized residuals

The *standardized residual* is the raw residual divided by the residual standard deviation

$$e_{std,i} = e_i / s,$$

where  $s$  is the residual standard deviation.

Residual plots for PLS regression present standardized residuals by default.

### Deleted studentized residuals

MODDE defaults to plotting *Deleted studentized residuals* when fitting with MLR and the model has at least three (3) *real* degrees of freedom. Deleted studentized residuals are not available when fitting with PLS regression.

The Deleted studentized residual is the raw residual ( $e_i$ ) divided by an estimate of its standard deviation. The estimation of the standard deviation is computed from the “deleted” standard deviation ( $s_i$ ), which is the residual standard deviation ( $s_i$ ) computed with observation  $i$  left out. When the  $i$ th residual is excluded like this, the residual is sometimes said to be *externally studentized*. The variance of the  $i$ th residual is defined as

$$\text{Var}(e_i) = \sigma^2(1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'),$$

where  $\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'$  is the  $i$ th element of the hat matrix, sometimes denoted *leverage*. Thus the standard deviation of the residual is

$$\text{sd}(e_i) = \sqrt{\text{Var}(e_i)} = \sigma \sqrt{1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'}.$$

Hence, the Deleted studentized residual is computed as

$$e_{del,i} = \frac{e_i}{\text{sd}(e_i)} = \frac{e_i}{s_{-i}\sqrt{1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'}}$$

where  $s_{-i}$  is the estimate of the residual standard deviation with the  $i$ th sample left out.

For more information see *e.g.*, Belsley, Kuh and Welsch (1980).

---

**Note:** Deleted studentized residual requires at least three degrees of freedom.

---

### References

Belsley, David A.; Kuh, Edwin and Welsch, Roy E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley and Sons, New York, 1980.

### Acceptance limit

Enter the default **Acceptance limit** for the investigation in percentage or DPMO.

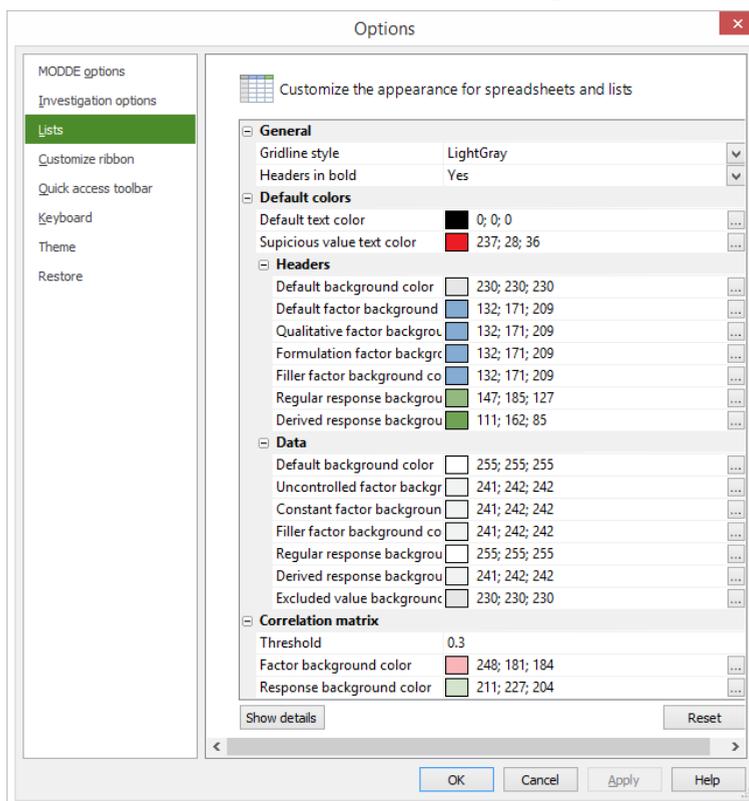
This default Acceptance limit is used in all Design space plots and Setpoint features. The default value is 1% outside (10 000 DPMO).

DPMO = Defects Per Million Opportunities.

### List options

Use this page to change the appearance and properties of lists.

- General options such as gridline style and header in bold.
- Default colors for text and suspicious value.
- Default background color for factors and responses.
- Default colors for header background for different types of factors and responses.
- Default background colors for different types of data.
- Correlation matrix threshold and factor and response background colors.

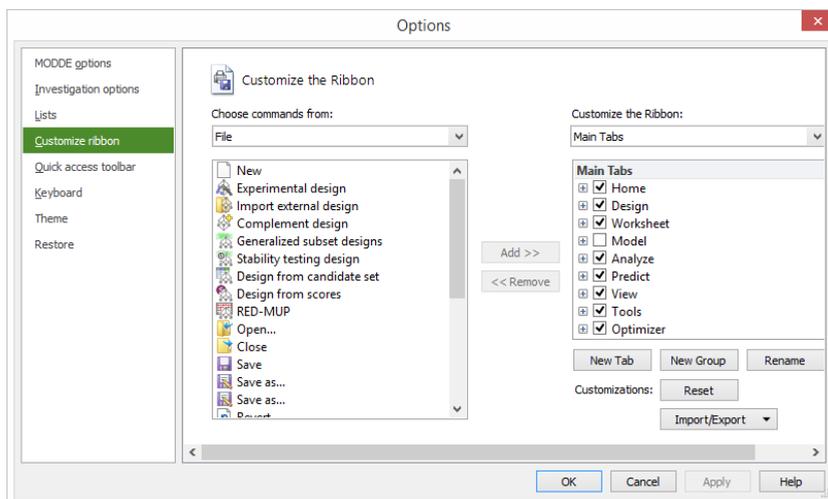


### Customize ribbon

On the **File** tab, click **Options**, then click **Customize ribbon** to access the page.

1. Select the command in the left box that you want to place.
2. Click the location in the right box where you would like the command to be shown.

3. Click **Add**.
4. Click **Close** to return to MODDE.

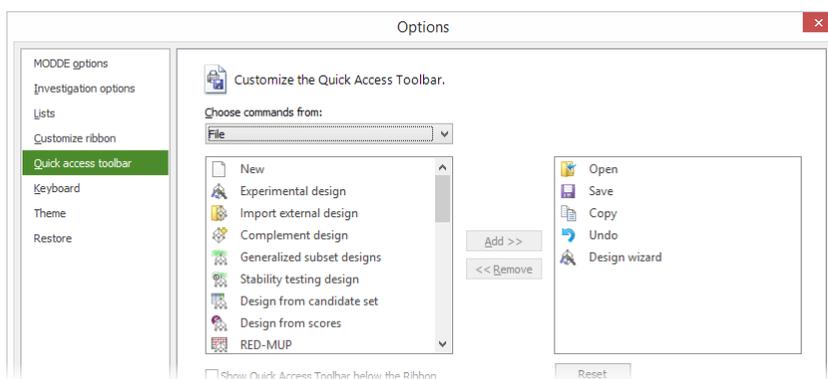


*Note: With this dialog open, all buttons and tabs currently displayed can be dragged to new positions. Buttons can be removed by pulling them down.*

## Customize quick access toolbar

On the **File** tab, click **Options**, then click **Quick access toolbar**.

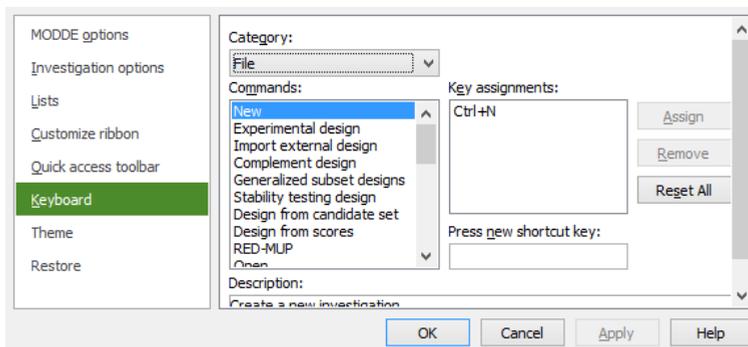
1. Select the command in the left box that you want to add.
2. Click **Add**.
3. Click **OK** to return to MODDE.



## Keyboard shortcuts

MODDE includes many keyboard shortcuts in order to access commonly used functions. Many of these shortcuts will be familiar to Windows users, while some are MODDE specific.

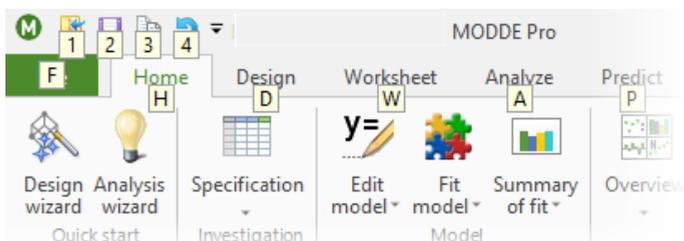
Key assignments can be modified as desired in the **Options** dialog, Keyboard page.



## KeyTips

In order to quickly navigate around MODDE using almost exclusively the keyboard instead of the mouse, MODDE implements the same system of KeyTips as Microsoft Office.

1. Press the Alt key to see the KeyTips.
2. Then press the indicated number or letter to run the associated command.



## General keyboard shortcuts

Copy: Ctrl+C, Ctrl+Insert

Paste: Ctrl+V, Shift+Insert

Cut: Ctrl+X, Shift+Delete

Print: Ctrl+P

Undo: Ctrl+Z, Alt+Backspace

Save: Ctrl+S

Delete: Delete

Select all: Ctrl+A

Insert: Insert

Full screen: F11

Start a new investigation: Ctrl+N

Open a different investigation: Ctrl+O

## List/spreadsheet specific

\* Ctrl+arrow up/down/left/right moves the cursor to corresponding edge of the worksheet.

\* Shift+arrow up/down/left/right - same as above but also selects cells.

- \* Alt+PgUp/PgDn scrolls left/right by page.
- \* Shift+Home/End scrolls the worksheet to the left/right marking cells.
- \* Ctrl+Shift+Home/End marks from current cell to top left/bottom right.
- \* In RED-MUP and Specification tabbed windows: Ctrl+PgDn/PgUp selects next/previous tab.

### **MODDE specific keyboard shortcuts**

Open the Design wizard: Ctrl+W

Open the Analysis wizard: Alt+W

Add to favorites: Ctrl+D

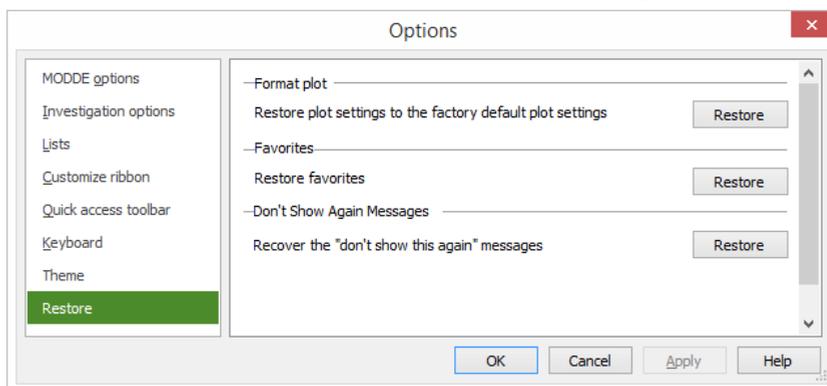
Add to report: Ctrl+R

Properties: Alt+Enter

### **Restore**

On the **File** tab, click **Options**, then click **Restore** to restore to MODDE factory defaults.

There are three sections, Format plot, Favorites, and Don't show again messages.

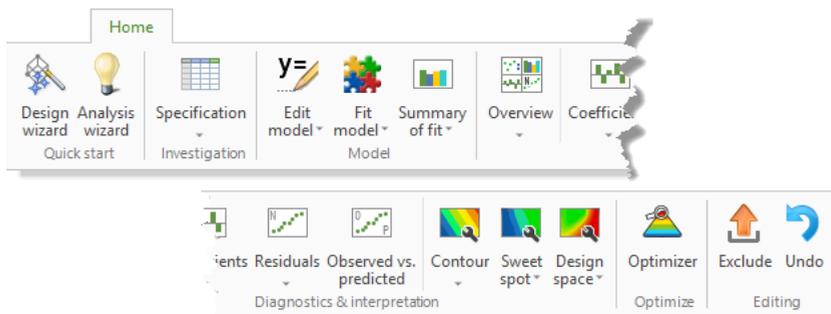


# 05-Home

---

## Introduction

The **Home** tab collects some of the most used functions together in one place, making them easy to find and access. Some functions are located only on this tab, while most functions are also located on other tabs. When a function is also located elsewhere, this chapter only provides a brief description and directions to the main description in another chapter.



The functions on the Home tab are divided into groups in order to make commonly used functions easier to find,

- **Quick start** group
  - **Design wizard** - provides guidance through the process of setting up a new investigation or changing an existing one.
  - **Analysis wizard** - provides guidance through the main steps of analyzing a model.
- **Investigation** group
  - **Specification** - opens the **Specification** window with worksheet, factors and responses, and constraints available.
- **Model** group
  - **Edit model** - to add and remove model terms, see the **Model list**.
  - **Fit model** - to refit the current model.
  - **Summary of fit** - provides access the summary plots and lists.
- **Diagnostics & interpretation** group
  - **Overview** - opens the **Overview plot** window displaying a customizable selection of plots.

- **Coefficients** - provides access to coefficient plots and lists
- **Residuals** - provides access to the residuals lists and plots.
- **Observed vs. predicted** - displays observed values vs. predicted values.
- **Contour** - opens the contour plot wizard.
- **Sweet spot** - creates a plot highlighting values that are within the user specified range.
- **Design space** - opens the **Design Space Plot** wizard.
- **Optimize** group
  - **Optimizer** - opens the optimizer window.
- **Editing** group
  - **Exclude** - excludes the currently selected experiment or data point.
  - **Undo** - reverses the last action.

---

## Design wizard

The **Design wizard** is a set of pages that guide you through the steps of setting up your investigation. The **Design wizard** covers factors, constraints, responses, objective, and model.

To access the **Design wizard**, on the **Home** tab, in the **Quick start** group, click **Design wizard**.



---

---

**Hint:** Press Ctrl+W to open the **Design wizard**.

---

---

For more information, see Chapter 3, Design wizard.

---

## Analysis wizard and One-Click

The **Analysis wizard** can be executed in two modes:

1. Clicking **One-Click** to automatically transform the data and tune the model if appropriate, or
2. Stepping through the pages in a manual mode using **Next** and investigating each page.

### **One-Click**

When stepping through the Analysis wizard using **One-Click** the tests available on each page are automatically performed and transformation of the data and/or tuning the model is done automatically. If there is a warning that cannot be handled automatically, the wizard stops at that page and you can decide how to proceed. More about the tests and warnings in the description of the individual pages later in this section.

### **Manual steps**

When stepping through the Analysis wizard clicking **Next** to investigate each page, the Analysis wizard provides guidance through the main steps of analyzing a model and is the recommended method for making changes to and adjusting the model. The

**Analysis wizard** covers:

- Reviewing raw data
- Fitting data
- Diagnostics
- Refining the model.

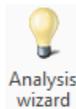
---

*Note: When opening the **Analysis wizard**, the investigation is fitted using the default fit method.*

---

### **Accessing the Analysis wizard and One-Click**

On the **Home** tab, in the **Quick start** group, click **Analysis wizard**.



### **Included graphs and plots**

- Replicates
- Histogram
- Coefficients
- Summary of Fit
- Residuals Normal Probability
- Observed vs. Predicted.

### **Toolbar functions**

The functions available in the **Analysis wizard**,

Response:

**Response:** Swaps between active responses.



**Exclude:** Excludes model terms or experiments.



**Undo:** Reverses the most recent change to the model.



**Regression line:** Toggles the regression line.

 **Show limits:** Toggles lines for applicable Min, Max, and Target levels.

 **Auto transform** **Auto transform:** Available when the value for Skewness test is higher than 2.

 **Transform** **Transform:** Available when looking at the histogram; opens the **Transform Response** dialog box.

 **Auto tune** **Auto tune:** Available when there are insignificant model terms that improve the model (increase the Q2-value) when removed.

 **Edit model** **Edit model:** Available when looking at the coefficients plot; opens the **Edit Model** dialog box.

 **Square test** **Square test:** Available for the coefficients plot when the square test detects significant square terms.

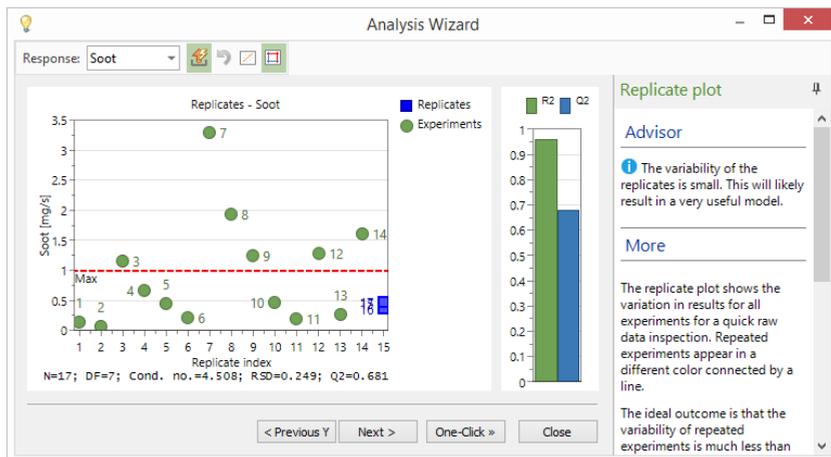
 **Interaction test** **Interaction test:** Available for the coefficients plot, when there are significant interactions detected and the design is resolution IV.

## One-Click simple case

Using One-Click in the Analysis Wizard aims at making the analysis process as simple for you as possible. Auto transform and Auto tune are performed without your involvement while there are other warnings that need your attention. This section explains a One-Click scenario where no interaction is necessary.

### Replicate plot

With default settings, One-Click stops at the Replicates plot for each response, helping you to take a quick look at the raw data before One-Clicking your way through the wizard.



### Summary page

After One-Clicking all responses there is a summary of the automatically performed actions on the Summary page.

Analysis Wizard

## Summary

Summary of changes done during the current session of the Analysis wizard.

**Auto transform<sup>1</sup>**

The following responses were automatically transformed:

Response	Transform
Soot	Log: 10Log(Y)

**Model auto tuning<sup>2</sup>**

Response	New model	Terms removed
Fuel	Air, EGR, NL	Air*EGR, EGR*NL, Air*NL
NOx	Air, EGR, NL, EGR*NL	Air*NL, Air*EGR
Soot	Air, EGR, NL	EGR*NL, Air*NL, Air*EGR

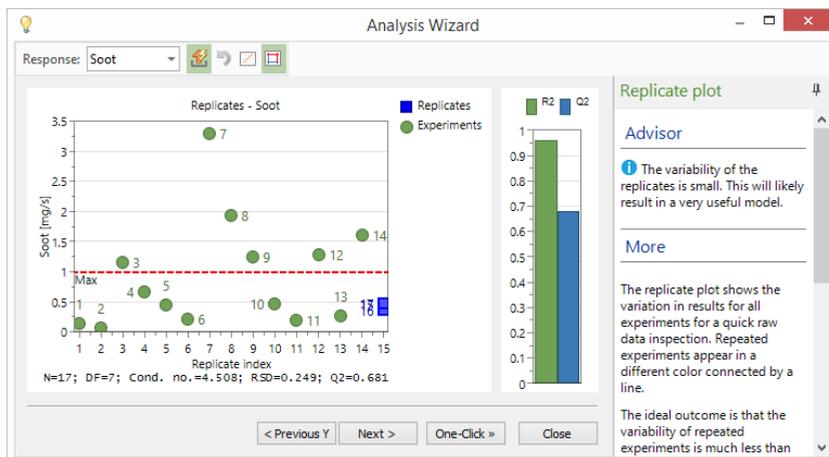
1) Auto transform log transforms a response if it is skewed and the transformation improves Q2 and skewness.  
2) Auto tune model automatically removes terms from the model if the p-value for the coefficient is greater than 0.05 and Q2 increases after removing the term.

< Back   Next >   Finish   Close

*Note: If manual changes were done, such as adding a term using the Square/Interaction test, that information is also listed here.*

## Replicates

Replicates is the first page shown when opening the **Analysis wizard** and provides a quick overview of the raw data. This plot is displayed for each response when using One-Click in default mode.



### Understanding the plot

Experiment points above the red *Max* line or under the *Min* line are generally undesirable.

MODDE shows replicated experiments as blue (default) points on the same replicate index.

The variation of repeated experiments should be less than the overall variation for the response.

The Advisor in the Analysis Wizard shows the results after running Tukey's test to identify outliers and the variability test to indicate possible expected results based on the relative replicate variability. For more about these tests, see the Tukey's and variability tests subsection in the Statistical appendix.

### **Working with the plot**

It is possible to identify outliers when examining the **Replicates** plot.

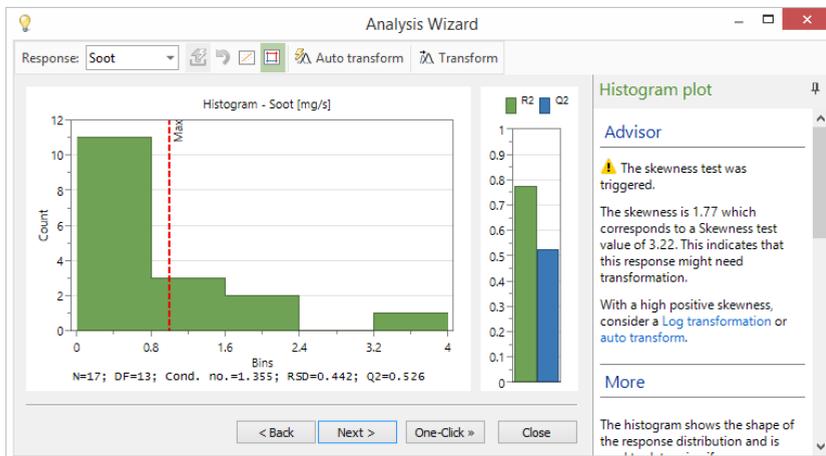
Some of the common causes and suggested solutions to outliers are:

- Incorrect data entry in the worksheet – In this situation correct the worksheet and refit.
- Experiment incorrectly carried out – If the experiment was not done correctly, repeat the failed experiment and refit.
- The deviating result is correct and the experiment does not fit the model – Empirical models are only valid within a limited interval. If the deviating point is correct but uninteresting, then remove it from the analysis and refit the model without it. If the deviating point is in an interesting area, it may be necessary to construct a finer-grained design in that area.

Click **Next** to see the **Histogram** for the current response.

## **Histogram**

Histogram is the second page of the **Analysis wizard**. This page provides the opportunity to see the shape of the response distribution and apply a transformation if required.



### **Understanding the plot**

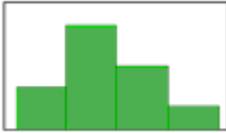
A proper estimate of the distribution requires a minimum of 11 observations. The important decision to make at this step in the Analysis wizard is to decide if a transformation is required. When the Skewness test value falls outside the -2 to 2 range there is a warning in the Advisor suggesting a transformation.

By selecting an appropriate transformation, a non normal distribution might be transformed to normal distribution. In general, normally distributed responses will give better model estimates and statistics.

If a response has already been transformed, the response will have a tilde ("~") after its name in the plot. To see what transformation has been applied, click **Transform**.

### ***Transforming the response***

If the distribution is not a normal "bell-shaped" distribution, then it is likely a transformation is required.



The desired distribution is a "bell shaped" normal distribution.



If there is a positive skewness, a log transformation is probably preferable.



If there is a negative skewness, a negative log transformation is probably preferable.

---



---

**Hint:** Biological systems almost always require a logarithmic transform.

---



---

### ***Auto transform***

For the case where the Skewness test value is larger than 2, Auto transform is available. Clicking **Auto transform** or **One-Click** log-transforms the response provided Q2 is increased and skewness decreased.

For details, see the Auto transform and Auto tune subsection in the Statistical appendix.

### ***Manual transformation***

To apply a transform manually:

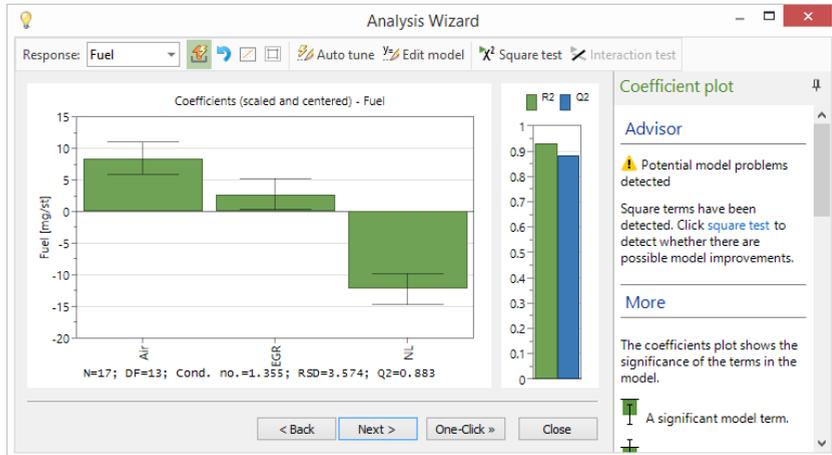
1. Determine the required transformation. For transformation suggestions, see the Box-Cox subsection in Chapter 8, Analyze.
2. Click **Transform**.
3. Click the desired transformation in the **Transformation** list.
4. Change the values of C1, C2, and C3 (if available) from default values if the data requires it.
5. Click **Apply** to see the transformed response.
6. Click **OK** to return to the **Analysis wizard**.

Click **Back** to go to **Replicates**.

Click **Next** to go to **Coefficients**.

## Coefficients

The coefficients page of the **Analysis wizard** provides access to the **Coefficients plot** for the current response (Y).



A significant term is one with a large distance from  $y=0$  as well as having an uncertainty level that does not extend across  $y=0$ . A non significant model term is a model term close to  $y=0$  and with an uncertainty level that crosses  $y=0$ . The trigger for the **Auto tune** feature in the Coefficient plot is that Q2 increases if the smallest non significant terms are excluded.

When confoundings are present, the coefficient plots and lists display a bracket ("#") after the term. Point to the column to view the confounded terms.

### Understanding the plot

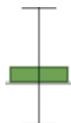
The coefficient plot presents a graphical representation of the model terms in order to determine their significance.

#### Significant model terms:



- Far away from  $y=0$  (either positive or negative)
- Uncertainty range does not cross  $y=0$ .

#### Non significant model terms:



- Close to  $y=0$
- Uncertainty range crosses  $y=0$ .

### **Working with the coefficient plot**

The coefficients plot provides support for tuning of the model by adding or removing model terms.

#### **Remove non significant model term using Auto tune**

When Auto tune is available, click it to remove all insignificant terms at once. Note here that Auto tune removes each term individually and checks the result before continuing. For more, see the Auto transform and Auto tune subsection in the Statistical appendix.

#### **Remove non significant model term manually**

Starting with the least significant model term

1. Check that no significant model term depends on the non significant model term, meaning that if a higher order model term (e.g. interaction) is significant a non significant linear term should not be removed.
2. Click **Exclude**.
3. Click the non significant model term.

---

**Hint:** When excluding terms manually, exclude only one model term at a time as it may make other model terms more significant. Q2 will probably increase as non significant model terms are removed, so note the before and after values as each term is excluded and aim to maximize Q2 to achieve the best possible prediction performance. DF (Degrees of Freedom) will also increase. A higher DF is better for a proper estimate of the confidence interval.

---

#### **Adding model terms**

Click **Edit model** to include any model terms to the investigation that are not visible. This can also be used to include a model term that was previously removed.

When applicable, click **Square test** or **Interaction test** to add model terms.

Click **Next** to go to **Summary of Fit**.

Click **Back** to return to **Histogram**.

#### **Square test**

The square test is performed when adding a square term does not result in a condition number that exceeds 100. When the square test has been performed and a significant square term has been found, the **Square test** becomes available. The test starts with the current model, adds one specific square term at a time and tests them individually.

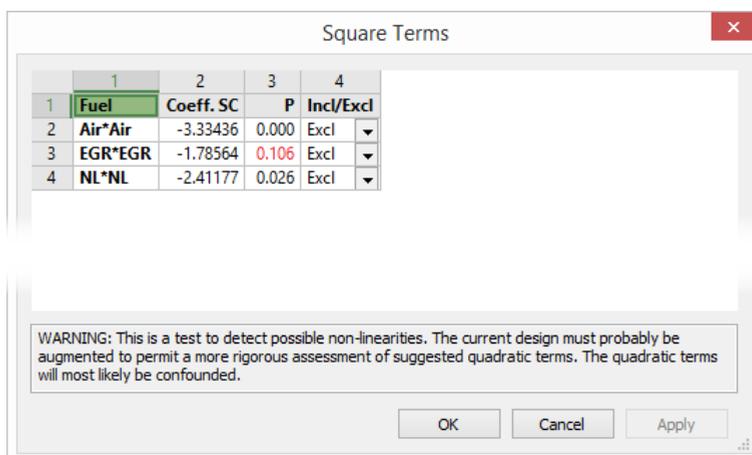
---

**Note:** *Quadratic terms are confounded when the design is a screening design. External knowledge or further experiments are required to permit a more rigorous assessment of which quadratic terms are necessary for the current design. Therefore addition of a square term is not done automatically in One-Click.*

---

To add a square term:

1. Click **Square test**.
2. Look for terms that are shown in black instead of red in the **P** column.
3. Change **Excl** to **Incl** in the **Incl/Excl** column for the square to include.
4. Click **OK** to return to the **Analysis wizard**. Note here that the addition of a square term may reactivate the interaction test and vice versa.




---

**Hint:** To check the correlation between square terms, open the correlation matrix.

---

### ***Interaction test***

For screening designs of resolution IV, the interaction test can help find significant interaction terms. In resolution IV designs, all interaction terms are confounded which means that you should consider which interaction term is most reasonable in the **Interactions** dialog box. This is also the reason why addition of interaction terms cannot be done automatically in One-Click but must be done manually by you.

Confounded terms are model terms that are mathematically identical to other model terms in the current design.

If significant interaction terms are detected, the Advisor will show a warning and the **Interaction test** becomes available. The interactions are tested individually and the values shown in the table are results with the individual interaction included.

To add significant interaction terms:

1. Click **Interaction test**.
2. Look for terms that are shown in black instead of red in the **P** column.
3. Based on which model terms are significant, change **Excl** to **Incl** in the **Incl/Excl** column.
4. Click **OK** to return to the Analysis wizard. Note here that the addition of an interaction term may reactivate the square test and vice versa.

---

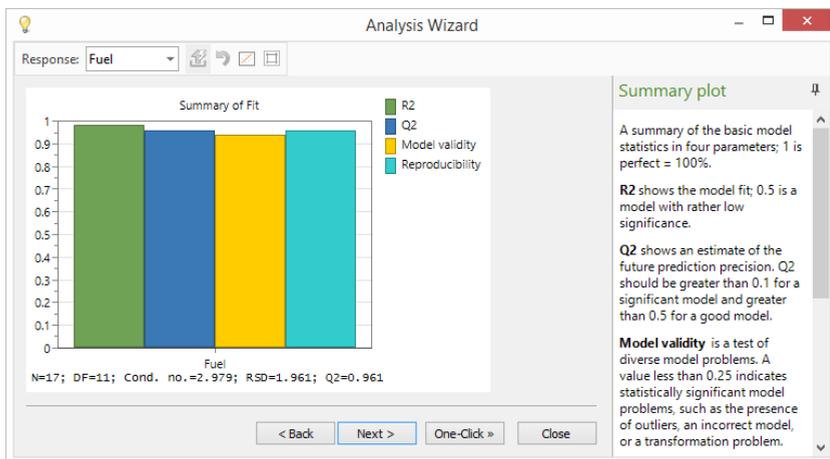
**Note:** The **Confounded with** column in the **Interactions** dialog box shows which model terms the model term is confounded with. For more, see the *Confoundings* subsection in Chapter 6, Design.

---

	1	2	3	4	5	6
1	T15	Coeff. SC	P	Confounded with	Confounded with	Incl/Excl
2	A/F*Dis	-0.125001	0.975	Surf*Lub		Excl
3	A/F*Surf	-0.75	0.853	F1/F2*Bin	Dis*Lub	Excl
4	A/F*Lub	-9.875	0.004	Dis*Surf		Excl
5	Dis*Surf	-9.875	0.004	A/F*Lub		Excl
6	Dis*Lub	-0.75	0.853	A/F*Surf	F1/F2*Bin	Excl
7	Surf*Lub	-0.125001	0.975	A/F*Dis		Excl

## Summary of Fit

The **Summary of Fit** page of the **Analysis wizard** provides a summary of the basic model statistics.



## Understanding the plot

Summary statistics are presented in four parameters ( $R^2$ ,  $Q^2$ , Model validity, and Reproducibility) where 1, or 100%, is perfect.  $R^2$  and  $Q^2$ , the first two columns, should be close in size. The difference shouldn't be more than 20% in most situations.

### $R^2$

Shows the model fit. A model with  $R^2$  of 0.5 is a model with rather low significance.

### $Q^2$

Shows an estimate of the future prediction precision.  $Q^2$  should be greater than 0.1 for a significant model and greater than 0.5 for a good model. The difference between  $R^2$  and  $Q^2$  should also be smaller than 0.3 for a good model.  $Q^2$  is the best and most sensitive indicator.

### Model validity

A test of diverse model problems. A value less than 0.25 for *Model validity* indicates statistically significant model problems, such as the presence of outliers, an incorrect model, or a transformation problem. A low value here may also indicate that a term, such as an interaction or square is missing. When the pure error is very small (replicates almost identical), the model validity can be low even though the model is good and complete. When the pure error is so small that the replicates are deemed identical by MODDE, the model validity is labeled *Missing* as it cannot be calculated.

---

*Note: Model validity might be low in very good models (Q2>0.9) due to high sensitivity in the test or extremely good replicates.*

---

### Reproducibility

The variation of the replicates compared to overall variability. The *Reproducibility* should be greater than 0.5.

### Improving summary statistics

Summary statistics can be improved by tuning the model in the various steps of the **Analysis wizard**.

- Select the appropriate transformation in the Histogram.
- Remove non-significant model parameters in the Coefficient plot.

Click **Next** to go to **Residuals Normal Probability**.

Click **Back** to return to **Coefficients**.

## Residuals Normal Probability

This page shows a plot with the residuals of a response vs. the normal probability of the distribution.



### Understanding the plot

If the experiments are on a straight line, then the residuals are normally distributed. Points outside +/- 4 SD are considered outliers and should be examined for errors.

A curved pattern indicates non modeled quadratic relations or incorrect transformation of the response. If the Degrees of Freedom is under 5 (DF < 5), the plot may display a strange pattern.

### Improving the Residuals Normal Probability plot

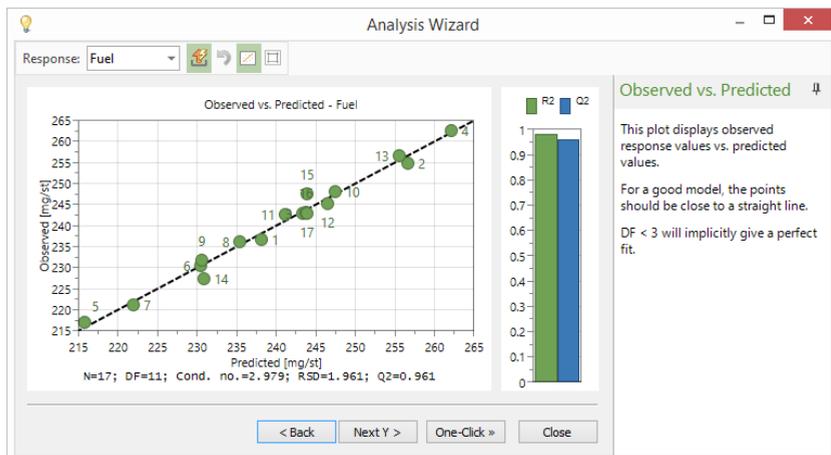
- If the points are not linear, the response may require a different transformation. Return to the histogram page and check.
- If the pattern is curved, return to the coefficients page and check the **Square test** for non modeled quadratic relations.
- Experiments can be excluded, but this is not recommended and is therefore never done automatically with One-Click. Outliers should instead be checked if they have been entered into the worksheet correctly. If a point proves to be a bad entry or large outlier but uninteresting point, any transformation should be removed and the model should be reset to the default before restarting the analysis after changing the value/excluding the point.

Click **Next** to go to **Observed vs. Predicted**.

Click **Back** to go to **Summary of Fit**.

### Observed vs. Predicted

This plot displays observed values vs. predicted values.



### Understanding the plot

Plots with points close to a straight line indicate good models. If Degrees of Freedom is under 3 ( $DF < 3$ ), the plot will implicitly give a perfect fit.

Click **Regression line** to see a perfect observed vs. predicted line.

Click **Next Y** to go to **Replicates** for the next response.

Click **Back** to return to **Residuals**.

## Specification

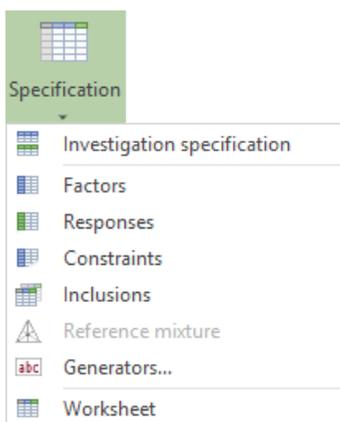
On the **Home** tab, click **Specification** to open the **Specification** window that shows views of the **Worksheet**, **Factors** and **Responses**.

Specification										
	1	2	3	4	5	6	7	8	9	
	Exp No	Exp Name	Run Order	Incl/Excl	Air	EGR%	Needlelift	Fuel	NO	
1	1	N1	10	Incl	241	6.1	-5.78	236.637	25	
2	2	N2	9	Incl	284	6	-5.78	254.722	26	
3	3	N3	5	Incl	241	12	-5.78	242.841	11	
4	4	N4	15	Incl	283	12.1	-5.78	262.626	12	
5	5	N5	6	Incl	241	6.2	0	217.069	29	
6	6	N6	14	Incl	283	6.1	0	230.527	32	
7	7	N7	17	Incl	242	12.1	0	221.274	13	
8	8	N8	7	Incl	284	11.8	0	236.25		
9	9	N9	1	Incl	241	9	-3.16	231.77	1	
10	10	N10	16	Incl	284	9.2	-3.16	248.064	19	

Worksheet Factors Responses

## Spreadsheet access

The arrow under **Specification** provides shortcuts to some of MODDE's spreadsheets. More info on these can be found in Chapter 6, Design. Information about adding, modifying and deleting factors, responses, etc, can be found in Chapter 3, Design wizard.

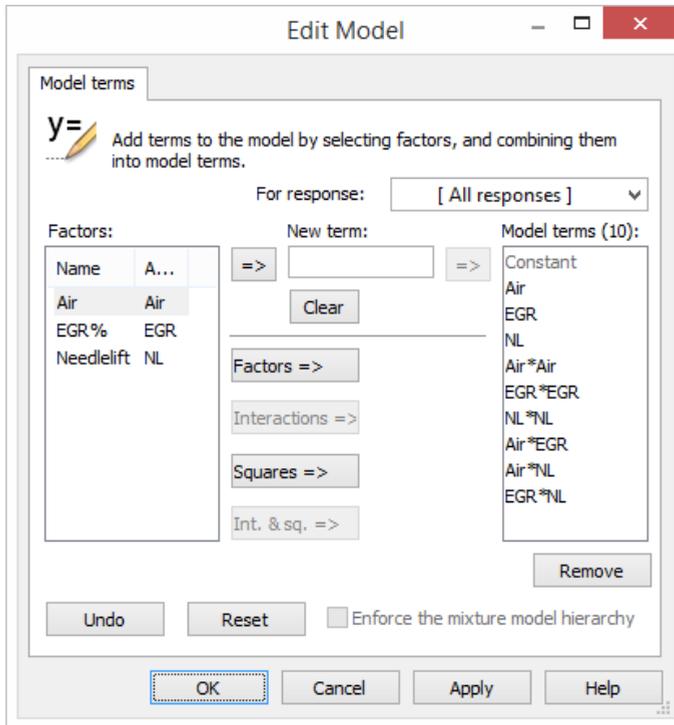


## Edit model

### Edit Model dialog box

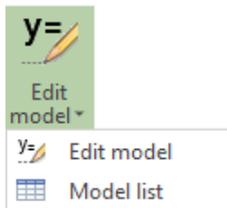
On the **Home** tab, in the **Model** group, click **Edit model** to open the **Edit Model** dialog box. Here you can view, add and remove model terms to the model for each or all responses.

Use the **For response** box to switch the response for which the model is displayed. [ *All responses* ] is available here to select all responses at once.



## Model list

On the **Home** tab, in the **Model** group, click the arrow under **Edit model** to open the **Edit model** menu.



Click **Model list** to open the **Model list**.

Model List							
	1	2	3	4	5	6	7
1		<b>Full Name</b>	<b>Abbreviation</b>	<b>Full Name</b>	<b>Abbreviation</b>	<b>Full Name</b>	<b>Abbreviation</b>
2	Responses:	Fuel		NOx		Soot	
3		Constant	Cst	Constant	Cst	Constant	Cst
4	1	Air	Air	Air	Air	Air	Air
5	2	EGR%	EGR	EGR%	EGR	EGR%	EGR
6	3	Needlelift	NL	Needlelift	NL	Needlelift	NL
7	4	Air*Air	Air*Air	Air*Air	Air*Air		
8	5	Air*Needlelift	Air*NL	EGR%*EGR%	EGR*EGR		
9	6			Needlelift*Ne	NL*NL		
10	7			EGR%*Needle	EGR*NL		

If you have qualitative factors at more than two levels and want to display the model with the qualitative factors extended:

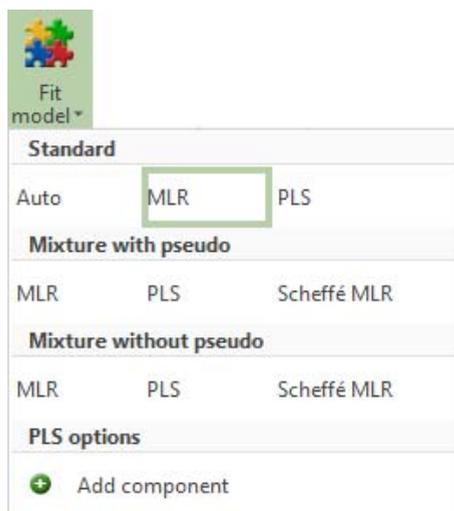
1. Right-click the **Model** list.
2. Click **Properties**.
3. Click **Extended – Show all settings**.

---

## Fit model

On the **Home** tab, in the **Model** group, click **Fit model** to fit or refit the model.

Click the arrow under **Fit model** to show the available fitting methods.



### Standard fit

When the investigation does not contain mixture factors, the only available fitting methods are standard.

#### Auto

Click **Auto** to use the default fit method.

MODDE defaults to using Multiple Linear Regression (MLR) as long as the condition number permits. When the condition number becomes too large, MODDE defaults to using PLS. You may override the default by selecting the desired fit on the **Home** tab, in the **Model** group; click the arrow under **Fit model** to open the **Fit Model** gallery.

---

*Note: If your X matrix has a condition number > 3000, MODDE will only fit the model with PLS and the condition number when selecting MLR is displayed as infinite.*

---

#### Fitting with MLR

When fitting with MLR, MODDE will separately but automatically fit all of the responses. Use the **Select responses** box to select the desired response.

If some response values are missing, MODDE excludes the rows with missing data for that specific response and keeps it for all others in the calculations.

## Fitting with PLS

With PLS all responses are fitted simultaneously.

PLS handles missing values in the responses, without excluding the runs from the analysis when the same model is used for all responses. When the models are not identical, the fit is done separately for each response and missing is handled as for MLR.

When fitting the model with PLS, MODDE computes as many PLS components as significant by cross validation. See the statistical appendices for significance rules.

To add more PLS components:

1. On the **Home** tab, in the **Model** group, click the arrow under **Fit model**.
2. Click **Add component**.

Once the model is fitted the commands to display results and perform diagnostics are available. Specific commands and features pertaining only to PLS are unavailable when fitting the model with MLR.

To exclude responses from the analysis, set their unit variance modifier to zero, in the response dialog box. This will give the responses zero variance, and hence exclude them from the analysis.

Note that with PLS the X matrix is always scaled and centered to unit variance. The centered responses are scaled as you selected in the response definition. The default is unit variance.

---

***Note:** The default method of fit with the Cox reference mixture model is PLS. When the model obeys mixture hierarchy you can if you want fit the model with MLR. When fitting the model with PLS, the condition number refers to the X matrix, with unit variance coding.*

---

## Mixture fit

A design containing mixture factors can be fitted either with or without pseudo components.

To fit the model containing mixture factors, on the **Home** tab, in the **Model** group, click the arrow under **Fit model**. Then select the desired fit option.

When you select a fit method with pseudo components, MODDE displays all mixture designs (the design matrix not the worksheet) with the mixture factors transformed to pseudo components. When the mixture region is a simplex, transforming to pseudo components gives all mixture factors the range 0 to 1. When the mixture region is not a simplex, pseudo components stretch the experimental region.

### Scheffé MLR

Click **Scheffé MLR** to fit the mixture data with a Scheffé type model. When you select this fit method, the model is restored to its default specification. Scheffé MLR is available when your investigation contains only mixture factors.

---

## Summary of fit

Summary plots and lists can be accessed on the **Home** tab. Click **Summary of fit** to open the **Summary of Fit Plot**. Alternatively click the arrow under **Summary of fit** in order to choose between various summary plots and lists.

Available plots and lists:

- Summary of Fit Plot
- PLS Total Summary Plot
- PLS Response Summary Plot
- Summary List
- PLS Summary List

---

*Note: PLS summaries are only available when fitting the model with PLS.*

---

### **Understanding the plot**

Summary statistics are presented in four parameters ( $R^2$ ,  $Q^2$ , Model validity, and Reproducibility) where 1, or 100%, is perfect.  $R^2$  and  $Q^2$ , the first two columns, should be close in size. The difference shouldn't be more than 20% in most situations.

#### **$R^2$**

Shows the model fit. A model with  $R^2$  of 0.5 is a model with rather low significance.

#### **$Q^2$**

Shows an estimate of the future prediction precision.  $Q^2$  should be greater than 0.1 for a significant model and greater than 0.5 for a good model. The difference between  $R^2$  and  $Q^2$  should also be smaller than 0.3 for a good model.  $Q^2$  is the best and most sensitive indicator.

#### **Model validity**

A test of diverse model problems. A value less than 0.25 for *Model validity* indicates statistically significant model problems, such as the presence of outliers, an incorrect model, or a transformation problem. A low value here may also indicate that a term, such as an interaction or square is missing. When the pure error is very small (replicates almost identical), the model validity can be low even though the model is good and complete. When the pure error is so small that the replicates are deemed identical by MODDE, the model validity is labeled *Missing* as it cannot be calculated.

---

*Note: Model validity might be low in very good models ( $Q^2 > 0.9$ ) due to high sensitivity in the test or extremely good replicates.*

---

#### **Reproducibility**

The variation of the replicates compared to overall variability. The *Reproducibility* should be greater than 0.5.

### **Improving summary statistics**

Summary statistics can be improved by tuning the model in the various steps of the **Analysis wizard**.

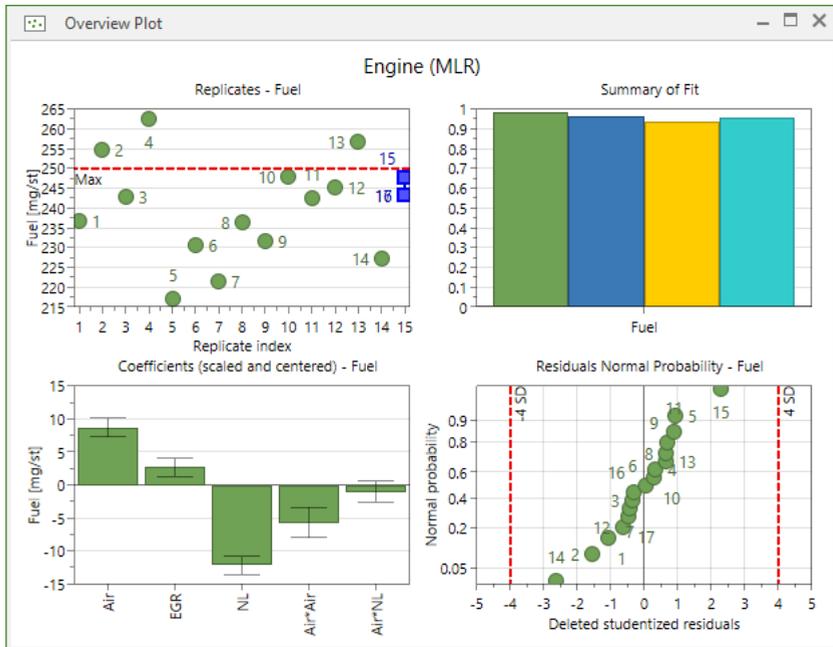
- Select the appropriate transformation in the Histogram.
- Remove non-significant model parameters in the Coefficient plot.

More information on the Summary of fit plots and lists can be found in Chapter 8, Analyze.

---

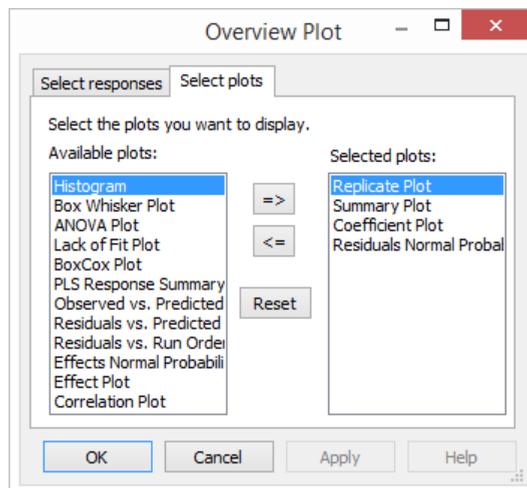
## **Overview**

On the **Home** tab, click **Overview** to open the **Overview Plot** window.



This window shows multiple graphs at once in order to better give an overview of a response. The plots shown can be changed.

1. Right-click the **Overview Plot** window.
2. Click **Properties**.
3. In the **Overview Plot** dialog box, click the **Selected plots** tab.



4. Use the arrows  to move plots back and forth between **Available plots** and **Selected plots**.
5. Click **OK** to return to the **Overview Plot**.

## Coefficients

There are two coefficient plots and two coefficient lists available in MODDE.

On the **Home** tab, in the **Diagnostics & interpretation** group, click the arrow under **Coefficients** to open the gallery.

See Chapter 8, Analyze for more information.

## Residuals

On the **Home** tab, in the **Diagnostics & interpretation** group, click **Residuals** to open the **Residuals Normal Probability Plot**.

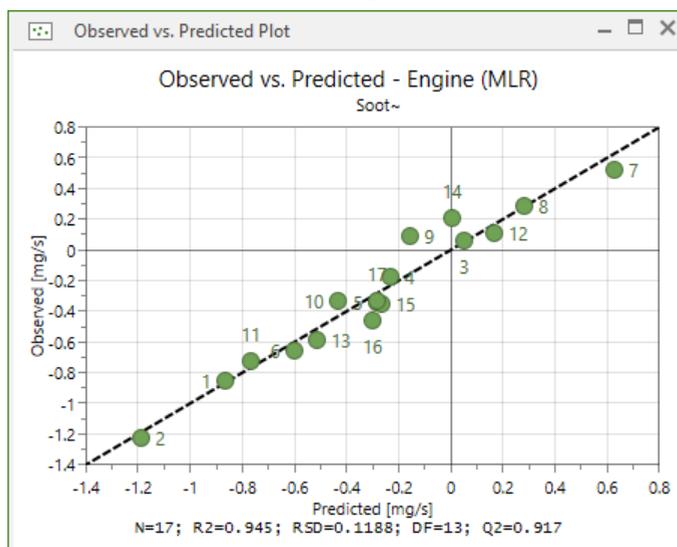
Other residuals plots and lists are available by clicking the arrow under **Residuals** and then clicking a plot or list.

See Chapter 8, Analyze for more information.

## Observed vs. predicted

On the **Home** tab, in the **Diagnostics & interpretation** group, click **Observed vs. predicted** to open the **Observed vs. Predicted Plot**.

The **Observed vs. Predicted Plot** displays observed values vs. predicted values.

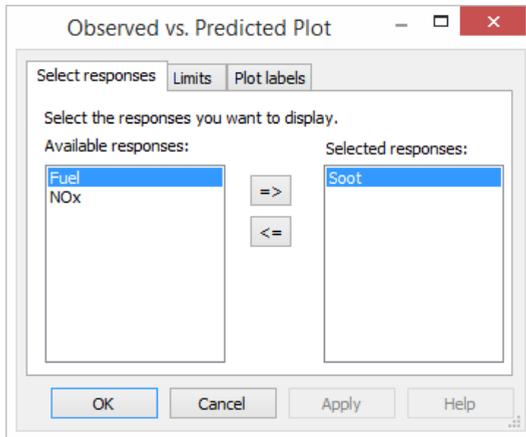


### Understanding the plot

Plots with points close to a straight line indicate good models. If Degrees of Freedom is under 3 ( $DF < 3$ ), the plot will implicitly give a perfect fit.

### Properties

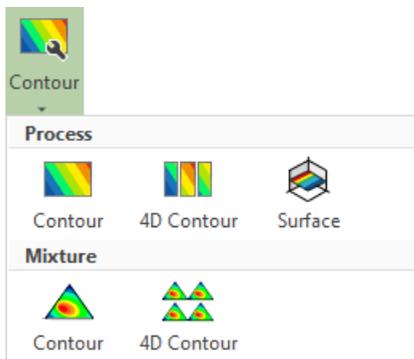
The properties dialog box of the **Observed vs. Predicted Plot** allows for the responses to be selected, limits to be shown or not, and what plot labels to display.




---

## Contour

On the **Home** tab, in the **Diagnostics & interpretation** group, click **Contour** to open a dialog box with settings for creating contour plots. Click the arrow under **Contour** to choose to show a specific contour plot directly.



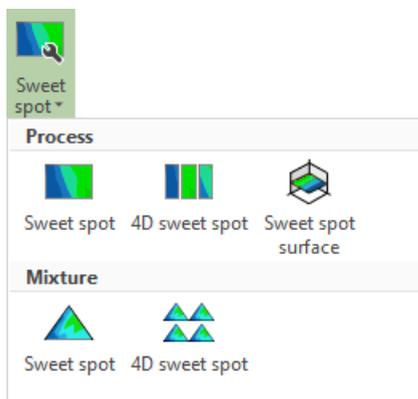
More information for each plot is available in Chapter 9, Predict.

---

## Sweet spot

The **Sweet Spot** plot highlights the areas where the responses are within the user specified ranges. The sweet spot plot can be displayed as 2D, 3D or 4D for process factors and as 2D or 4D for mixture factors.

On the **Home** tab, in the **Diagnostics & interpretation** group, click **Sweet spot** to open a dialog box with settings for creating a sweet spot plot. Click the arrow under **Sweet spot** to show a specific sweet spot plot directly.



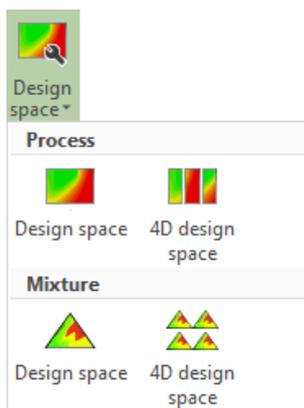
For more information, see the Sweet spot section in Chapter 9, Predict.

---

## Design space

The **Design space plot** is a type of sweet spot plot that shows the probability estimation.

On the **Home** tab, in the **Diagnostics & interpretation** group, click **Design space** to open the **Design Space Plot** dialog box with settings for creating a design space contour plot. Click the arrow under **Design space** to choose to show a specific design space plot directly.



For more information on the various design space plots, see Chapter 9, Predict.

---

## Optimizer

On the **Home** tab, in the **Diagnostics & interpretation** group, click **Optimizer** to open the **Optimizer** window and provide access to the **Optimizer** contextual tab.

More information about the optimizer is available in Chapter 12, Optimizer.

---

## Exclude

**Exclude** is located on the **Home** tab, in the **Editing** group.



Use **Exclude** to remove unwanted model terms or experiments from plots and charts. **Exclude** can either be clicked before selecting points or after selecting points. When clicked before selecting points, **Exclude** remains active allowing for the repeated removal of model terms.

Click **Undo** to read newly removed model terms or experiments.

## Undo

**Undo** is located on the **Home** tab, in the **Editing** group.



Click **Undo** or press Ctrl+Z to undo changes in MODDE. Such changes can be changing values or text (factor name for instance), cutting, copying, pasting, deleting, sorting in spreadsheets, or editing the model.

**Undo** is activated after changes in:

- The spreadsheets **Worksheet**, **Factors**, **Responses**, **Constraints**, **Inclusions**, **Prediction**, and **Optimizer**.
- The model terms in the **Edit Model** dialog box or by clicking **Exclude** and marking model terms to exclude in a plot.
- The experiments (observations) by clicking **Exclude** and marking experiments to exclude in a plot.

**Undo** remembers the ten last actions in the worksheet, in the **Edit Model** dialog box, and in plots displaying experiments (observations) or model terms.

**Undo** works on the active plot or spreadsheet.

***Note:** After editing the values in the worksheet the undo feature of plots displaying experiments stops working as manual changes in the worksheet empty the undo-memory for plots.*



# 06-Design

---

## Introduction

The **Design** tab provides access to the various worksheets that are part of the design such as factors, responses, constraints, and inclusions. D-Optimal design functions are also available on this tab (if the current design is D-Optimal) as well as functions concerning the objective of the investigation.



### Specification group

- **Factors** - opens the **Factors** spreadsheet with all factors.
- **Responses** - opens the **Responses** spreadsheet with all responses.
- **Constraints** - opens the **Constraints** window with the opportunity to define and modify constraints graphically.
- **Inclusions** - opens the **Inclusions** spreadsheet.
- **Reference mixture** - opens the **Reference mixture** spreadsheet.
- **Generators** - opens the **Generators** dialog box.
- **Objective** - goes to the **Select objective** page of the **Design wizard**.

### Show group

- **Design region** - opens the **Design Region** plot.
- **Design matrix** - opens the **Design Matrix**.
- **Design summary** - opens the **Design Summary** list, summarizing the design specifications.
- **Confoundings** - is available if confoundings are present in the investigation, opens the **Confoundings** list.

### D-Optimal group

- **Candidate set** - opens the **Candidate Set** spreadsheet with the discrete set of “all potential good runs.”
- **New design** - opens the **D-Optimal results** page of the **Design wizard**.
- **Onion** - provides access to the **Onion** and **Onion 3D** plot.

---

## Factors

Factors are variables that can be varied, or vary, during an experiment. Typical examples of factors are amount of raw material or temperature.

The **Factors** spreadsheet is located on the **Design** tab, in the **Specification** group, alternatively on the **Home** tab, under **Specification**.

In the **Factors** spreadsheet (window), you define (enter), modify, and delete factors. MODDE supports quantitative, qualitative, and mixture factors.

Quantitative factors may be used in a transformed metric. When factors are transformed, the design is created in the transformed units, but the worksheet is expressed in original units and so are the plots by default.

---

**Hint:** To set the units of a factor as °C (degrees Celsius), use Alt+0176 for the ASCII code. Hold down Alt, then press 0176 in succession, using the number pad on your keyboard.

---

### Factors spreadsheet

When a factor has been defined, the **Factors** spreadsheet provides an overview of the factor definitions, with one factor in each row. In the factor definition spreadsheet, the fields **Name**, **Abbr**, **Units**, and **Settings** can be modified directly by typing in the worksheet.

#### *Modify*

Modifying the settings when a design was created opens a dialog allowing you to:

- **Delete the current design** - the design and all settings and results are deleted.
- **Update factor scaling** - updates the factor scaling used in for example the calculations of coefficients and selection of axis length. Note that constraints are updated with the new setting.
- **Update factor settings in the worksheet for experiments not yet performed. Update factor scaling** - alters the factor settings in the worksheet for experiments not performed and also updates the factor scaling used.

To modify any of the other fields, double-click one of them and the **Factor Definition** dialog box opens. Factors can be added by double-clicking the last row of the factor definition spreadsheet, or right-clicking the spreadsheet and clicking **Add factor...**

#### *Copy*

To copy factors, mark the factors to copy, then press Ctrl+C and Ctrl+V. MODDE copies the factors and adds a digit after the name when pasting to make it unique.

---

**Hint:** You can copy factors in one investigation and paste them in another new investigation. Other transformations than Log cannot be pasted.

---

#### *Spreadsheet*

The factor columns that are present by default are,

- Name

- Abbr
- Units
- Type (quantitative, quantitative multilevel, qualitative, formulation, or filler)
- Settings (high and low vales displayed except for quantitative multilevel and qualitative where all levels are displayed)
- Transform
- Factor setting precision.

	Name	Abbr.	Units	Type	Settings	Transform	Precision
1	Air	Air	kg/h	Quantitative	240 to 284	None	4
2	EGR%	EGR	%	Quantitative	6 to 12	None	0.4
3	Needlelift	NL	BTDC	Quantitative	-5.78 to 0	None	0.4
		<i>Double-click here</i>	<i>to add</i>	<i>a new factor</i>			

More factor columns can be shown by changing the **Properties** of the **Factors** spreadsheet (right-click the spreadsheet and click **Properties**);

- Use (controlled, uncontrolled, or constant)
- No. of decimals
- MLR scale
- PLS scale

Click **Reset** in the **Properties** dialog box of the **Factors** spreadsheet to reset the **Factor columns** to MODDE's default.

The screenshot shows the 'Properties' dialog box with the 'Factor columns' tab selected. It contains a table with the following columns: 'Show or hide', 'Show Abbr.', 'Show Units', 'Show Type', 'Show Use', 'Show Settings', 'Show Transform', 'Show 'No. of decimals'', 'Show MLR scale', 'Show PLS scale', and 'Show factor setting precision'. The 'Reset' button is highlighted in blue.

Show or hide	
Show Abbr.	Yes
Show Units	Yes
Show Type	Yes
Show Use	No
Show Settings	Yes
Show Transform	Yes
Show 'No. of decimals'	No
Show MLR scale	No
Show PLS scale	No
Show factor setting precision	Yes

### Printing the factor spreadsheet

Print is available for the active plot or list. MODDE provides a variety of ways to print,

- Ctrl+P, the Windows keyboard shortcut for print,
- Right-click the desired plot or list and click **Print**,
- On the **File** tab, click **Print**.

## Responses

A response is the result from an experiment. A typical example of a response is yield.

On the **Design** tab, in the **Specification** group, click **Responses** to open the **Responses** spreadsheet. The **Responses** spreadsheet can also be opened on the **Home** tab, click **Specification**, then click **Responses**.

In the response definition spreadsheet, you define (enter), modify, delete, copy, print, and list responses. MODDE supports only quantitative responses.

Responses may be transformed, and MODDE supports several transformations.

*Note: For transformed responses predictions, contour plots and 3D surface plots are back transformed to original units.*

### Responses spreadsheet

When the responses are defined, the response spreadsheet provides an overview of the response definitions, with one response in each row. The response properties are name, abbreviation, unit, transformation, MLR Scale, PLS Scale, and type of response (regular or derived). A response is selected by clicking when pointing to it or by using the keyboard arrow keys to move in the spreadsheet.

The fields **Name**, **Abbr**, **Units**, **Min**, **Target**, and **Max** can be edited directly in the spreadsheet. To modify any other fields double-click one of them, or mark the response and press **Enter**, to open the **Response Definition** dialog box. **Type** cannot be modified after a response has been defined.

	Name	Abbr.	Units	Transform	Type	Min	Target	Max
1	Fuel	Fue	mg/st	None	Regular			250
2	NOx	NOx	mg/s	None	Regular			25
3	Soot	So	mg/s	Log: 10Log(Y)	Regular			1
		Double-click here to add a new response						

Responses can be added to the **Responses** definition spreadsheet by typing the new response information directly into the last row or by double-clicking the last row.

To copy responses, mark the responses to copy, then press Ctrl+C and Ctrl+V. MODDE copies the responses and adds a digit after the name when pasting to make it unique.

### Printing the response spreadsheet

Print is available for the active plot or list. MODDE provides a variety of ways to print,

- Ctrl+P, the Windows keyboard shortcut for print,
- Right-click the desired plot or list and click **Print**,
- On the **File** tab, click **Print**.

## Constraints

A common problem is that experimentation may not be possible in some region of the experimental space. For example it may not be possible to have high temperature and simultaneously low pH, and you want to cut-off the corner High temp, Low pH. In MODDE this is solved by adding a constraint.

A linear constraint is a function of the factors that specify a part of the experimental region to be included or excluded.

The resulting experimental region is an irregular polyhedron. The corners of this region are called the extreme vertices; they constitute part of the candidate set, i.e. a discrete set of potentially good runs.

D-Optimal designs are the only designs available when the experimental region is constrained to an irregular polyhedron.

Constraints can be defined for quantitative or formulation factors.

## Specifying constraints

Enter your constraints in the **Constraints** spreadsheet. To open the **Constraints** window, on the **Design** tab, in the **Specification** group, click **Constraints**. You can also access the **Constraints** on the **Home** tab; click **Specification**, then click the **Constraints** tab.

In the upper part of the spreadsheet, you define each constraint (one per row) as a mathematical relation. In the lower part, the graphical view, you can define constraints, to be added to the upper part, geometrically. Such constraints may include two factors only and are shown in the upper part after clicking **Add**.

	1	2	3	4	5	6	7	8
1		butane	isopent	reformt	catcrk	alkylat	Operator	Limit
2	Exclude:	1	1				>	0.3
3	Exclude:				1	1	>	0.7
4	Exclude:						<	

Graphical view for simple constraints. Click on a row in the above spreadsheet to see the constraint's graphical view.

Factor on the X-axis: Low: 0.15 High: 0.3 Exclude area:  Above line  Below line

Factor on the Y-axis: Low: 0.15 High: 0

Update constraint Add Help

## Defining a constraint graphically

The graphical interface in MODDE helps you define the constraints to exclude a region of the experimental space defined by the intersection of a *line* with the experimental region. Only two factor constraints can be defined graphically.

1. In the **Factor on the X-axis** and **Factor on the Y-axis** boxes select the two factors defining the constraint.
2. Define the coordinates of the extreme vertices (intersection of the line with the experimental region) or pull the end of the line along the side to select the region to cut off. When pulling, MODDE enters the current extreme vertices in **Low** and **High** of the selected X and Y-axis factors.
3. Under **Exclude area**, click **Above line** or **Below line** to exclude the correct area.

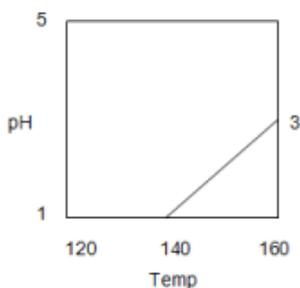
- Click **Add**. MODDE computes the equation of the line and enters the coefficients  $A_k$  of the two factors in the constraints spreadsheet in the upper part of the constraints window.

### **An example of entering a constraint graphically**

In an experiment with temperature and pH, temperature varies between 120 °C and 160 °C and pH between 1 and 5. You want to exclude the corner Temperature = 160 and pH = 1. Define the extreme acceptable conditions, that is the lowest pH when temperature is 160, for example pH = 3, and the highest temperature when pH = 1 for example temperature = 140.

These are the coordinates of the extreme vertices, the intersection of the line that cuts off the undesirable corner.

Enter these in the **Low** and **High** boxes, and the coefficients of the intersecting line are computed when you click **Add**.



### **Modifying a constraint graphically**

To modify a constraint graphically,

- Mark the row showing the constraint in the spreadsheet
- Change the constraint by pulling the end points of the line, or by modifying the values in **Low** and/or **High**
- Click **Update constraint** to update the constraint formula in the constraints spreadsheet.

---

**Hint:** Click a row in the spreadsheet defining a constraint in two factors and MODDE displays the graphical constraint.

---

### **Candidate set with a constraint**

When defining constraints the only designs available are D-Optimal designs. To create a D-Optimal design, a candidate set is created. When there is a constraint present, the resulting candidate set is formed by the extreme vertices of the irregular region, defined by the linear constraints.

If there in addition are qualitative or quantitative multilevel factors, the final candidate set is the product of the full factorial in the qualitative or quantitative multilevel factors times the candidate set resulting from the linear constraints (extreme vertices, center of edges, etc. of the irregular experimental region).

---

**Note:** To find undesirable factor combinations you need to create a D-Optimal design then sort the candidate set. On the **Design** tab, in the **D-Optimal** group, click **Candidate set**. Select a column to sort, then right-click the spreadsheet, and click **Sort**.

---

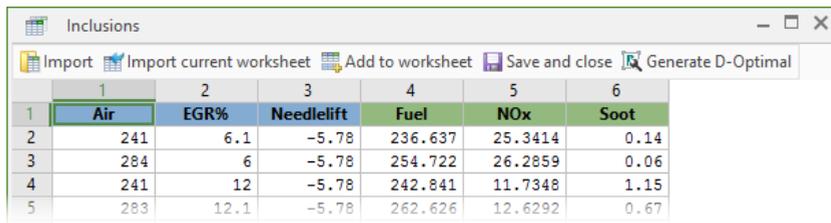
## Inclusions

In MODDE designs can be augmented using either **Complement design** or **Inclusions**.

Inclusions are extra runs that will be part of the worksheet. You can include a set of experimental runs (inclusions), either at the end of the worksheet or to be part of a D-Optimal design.

### Opening the inclusions spreadsheet

On the **Design** tab, in the **Specification** group, click **Inclusions** and add the experiments to add to the worksheet or design generation.



	1	2	3	4	5	6
1	Air	EGR%	Needlelift	Fuel	NOx	Soot
2	241	6.1	-5.78	236.637	25.3414	0.14
3	284	6	-5.78	254.722	26.2859	0.06
4	241	12	-5.78	242.841	11.7348	1.15
5	283	12.1	-5.78	262.626	12.6292	0.67

### Inclusions vs. complement design

Using **Inclusions** to augment a design is preferred when:

- The extra experiments to include are found in another investigation or in a text-file – Complement design can only complement the current investigation.
- The experiments were not saved in a MODDE investigation or a text file – In inclusions such experiments can be entered manually or pasted.
- When adding experiments after the design has already been created – This means that the inclusions should not be part of the design generation.

Using **Complement design** to augment a design is preferred when:

- The desired design should be a classical design – Use **Fold over** or **Estimate square terms in a screening design**. When using inclusions a D-optimal design is the only available choice.
- The desired design should include star points – Use **Estimate square terms in a screening design** and change the **Star distance**.

### Adding inclusions to the worksheet

Inclusions can be specified before or after the worksheet is generated.

If the worksheet already exists when you enter the inclusions, click **Add to worksheet** at the top of the **Inclusions** spreadsheet to add the inclusions last in the worksheet.

If you enter the inclusions before the generation of the worksheet, click **Save and close**, reopen the **Inclusions** spreadsheet after creating the worksheet and click **Add to worksheet**.

If the inclusions are entered before the generation of a D-Optimal design, the **Include in design** check box on the D-Optimal page of the design wizard has to be cleared to avoid including the inclusions. Then after generating the D-Optimal design the inclusions can be added to the worksheet.

---

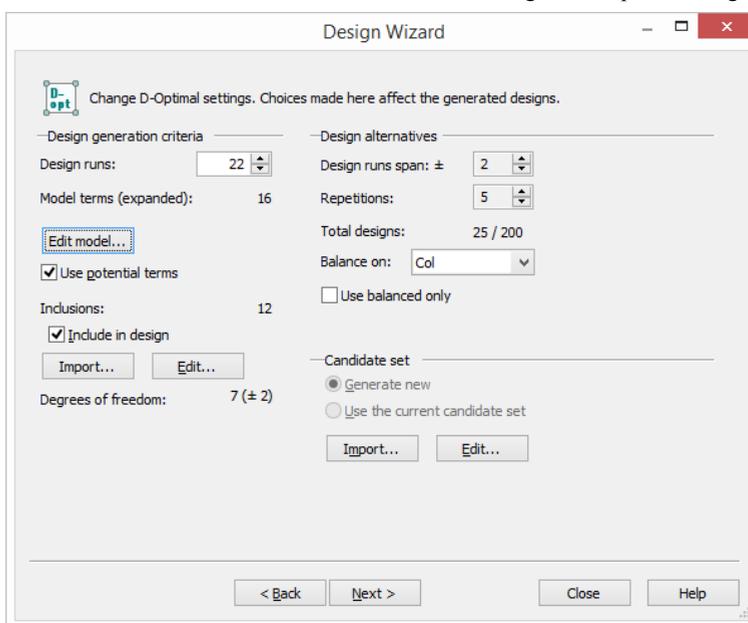
*Note: The inclusions are added to the worksheet only when you click **Add to worksheet**.*

---

## Inclusions as part of the design

With D-Optimal designs inclusion runs can be a part of the design or added at the end of the worksheet.

If the inclusions are entered before the generation of a D-Optimal design the **Include in design** check box on the D-Optimal page of the **Design wizard** is by default selected and the inclusions are then used when creating the D-Optimal design.




---

*Note: When generating D-Optimal designs, and the **Include in design** check box is selected, the inclusions are a part of the design and included in the number of runs.*

---

## Editing inclusions

Inclusions can be edited/imported as follows:

- Edit or specify new inclusions by pasting or typing values directly in the spreadsheet.
- **Import** - Import a tab separated text-file or another MODDE investigation with the same factors as the current investigation. When importing, all the factors defined in the MODDE investigation have to be present in the file, including uncontrolled, filler, and constant factors.
- **Import current worksheet** - The current worksheet is added to inclusions.

- **Add to worksheet** - The current inclusions content is added last in the worksheet.
- **Save and close** - The current inclusions content is saved and the Inclusions window is closed.
- **Generate D-Optimal** - Saves the current inclusions and opens the Design Wizard on the Change D-Optimal settings-page.

To delete rows, mark them, press the **Delete** key or right-click the spreadsheet and click **Delete**.

## Reference mixture

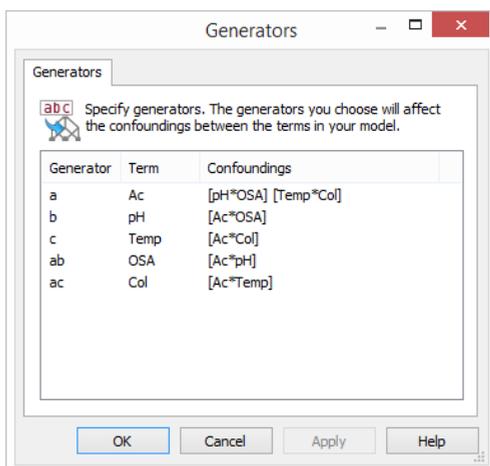
If the current investigation includes mixture factors, then the **Reference mixture** spreadsheet is available.

To access the **Reference mixture** spreadsheet, on the **Design** tab, in the **Specification** group, click **Reference mixture**. Another method to access this spreadsheet is on the **Home** tab, click **Specification**, then click **Reference mixture**.

	1	2	3	4	5
1	butane	isopent	reformt	catcrk	alkylat
2	0.05	0.15	0.2	0.3	0.3

## Generators

To access the **Generators** dialog box, on the **Design** tab, in the **Specification** group, click **Generators**. Alternatively on the **Home** tab, click **Specification**, and then click **Generators**.



A generator is a column of signs in the extended design table of the basic factors. It is used to introduce additional factors in the fractional factorial designs.

For example, let us assume that 5 factors are to be investigated in 8 runs. The extended design table is the table of the full factorial in three factors (basic factors), symbolically named a, b and c plus the additional columns for all the interactions. Any interaction column can be used to introduce additional factors. Let us say that to introduce the 2 additional factors, d, and e, the column of signs of a\*b and a\*c are selected. Then d = ab and e = ac are the generators of the fractional factorial design  $2^{5-2}$  (see Box, Hunter and Hunter for further information).

When MODDE generates fractional factorial designs the default generators used are those published in Box, Hunter and Hunter.

Editing and/or changing the default generators of a design is done in order to estimate selected interactions in a fractional factorial design of resolution III or IV instead of the default.

MODDE supports the choice of positive or negative generators.

To edit a generator, click in the generator column of the desired row, and enter a new generator. The confounding, in the **Confoundings** column, is updated.

When you click **OK**, your design and worksheet are deleted and new ones are generated.

---

## Objective

The **Objective** is the purpose for creating the design. MODDE recognizes two objectives: **Screening** (first stage of an investigation when little is known) and **Optimization (RSM)** (optimization with the important factors.). The **Split objective** supports both screening and optimization, as does **Paste data**.

After defining your factors and responses, on the **Design** tab, in the **Specification** group, click **Objective**. You can also click **Design Wizard** on the **Home** tab, in the **Quick start** group.

This opens the objective page of the **Design wizard** which guides you through the selection of objective, design, and model of the investigation. For more information about the objective page of the **Design wizard**, see Chapter 3, Design wizard.

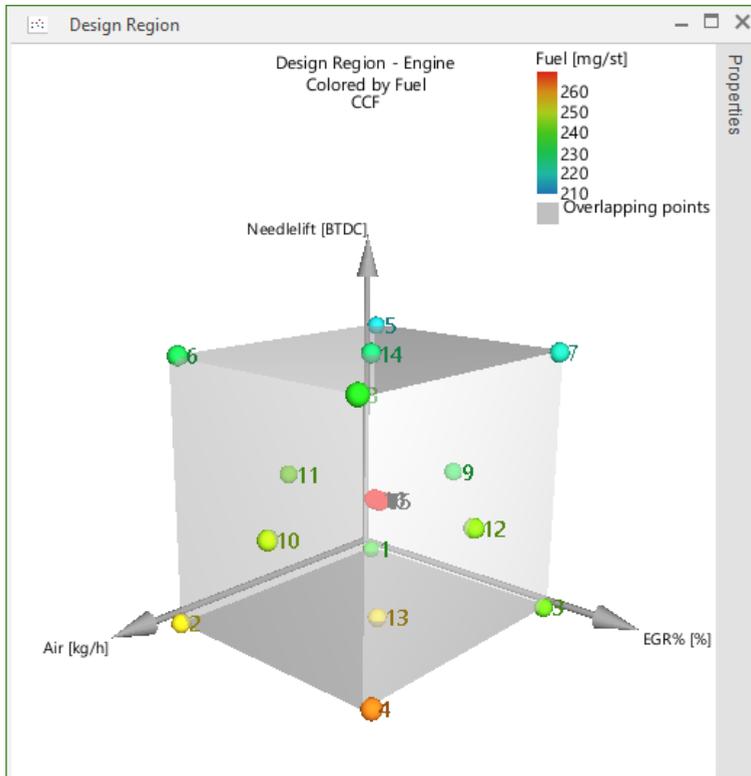
---

## Design region

To display the **Design Region**, on the **Design** tab, in the **Show** group, click **Design region**.

The plot gives an overview of your experimental plan.

Use the **Select responses** box or right-click the plot and click **Properties** to select a response to color by.



The plot displays the experiments as listed in the worksheet. This means that all experiments with factor settings are displayed, whether they have response values or not.

*Note: The design region plot is illustrative for designs with 3 factors. With 4 or more factors, the factors held constant limit the points possible to be displayed.*

## Design region properties

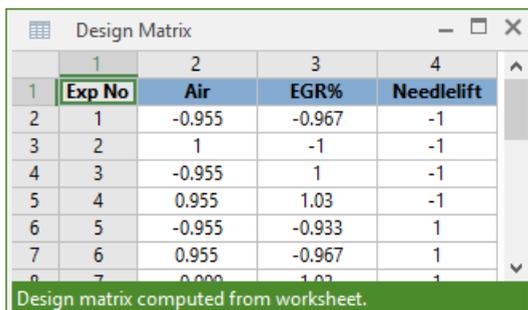
The Design Region plot:

- Is displayed as a cube where all other factors are held constant.
- Can be displayed for all designs.
- Can display the points in the design region plot color coded according to the response values entered in the worksheet.
- Displays a gray point, when more than one point is placed on exactly the same position. These points are called *overlapping points*.
- Cannot display qualitative and mixture factors on the axes.

## Design matrix

To display the **Design Matrix** window, on the **Design** tab, in the **Show** group, click **Design matrix**.

The **Design Matrix** displays the experimental plan in coded unit for quantitative factors, as in the worksheet for qualitative factors and in pseudo components for formulation factors.



	1	2	3	4
1	Exp No	Air	EGR%	Needlelift
2	1	-0.955	-0.967	-1
3	2	1	-1	-1
4	3	-0.955	1	-1
5	4	0.955	1.03	-1
6	5	-0.955	-0.933	1
7	6	0.955	-0.967	1
8	7	0.955	1.03	1

Design matrix computed from worksheet.

If you have qualitative factors at more than two levels, you can select to display the design with the qualitative factors extended.

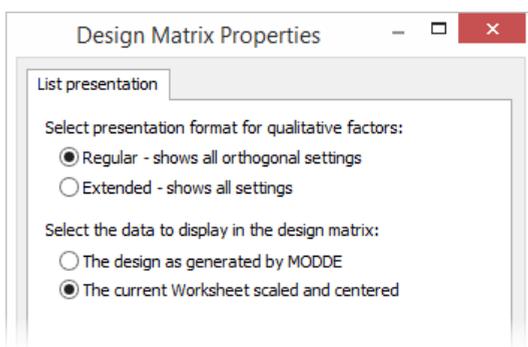
To display the design matrix with the qualitative factors extended:

1. Right-click the matrix and click **Properties**.
2. Click **Extended – shows all settings** and click **OK**.
3. Optionally switch back by clicking **Regular – shows all orthogonal settings**.

By default the **Design Matrix** is derived from the worksheet, and reflects any changes (excluded runs, changed values, additional runs etc..) done to the factor part of the worksheet after its generation.

To display the original design matrix generated by MODDE:

1. Right-click the matrix and click **Properties**.
2. Click **The design as generated by MODDE** and click **OK**.
3. Optionally switch back by clicking **The current Worksheet scaled and centered**.




---

**Note:** The defaults for the design matrix can be changed in Investigation options. For stability testing designs and generalized subset designs, the properties apply to the All page only.

---

## Design matrix for Stability testing

To display the **Design Matrix** window, on the **Design** tab, in the **Show** group, click **Design matrix**.

The Design Matrix for Stability testing designs has several tabs, the *Overview* and *All* tabs displaying all experiments in different layouts and then the tabs with the individual designs for each time.

### Overview tab

In the *Overview* tab each experiment is displayed as cross in a grid defined by the combination ID and the time. Each time point is connected to the name of the design that will be run at that time point. These sub designs are denoted A, B:1, B:2, C:1, etc.

	1	2	3	4	5	6	7	8	9
1	Combination ID	0	3	6	9	12	18	24	36
2		A:1	C:1	C:2	C:3	A:1	B:1	B:2	A:1
3	c1	✓			✓		✓		✓
4	c2	✓			✓	✓		✓	✓
5	c3	✓	✓			✓		✓	
6	c4	✓		✓		✓	✓		✓
7	c5	✓			✓	✓	✓		✓
8	c6	✓	✓			✓		✓	✓
9	c7	✓			✓	✓		✓	✓
10	c8	✓	✓			✓	✓		✓
11	c9	✓		✓		✓	✓		✓
12	c10	✓			✓	✓		✓	✓
13	c11	✓	✓			✓		✓	✓

### All tab

The *All* tab displays the design matrix in the same form as for a regular design. This page has a property page with options, see Design matrix for details.

	1	2	3	4	5
1	Exp No	Combination ID	Strength	Batch	Prim Pack
2	1	c1	-1	A	A
3	2	c2	1	A	A
4	3	c3	-1	C	A
5	4	c4	1	B	D
6	5	c5	-1	C	A
7	6	c6	1	C	A
8	7	c7	-1	A	B
9	8	c8	1	A	B
10	9	c9	-1	B	A

### A, B, C etc. tabs

The remaining tabs each list the sub design at each *Time* in the order the experiments are listed in the worksheet, that is, the order of the *Experiment number*.

	1	2	3	4
1	Combination ID	Strength	Batch	Prim Pack
2	c13	-1	A	C
3	c3	-1	B	A
4	c21	-1	B	D
5	c11	-1	C	B
6	c8	1	A	B
7	c16	1	B	C
8	c6	1	C	A
9	c24	1	C	D

Overview All A:1 B:1 B:2 C:1 C:2 C:3

## Design summary

To display the **Design Summary** list, on the **Design** tab, in the **Show** group, click **Design summary**.

The **Design Summary** list displays the selections made in the objective pages of the **Design wizard**.

	1	2
1	Objective	Optimization (RSM)
2	Process model	Quadratic
3	Mixture model	--
4		
5	Design	CCF
6	Runs in design	14
7	Center points	3
8	Replicates	0
9	N = actual runs	17
10	Maximum runs	12000
11	Constraints	No

The following information is listed for a classic design:

**Objective:** the selected objective.

- Process model: the type of model created for the process factors.
- Mixture model: the type of model created for the mixture factors.

**Design:** the selected design.

- Runs in design: the selected number of runs in the design created by MODDE. If you have added runs after creating the design they will not be included here.
- Center points: the number of center points selected.
- Replicates: the number of times the entire design has been replicated.
- N = actual runs: the number of runs created by MODDE.
- Maximum runs: the maximum number of runs that MODDE can include in a design created by MODDE.

- Constraints: *Yes* if constraints are present in the design.

---

## Design summary D-Optimal

To display the **Design Summary** list, on the **Design** tab, in the **Show** group, click **Design summary**.

The **Design Summary** list displays the selections made in the objective pages of the **Design wizard**.

When using a D-Optimal design, the **Design Summary** list additionally displays information concerning the candidate set and D-Optimal statistics.

### Candidate set

- Number of extreme vertices.
- Number of edge points.
- Number of centroids of high dimension surfaces.
- Total runs.

### D-Optimal

- Potential terms: The type of potential terms when selected; empty when the **Use potential terms** check box was cleared.
- Number of inclusions.
- Constraints: *Yes* if constraints are present in the design.
- Selected design number.
- Design statistics,
  - G-efficiency.
  - $\log(\text{Det. of } X'X)$ .
  - Norm.  $\log(\text{Det. of } X'X)$ .
  - Condition number.

	1	2	3
1	<b>Objective</b>	<b>Split Objective</b>	
2	Process model	Linear	
3	Mixture model	Quadratic	
4			
5	<b>Design</b>	<b>D-Optimal</b>	
6	Runs in design	17*	
7	Center points	3	
8	Replicates	0	
9	N = actual runs	20*	
10	Maximum runs	12000	
11	Constraints	No	
12			
13	<b>Candidate set</b>		
14	Extreme vertices	64	
15	Edge points	128	
16	Centroids of high dim. surfaces	97	
17	Total runs	289	
18			
19	<b>D-Optimal</b>		
20	Potential terms	Interaction	Quadratic
21	Number of inclusions	0	
22	Constraints	No	
23	Selected design number	22	
24	Design statistics	G-efficiency	71.8010254
25		log(Det. of X'X)	10.0177546
26		Norm. log(Det. of X'X)	-0.276978135
27		Condition number	7.66150236

\*) The number of runs shown is the specified number, not the actual number

When using a D-Optimal Onion design, the **Design Summary** list gives very similar information to the D-Optimal some additions,

- Design
  - Number of layers.
  - Path the scores were imported from.
  - SIMCA model name.
- Candidate set
  - Path the candidate set was imported from.

---

## Design summary, GSD and Stability testing

To display the **Design Summary** list, on the **Design** tab, in the **Show** group, click **Design summary**.

The **Design summary** list displays the selections made in the objective pages of the **Design wizard**.

With Generalized subset designs and Stability testing design, the **Design Summary** list additionally displays information about the different design sets.

### Design set

- Runs in design set.

- Center points.
- Replicates.
- N = actual runs.
- Condition number: condition number calculated using only this design set.
- Balanced: Equal number of runs for all factors not counting the replicates; *Yes* or *No*.
- OA: Orthogonal array; can be *Yes* or *No*.
- NOA: Near orthogonal array; can be *Yes* or *No*.
- #Dist1: Number of distances between two points that equal 1.
- #Dist2: Number of distances between two points that equal 2.
- #Dist3: Number of distances between two points that equal 3.
- #Dist4: Number of distances between two points that equal 4.

	1	2
1	<b>Objective</b>	<b>Screening</b>
2	Process model	Linear
3	Mixture model	--
4		
5	<b>Design</b>	<b>Stability</b>
6	Runs in design	120
7	Center points	0
8	Replicates	0
9	N = actual runs	120
10	Maximum runs	12000
11	Constraints	No
12		
13	<b>Design set</b>	<b>A:1</b>
14	Runs in design set	24
15	Center points	0
16	Replicates	0
17	N = actual runs	24
18	Condition number	2
19	Balanced	Yes
20	OA	Yes
21	NOA	Yes
22	#Dist1	46
23	#Dist2	58
24	#Dist3	24
25	#Dist4	20
26	<b>Design set</b>	<b>B:1</b>
27	Runs in design set	12

## Confoundings

Open the **Confoundings** list to see which terms that are mathematically identical in the current design. For instance, in the example below the term Ad\*Te is included in the model, but the effect of this term is confounded with the effect of St\*H2O. This means that using this design there is no way of telling whether the coefficient displayed for Ad\*Te reflects Ad\*Te, St\*H2O, or a mixture of both.

### To list the confoundings

On the **Design** tab, in the **Show** group, click **Confoundings**.

For factorial designs resolution III or IV, the **Confoundings** list displays the confounding pattern for the complete interaction model.

In the **Term** column, the background of the terms included in the model is colored.

In the coefficient plots and lists, confounded terms are marked with a bracket ("#").

	1	2	3
	<b>Term</b>	<b>Generator</b>	<b>Confounded with</b>
2	Ad	a	
3	St	b	
4	Te	c	
5	H2O	abc	
6	Ad*St		Te*H2O
7	Ad*Te		St*H2O
8	Ad*H2O		St*Te
9	St*Te		Ad*H2O
10	St*H2O		Ad*Te
11	Te*H2O		Ad*St

## Candidate set

	1	2	3	4	5	6
	<b>Design Run</b>	<b>AcN</b>	<b>pH</b>	<b>Temp</b>	<b>OSA</b>	<b>Column</b>
	center	26	4	21.5	0.1	ColB
1		25	3.8	18	0.09	ColA
2		27	3.8	18	0.09	ColA
3	1	25	4.2	18	0.09	ColA
4		27	4.2	18	0.09	ColA
5	2	25	3.8	25	0.09	ColA
6		27	3.8	25	0.09	ColA
7		25	4.2	25	0.09	ColA
8	3	27	4.2	25	0.09	ColA

D-Optimal designs are constructed by selecting N runs from a candidate set. This candidate set is the discrete set of “all potential good runs”.

MODDE generates the candidate set as follows:

**I)** For a regular process region, the candidate set consists of one or more of the following sets of points (depending on your model and the number of factors):

- The full factorial for up to 10 factors, reduced factorial for up to 32 factors.
- Centers of edges between hyper-cube corners.
- Centers of the faces of the hyper-cube.
- Overall centroid.

**II)** For constrained regions of mixture and/or process factors, the candidate set consists of one or more of the following set of points:

- The extreme vertices of the constrained region.
- The centers of the edges. If these exceed 200, the centers of the 200 longest edges.
- The centers of the various high dimensional faces.
- The overall centroid.

MODDE has implemented an algorithm to compute the extreme vertices, center of edges, center of faces etc. as described by Piepel (1988).

### Accessing the Candidate set

To access the candidate set, on the **Design** tab, in the **D-Optimal designs** group, click **Candidate set**.

## New design



On the **Design** tab, in the **D-Optimal designs** group, click the arrow under **New design** to display the menu.

See the D-Optimal pages of Chapter 3, Design wizard for more information.

### New D-Optimal design

Click **New design** to open the D-Optimal settings page of the Design wizard.

### Select from already generated D-Optimal designs

Click **New design** | **Select from already generated D-Optimal designs** to open the D-Optimal results page of the Design wizard.

# Onion

On the **Design** tab, in the **D-Optimal designs** group, click **Onion** to open the onion plot gallery.

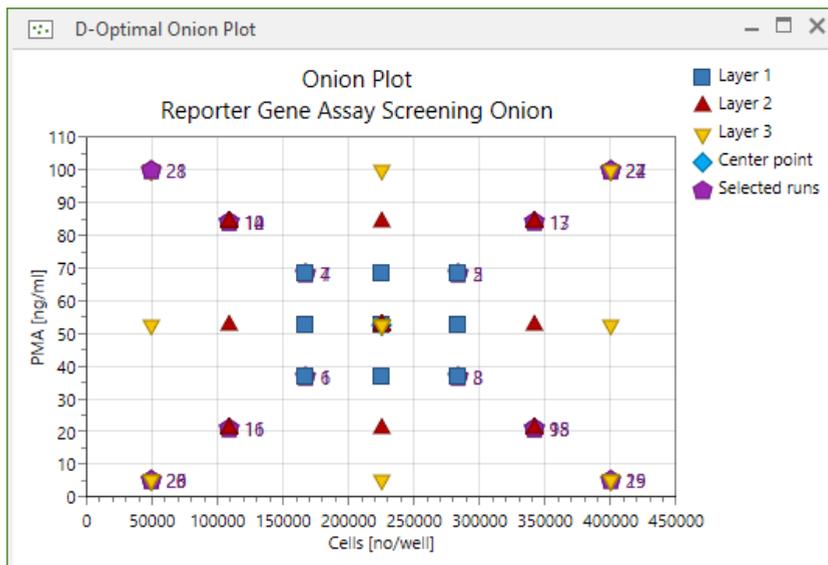


There are two onion plots to visualize the candidate set and the selected D-Optimal design, the Onion plot and the Onion 3D scatter.

For the onion scatter plots, the property page has two tabs: **Select factors** and **Plot labels**. Use **Select factors** to select which factors to display on the X, Y, and Z-axes. Use **Plot labels** to select which labels to display in the plot.

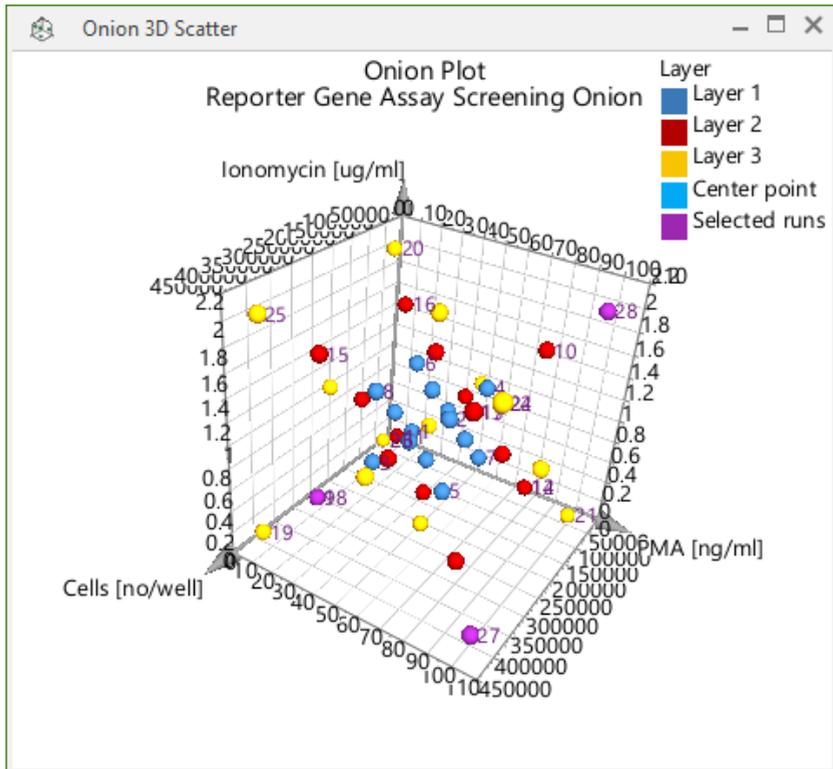
## Onion Plot

The **D-Optimal Onion Plot** is a 2D scatter plot of the candidate set. The candidate set runs are colored by layer, and the selected design runs are colored differently according to the legend.



## Onion 3D Plot

The **Onion 3D Scatter** plot displays the candidate set colored by layer with the selected design runs in a different color according to the legend.



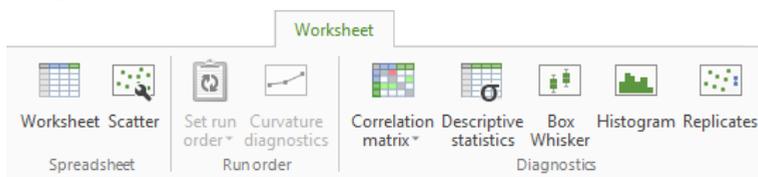


# 07-Worksheet

---

## Introduction

The **Worksheet** tab contains a number of functions for manipulating and working with the worksheet that has been generated. These are divided into three groups in MODDE, the **Spreadsheet**, **Run order** and **Diagnostics** groups.



### Spreadsheet group

- **Worksheet** - shows the worksheet itself, allowing for data to be seen or changed.
- **Scatter** - opens a dialog box to allow for the creation of a 2D or 3D scatter plot.

### Run order group

- **Set run order** - allows the experiment run order to be changed or randomized.
- **Curvature diagnostics** - facilitates detection curvature.

### Diagnostics group

- **Correlation matrix** - shows correlation coefficients as either a plot or a matrix.
- **Descriptive statistics** - summarizes the descriptive statistics for all responses.
- **Box Whisker** - illustrates how the response values are distributed around the response mean.
- **Histogram** - shows the response distribution and is used to determine if a transformation is needed.
- **Replicates** - shows the variation in results for all experiments.

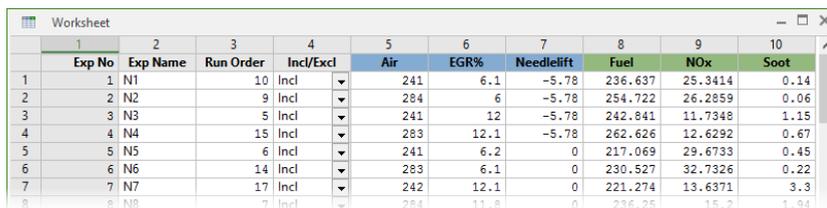
---

## Worksheet

The **Worksheet** presents a summary of all the data entered into the investigation in the form of factors and responses.

## Accessing the Worksheet

On the **Worksheet** tab, click **Worksheet**.



	1	2	3	4	5	6	7	8	9	10
	Exp No	Exp Name	Run Order	Incl/Excl	Air	EGR%	Needlelift	Fuel	NOx	Soot
1	1	N1	10	Incl	241	6.1	-5.78	236.637	25.3414	0.14
2	2	N2	9	Incl	284	6	-5.78	254.722	26.2859	0.06
3	3	N3	5	Incl	241	12	-5.78	242.841	11.7348	1.15
4	4	N4	15	Incl	283	12.1	-5.78	262.626	12.6292	0.67
5	5	N5	6	Incl	241	6.2	0	217.069	29.6733	0.45
6	6	N6	14	Incl	283	6.1	0	230.527	32.7326	0.22
7	7	N7	17	Incl	242	12.1	0	221.274	13.6371	3.3
8	8	N8	7	Incl	284	11.8	0	236.25	15.2	1.94

## Description of the worksheet

**Experiment number (Exp No)** - starts with the number one then assigned sequentially. *This column cannot be edited.*

**Experiment name (Exp Name)** - MODDE assigns a default experiment name, in the form Nxx, where xx is the experiment number. You may edit the name and enter your own identification.

**Run order** - order in which the experiments should be performed. MODDE suggests a randomized run order. Sort the worksheet according to run order before performing the experiments.

**Include or exclude (Incl/Excl)** - indicates if the experiment is included or excluded from the analysis. When the worksheet is generated all experiments are marked **Incl** and are included in the analysis. To exclude a failed experiment from the analysis, select **Excl** in the worksheet.

---

**Note:** *Excluded rows are excluded from the analysis for all responses. To exclude the response value for only one response, right-click the cell and click **Exclude values**.*

---

**Factors:** In the columns to the right of the Incl/Excl column the factors are listed in original units.

**Blocking:** When you have selected blocks in the **Select model and design page** of the design wizard, the column *\$BlockV* displays which block each experiment has been assigned to. For details on blocking, see the **Orthogonal blocking** section in the Statistical appendix.

**Responses:** In the columns to the far right all responses are found. The response values are listed in original units.

## Excluding qualitative settings

In Generalized Subset Designs it is common to start with one design set. This means that the qualitative factor Design set will be constant for that factor. MODDE then automatically excludes the factor and when you add design sets you have to manually add the Design set factor to the model.

This feature is general to all qualitative factors and all designs.

## Missing values in the worksheet

When responses have missing values in the analysis, MODDE creates individual models for all responses excluding only the row with the missing value for the relevant response.

Missing values in controlled factors are not allowed.

---

*Note: Missing values and excluded values are handled identically for MLR but not for PLS. PLS fits a common model when there are missing values but individual models when there are excluded values.*

---

## Deleting the worksheet

You cannot delete the worksheet; MODDE will automatically delete the worksheet when you make modifications to the factors.

## Adding experiments in the worksheet

To add extra runs to the worksheet:

- Open the **Worksheet** window, right-click and click **Add experiment**.
- On the **Design** tab, click **Inclusions**. Type the values or import the experimental runs you want added to the worksheet. Then click **Add to worksheet**. The inclusions will be added at the end of the worksheet.

The first four columns of the worksheet are automatically filled in when adding experiments.

## Sorting the worksheet

To sort a column in the worksheet:

1. Right-click the desired column,
2. Click **Sort ascending**, **Sort descending**, or **Custom sort...**

For more, see the Sorting spreadsheet subsection later in this chapter.

## Colors in the worksheet

Suspicious values in the worksheet are colored in red. Non-transformable values have a red background. When a colored cell is marked, the status bar will display a message about why it is colored.

The RED-MUP Worksheet response spreadsheet has special coloring, see the RED-MUP Worksheet subsection in Chapter 4, File.

---

## Sorting spreadsheet

Sorting is available for the worksheet, constraints, inclusions, candidate set, and prediction spreadsheet.

To sort a spreadsheet, right-click and click **Custom sort**, **Sort ascending**, or **Sort descending**. Sort ascending and descending immediately sorts the spreadsheet on the current column while **Custom sort** opens the **Sort Worksheet** dialog.

### Custom sort

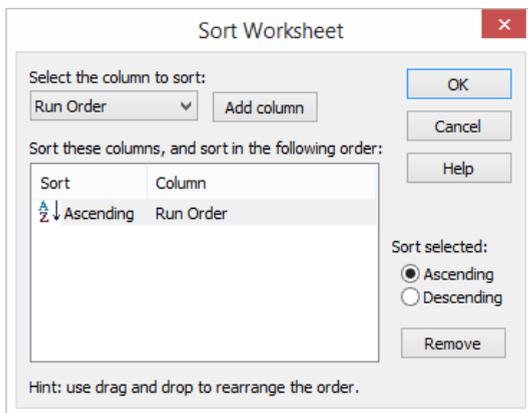
In the **Sort Worksheet** dialog box, select the columns to sort by from the **Select the column to sort** box and click **Add column**. The column appears in the list with the default sort type. Click the sort order you want under Sort selected, **Ascending** or **Descending**.

If a column of the spreadsheet is marked when sort is activated it becomes, by default, the primary column to sort by.

Add more columns to select secondary and tertiary etc., columns to sort by. Specify for each column the sort order.

Use **Remove** to remove a column from the sort list or drag the item outside the list.

The sorting starts when **OK** is clicked.



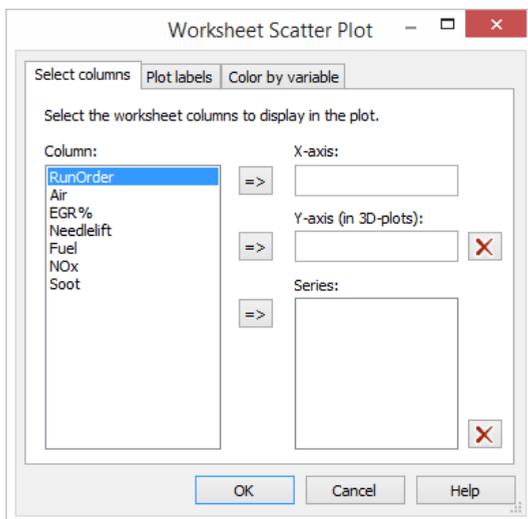
### Sorting the candidate set

Sorting the candidate set is useful if you want to exclude some rows that correspond to undesirable runs.

A candidate set imported to create an onion design cannot be sorted.

## Scatter

On the **Worksheet** tab, in the **Diagnostics** group, click **Scatter** to open the **Worksheet Scatter Plot** dialog box.



To add a variable to an axis or series, click the variable in the **Column** field, then click the relevant arrow . Click  to remove factors or responses from the **Y-axis** or from **Series**. To change what to display on the **X-axis**, mark a new variable and click .

For **2D** plots, select the variable to be plotted on the **X** and **Series (Z)** axes.

For **3D** plots, select the variables to be plotted on the **X**, **Y**, and **Series (Z)** axes.

Click the **Color by variable** tab to color by a factor or a response.

## Run order

To randomize the run order in different manners, on the **Worksheet** tab, click the **Set run order** or click the arrow under **Set run order** and then **Randomize run order to detect curvature**.

### Randomize run order

With RSM designs and Screening designs at more than 2 levels, *randomized* is the order in which you should perform the experiments and this is also the only available order for these cases.

To re-randomize the run order after the worksheet has been created but before performing any experiments, on the **Worksheet** tab, click **Set run order**.

Randomizing the run order is done to avoid that the effect of external variability, such as room temperature or who performs the experiment, coincides with the effect of a factor. By default the worksheet is randomized.

*Note: The run order can only be randomized for experiments with no results in the worksheet.*

### Run order to detect curvature

With 2 level screening designs, it is desirable to guard against strong curvature in the response due to overly wide ranges in the factors. Strong curvature in the response masks the effect of the factors. In this situation randomizing to detect curvature should be done before performing any experiments.

To detect curvature, one should first perform the following experiments:

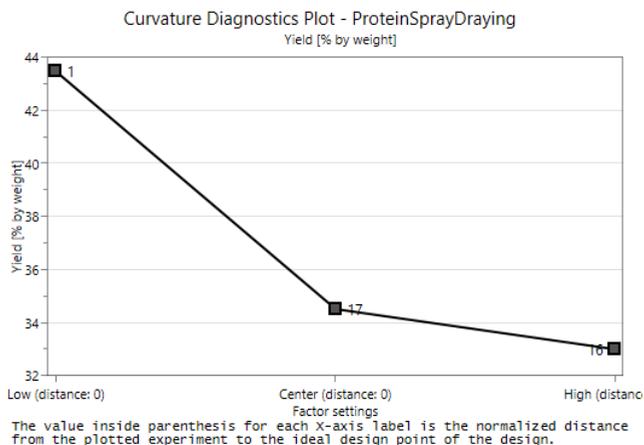
- A center point
- A point with as many factors as high (+) as possible
- A point with as many factors as low (–) as possible.

After clicking **Randomize run order to detect curvature** you will find the three points are given the run order numbers of 1-3 and the rest of the runs are randomized.

## Curvature diagnostics plot

Once the experiments are performed and the results are entered in the worksheet it is possible to examine the curvature diagnostics plot.

To open the plot, on the **Worksheet** tab, click **Curvature diagnostics**.



### Strong curvature

If the **Curvature Diagnostic Plot** exhibits strong curvature as is shown above, you should first re-measure the center point and re-do the plot. If the plot still exhibits strong curvature, drastically reduce the ranges of the factors and restart the project. You can also change the objective to RSM to try and reduce the curvature.

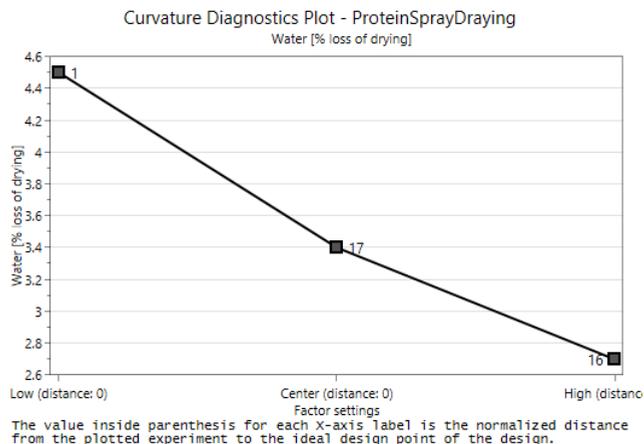
---

*Note: In the case of a strong interaction, it may falsely appear that there is strong curvature.*

---

### No curvature

If the **Curvature Diagnostic Plot** does not exhibit curvature, as the one below, continue performing the rest of the experiments.



## Correlation

The **Correlation plot** and **Correlation matrix** are available on the **Worksheet** tab; in the **Diagnostics** group, click the arrow under **Correlation matrix**.

The linear correlation coefficients  $R$  between all the terms in the model and all the responses are displayed in the **Correlation matrix** and **Correlation plot**.

Process factors are transformed, scaled, and centered as specified in the factor definition for MLR (default = orthogonal scaling). Responses are transformed as specified in the response definition.

Formulation factors are always scaled orthogonally.

The value of the correlation coefficient  $R$  represents the extent of the linear association between two terms. The value of  $R$  ranges from -1 to 1. When  $R$  is near zero there is no linear relationship between the terms.

### Correlation matrix

On the **Worksheet** tab, in the **Diagnostics** group, click **Correlation matrix**.

Correlation coefficients above the threshold, between a term in the model and the responses are colored green and those between terms of the model are colored red.

	1	2	3	4	5	6	7	8	9
1		Air	EGR%	Needlelift	Air*Air	Air*NL	Fuel	NOx	Soot~
2	Air	1	-0.00211416	0.00628096	0.0925578	-0.031827	0.520713	0.100585	-0.26335
3	EGR%	-0.00211416	1	-0.010124	-0.0210034	0	0.181283	-0.950098	0.795577
4	Needlelift	0.00628096	-0.010124	1	0.0156455	0.0159197	-0.790085	0.251371	0.48337
5	Air*Air	0.0925578	-0.0210034	0.0156455	1	0.014029	-0.185609	0.0681562	0.0150444
6	Air*NL	-0.031827	0	0.0159197	0.014029	1	-0.0908503	0.0243728	0.0403177
7	Fuel	0.520713	0.181283	-0.790085	-0.185609	-0.0908503	1	-0.322205	-0.399259
8	NOx	0.100585	-0.950098	0.251371	0.0681562	0.0243728	-0.322205	1	-0.673608
9	Soot~	-0.26335	0.795577	0.48337	0.0150444	0.0403177	-0.399259	-0.673608	1

Threshold 0.3

### Correlation matrix properties

**Correlation Matrix** properties allows control over threshold, colors, and number format. To access the properties, right-click the **Correlation Matrix** and click **Properties**.

*Note: By setting **Threshold** to '0' all correlations above 0 are colored.*

Correlation Matrix - □ ×

Correlation colors Number format

Select colors for cells with absolute correlations above the specified threshold.

Threshold:

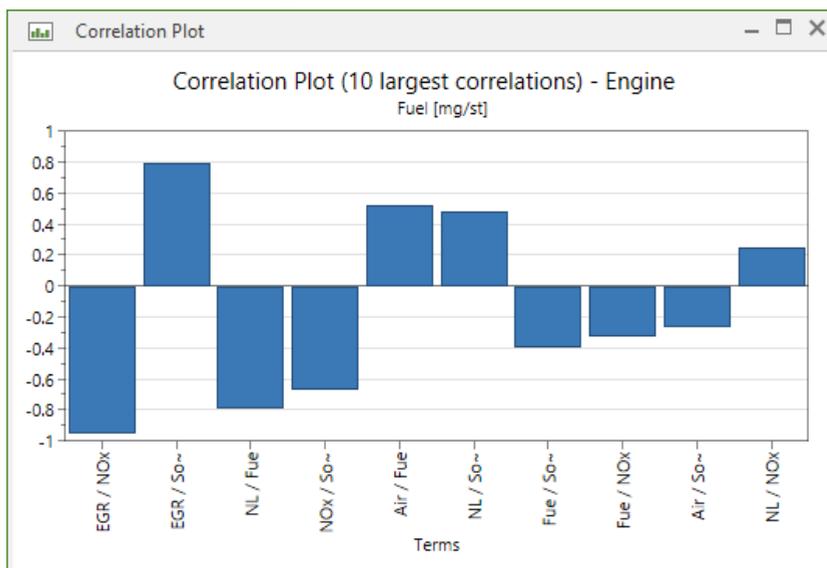
Factor color:

Response color:

Default threshold and colors are set in MODDE options.

### Correlation plot

1. On the **Worksheet** tab, in the **Diagnostics** group, click the arrow under **Correlation matrix**.
2. Click **Correlation plot**.



The default plot displays the 10 largest correlation coefficients.

To change number of correlations to display or limit the number of correlations according to a threshold:

1. Right-click the plot and click **Properties**.
2. Make the change, for instance click **Show absolute correlations above threshold** and enter a value. With '0' all correlations are displayed.

---

*Note: By setting **Threshold** to '0' all correlations above 0 are displayed. By default, the 10 largest correlations in absolute values are displayed.*

---

## Descriptive statistics

The descriptive statistics list summarizes the worksheet and model statistics for all selected responses. To display the list, on the **Worksheet** tab, in the **Diagnostics** group, click **Descriptive statistics**.

Descriptive Statistics		
	1	2
1		Fuel
2	<b>Worksheet statistics</b>	
3	Worksheet runs	17
4	N	17
5	Min	217.069
6	Max	262.626
7	Mean	240.422
8	Q(25%)	231.77
9	Q(75%)	247.479
10	Median	242.841
11	Std. dev.	12.2615
12	Min/Max	0.826533
13	Std. dev./Mean	0.0509999
14	Skewness	-0.180455
15	Skewness test	-0.328251
16	Kurtosis	-0.282755
17		
18	<b>Model statistics</b>	
19	Model type	Evaluation of MLR model
20	Scaling type	All factors are orthogonally scaled
21	DF	11
22	R2	0.982412
23	R2 adj	0.974417

### Worksheet statistics

The Worksheet statistics section holds statistics for the response as displayed in the worksheet:

- Worksheet runs - Number of rows included in the worksheet. Individually excluded values are not subtracted.
- N - Number of values included in the worksheet. Individually excluded values are subtracted.
- Min - Smallest value.
- Max - Largest value.
- Mean - Average.
- Q(25%) - The 25% quartile.
- Q(75%) - The 75% quartile.
- Median - Median of the values.
- Std. dev. - Standard deviation.
- Min/Max - Division of the smallest value with the largest.
- Std. dev./Mean - Division of the standard deviation with the mean (average).
- Skewness - The degree of asymmetry of a distribution around the mean.
- Skewness test -  $sTest = Skewness / (\sqrt{6.0 * dN * (dN - 1.0) / ((dN - 2.0) * (dN + 1.0) * (dN + 3.0))})$ , where dN = number of non missing values in the variable and has to be equal to or larger than 3.
- Kurtosis - Descriptor of the shape of the distribution.

### Model statistics

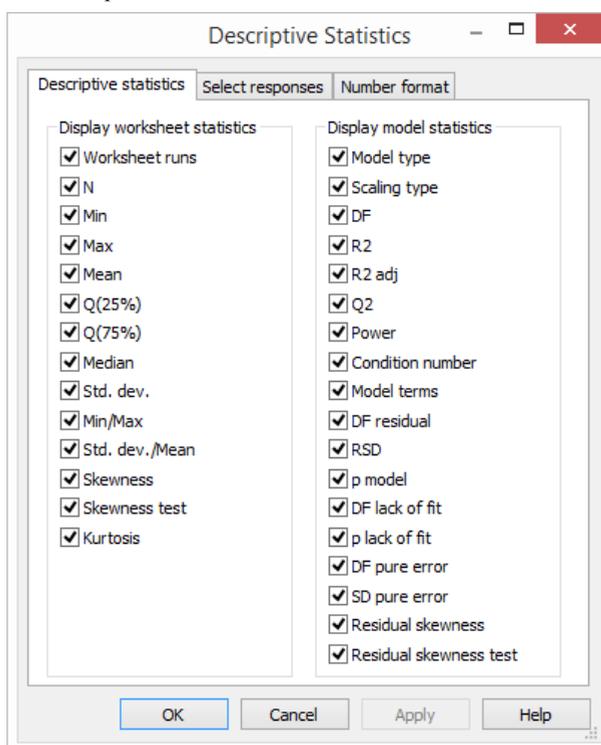
The Model statistics section holds statistics for the modelled response. All other statistics are available in the ANOVA table.

- Model type - Fit method.
- Scaling type - Factor scaling.
- Power (post-hoc) - Estimate of whether the number of runs are sufficient.
- Residual skewness - The degree of asymmetry of a distribution around the mean. Here for the residual vector.
- Residual skewness test -  $sTest = Skewness / (\sqrt{6.0 * dN * (dN - 1.0) / ((dN - 2.0) * (dN + 1.0) * (dN + 3.0))})$ , where  $dN$  = number of non missing values in the variable and has to be equal to or larger than 3. Calculated on the residual vector.

### Descriptive statistics properties

To open the **Descriptive Statistics** dialog box, right-click the **Descriptive Statistics** list and click **Properties**.

The Descriptive Statistics dialog box allows you to select which statistics to show in the Descriptive Statistics list.

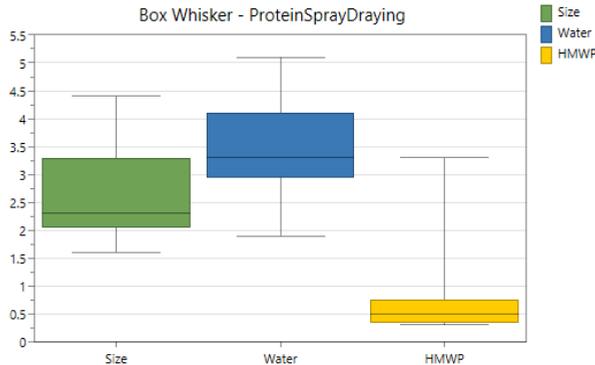


## Box Whisker

The Box Whisker plot illustrates how the response values are distributed around the response mean. The plot uses a box defined by the 25<sup>th</sup> and 75<sup>th</sup> percentiles and whiskers ending at the maximum and minimum values.

Select the desired response or add more responses by:

- Selecting in the **Select responses** box.
- Right-clicking and then clicking **Properties**.



## Histogram

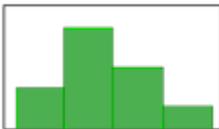
The **Histogram** shows the response distribution and is used to determine if a transformation is needed. If a transformation has already been applied, the response name is followed by tilde ("~").

### Accessing the Histogram

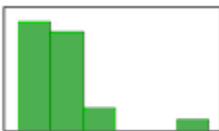
To access the histogram plot, on the **Worksheet** tab, in the **Diagnostics** group, click **Histogram**.

### Transforming the response

If the distribution is not a normal "bell-shaped" distribution, then it is likely a transformation is required.



The desired distribution is a "bell shaped" normal distribution.



If there is a positive skewness, a log transformation is probably preferable.



If there is a negative skewness, a negative log transformation is probably preferable.

---

**Hint:** Biological systems almost always require a logarithmic transform.

---

The ideal distribution is a "bell shaped" normal distribution. If the **Histogram** is not normally distributed, the response may require a transformation.

To transform a response,

1. Right-click the plot
2. Click **Transform**
3. Choose the desired transformation in the **Transform Response** dialog box
4. Click **OK**.

For more, see the Histogram subsection in Chapter 5, Home.

---

## Replicates

### Accessing the Replicate plot

To access the replicates plot, on the **Worksheet** tab, in the **Diagnostics** group, click **Replicates**.

### Plot information

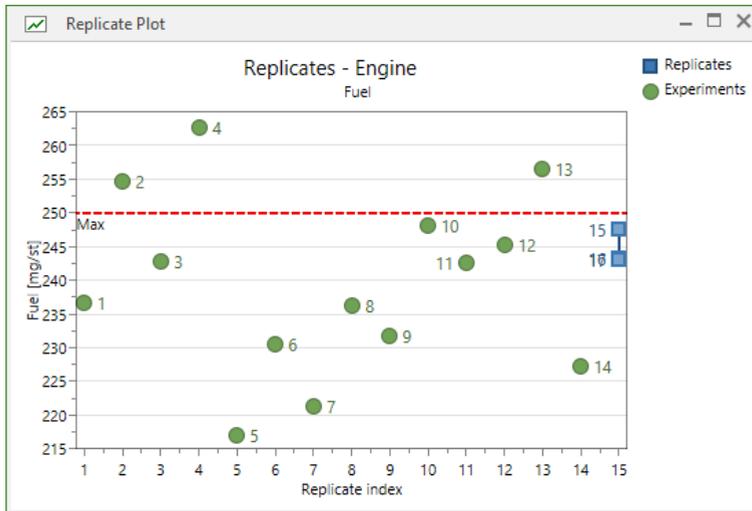
The plot shows the variation in results for all experiments in order to provide a quick overview of raw data. The values of the responses (green and blue points) are plotted vs. experimental runs displaying the variation in the response for replicated experiments.

The ideal outcome is that the variability of the repeated experiments is much less than the overall variability. Experiments deviating significantly from others should be checked. It is not recommended to exclude experiments. Failed experiments can be excluded from the investigation by using **Exclude** once verified with a repeated experiment.

---

***Note:** When the response has been transformed the **Replicate plot** by default displays the back transformed values. To display the plot in the transformed metric, select the **Show transformed values** check box in the **Options** tab in **Properties**.*

---



### Replicated experiments

MODDE checks the rows of all the factors (both included and excluded) in the worksheet for replicates. Rows in the worksheet with the same factor values plus or minus the replicate tolerance are considered replicates.

The default **Replicate tolerance** is 0.1 (10%). You can change the **Replicate tolerance** in **MODDE options** after clicking **File | Options**.

### Understanding the plot

Experiment points above the red *Max* line or under the *Min* line are generally undesirable.

MODDE shows replicated experiments as blue (default) points on the same replicate index.

The variation of repeated experiments should be less than the overall variation for the response.

The Advisor in the Analysis Wizard shows the results after running Tukey's test to identify outliers and the variability test to indicate possible expected results based on the relative replicate variability. For more about these tests, see the Tukey's and variability tests subsection in the Statistical appendix.

### Working with the plot

It is possible to identify outliers when examining the **Replicates** plot.

Some of the common causes and suggested solutions to outliers are:

- Incorrect data entry in the worksheet – In this situation correct the worksheet and refit.
- Experiment incorrectly carried out – If the experiment was not done correctly, repeat the failed experiment and refit.

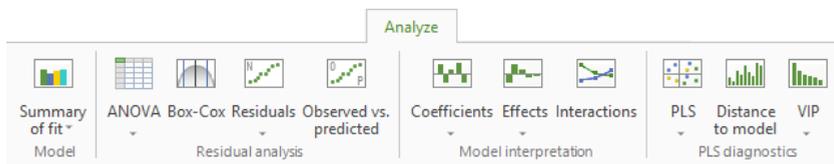
- The deviating result is correct and the experiment does not fit the model – Empirical models are only valid within a limited interval. If the deviating point is correct but uninteresting, then remove it from the analysis and refit the model without it. If the deviating point is in an interesting area, it may be necessary to construct a finer-grained design in that area.

For more, see the Replicates subsection in Chapter 5, Home.

# 08-Analyze

---

## Introduction



The **Analyze** tab presents a number of functions for fitting and analyzing the data gathered in the worksheet.

### Model group

- **Summary of fit** - provides access to a number of summary of fit plots.

### Residual analysis group

- **ANOVA** - opens the **ANOVA table**, **ANOVA plot**, and **Lack of fit**.
- **Box-Cox** - opens the **Box-Cox Plot**.
- **Residuals** - show a residuals plot or list.
- **Observed vs. predicted** - opens the **Observed vs. Predicted Plot**.

### Model interpretation group

- **Coefficients** - provides access to various coefficient lists and plots.
- **Effects** - provides access to various effect plots and the **Effects List**.
- **Interactions** - opens the **Interaction Plot** for the currently selected response.

### PLS diagnostics group, available when the model is fitted with PLS,

- **PLS** - provides access to score and loading plots.
- **Distance to model** - opens the **Distance to Model Plot**.
- **VIP** - provides access to the **VIP Plot** and **VIP List**.

---

## Summary of fit

On the **Analyze** tab, in the **Model** group, clicking **Summary of fit** opens the **Summary of Fit Plot**.

Click the arrow under **Summary of fit** to show a selection of summary plots and lists.

Available plots and lists:

- Summary of Fit Plot
- PLS Total Summary Plot
- PLS Response Summary Plot
- Summary List
- PLS Summary List.

---

*Note: PLS summaries are only available when fitting the model with PLS.*

---

### **Understanding the plot**

Summary statistics are presented in four parameters ( $R^2$ ,  $Q^2$ , Model validity, and Reproducibility) where 1, or 100%, is perfect.  $R^2$  and  $Q^2$ , the first two columns, should be close in size. The difference shouldn't be more than 20% in most situations.

#### **$R^2$**

Shows the model fit. A model with  $R^2$  of 0.5 is a model with rather low significance.

#### **$Q^2$**

Shows an estimate of the future prediction precision.  $Q^2$  should be greater than 0.1 for a significant model and greater than 0.5 for a good model. The difference between  $R^2$  and  $Q^2$  should also be smaller than 0.3 for a good model.  $Q^2$  is the best and most sensitive indicator.

#### **Model validity**

A test of diverse model problems. A value less than 0.25 for *Model validity* indicates statistically significant model problems, such as the presence of outliers, an incorrect model, or a transformation problem. A low value here may also indicate that a term, such as an interaction or square is missing. When the pure error is very small (replicates almost identical), the model validity can be low even though the model is good and complete. When the pure error is so small that the replicates are deemed identical by MODDE, the model validity is labeled *Missing* as it cannot be calculated.

---

*Note: Model validity might be low in very good models ( $Q^2 > 0.9$ ) due to high sensitivity in the test or extremely good replicates.*

---

#### **Reproducibility**

The variation of the replicates compared to overall variability. The *Reproducibility* should be greater than 0.5.

### **Improving summary statistics**

Summary statistics can be improved by tuning the model in the various steps of the **Analysis wizard**.

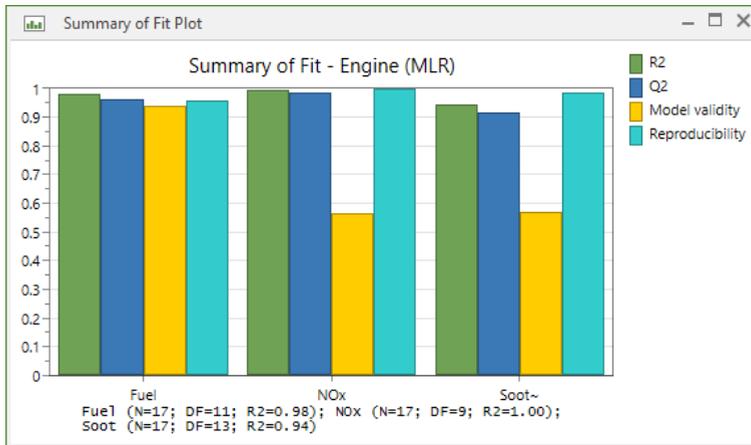
- Select the appropriate transformation in the Histogram.
- Remove non-significant model parameters in the Coefficient plot.

See the Analysis wizard section in Chapter 5, Home for more information.

## Summary of fit plot

To display the **Summary of Fit Plot**, on the **Analyze** tab, in the **Model** group, click **Summary of fit**.

The **Summary of Fit Plot** provides a summary of the basic model statistics, presented visually. For every response MODDE displays 4 columns: R2, Q2, Model validity, and Reproducibility.



You may select to plot R2 Adjusted instead of R2,

1. Right-click the plot
2. Click **Properties**
3. Click the desired type of R2 to display
4. Click **OK**.

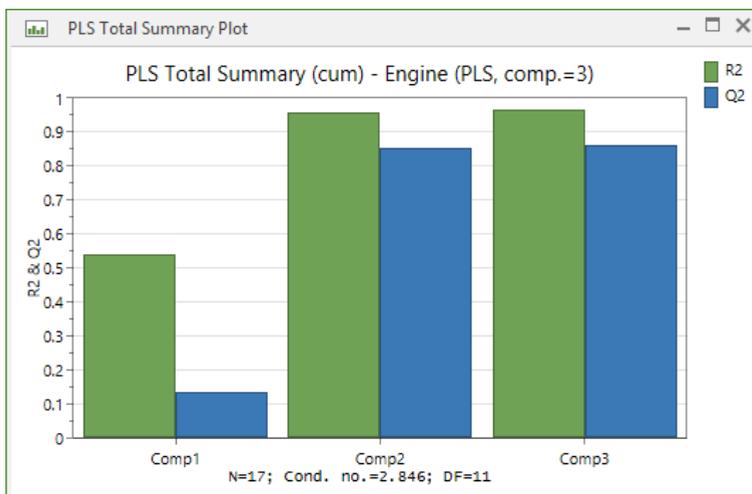
## PLS total summary plot

To display the **PLS Total Summary Plot**,

1. On the **Analyze** tab, click the arrow under **Summary of fit**.
2. Click **PLS total**.

For every fitted response the plot displays R2 and Q2. The definition for R2 and Q2 is the same as in the **Summary of Fit Plot**.

The condition number is calculated for the extended design matrix with the factors scaled to unit variance.



You may select to plot R2 Adjusted instead of R2,

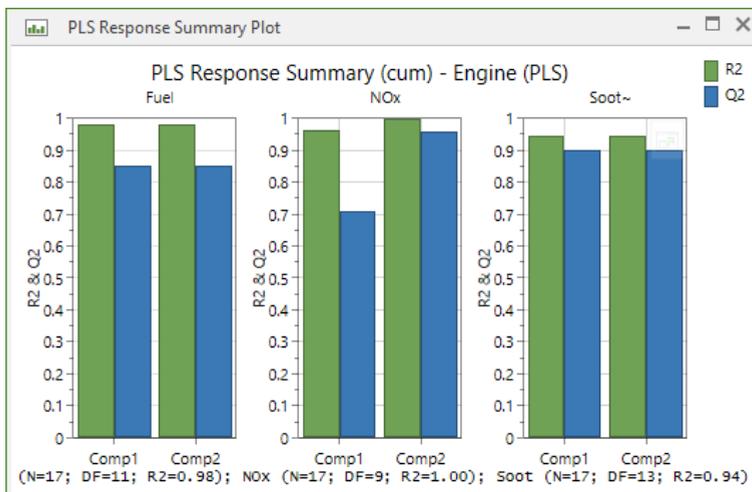
1. Right-click the plot
2. Click **Properties**
3. Click the desired type of R2 to display
4. Click **OK**.

The PLS Total Summary Plot can only be created when all responses are fitted with the same model, that is, when the same terms are included for all responses and there are no missing or excluded values.

## PLS response summary plot

To display the **PLS Response Summary Plot**,

1. On the **Analyze** tab, click the arrow under **Summary of fit**.
2. Click **PLS response**.



This plot displays the R2 and Q2 per PLS component for the selected responses. You can change the selected responses using the **Response** box.

You may select to plot R2 Adjusted instead of R2,

1. Right-click the plot
2. Click **Properties**
3. Click the desired type of R2 to display
4. Click **OK**.

## Summary of fit list

To display the **Summary of Fit List**,

1. On the **Analyze** tab, click the arrow under **Summary of fit**.
2. Under **List**, click **Summary**.

The **Summary of Fit List** shows a number of values for each response in a spreadsheet. The list displays for each response:

- R2
- R2 Adjusted
- Q2
- SDY=Standard Deviation of the Y (response)
- RSD=the Residual Standard Deviation
- N=number of experiments
- Model validity
- Reproducibility.

	1	2	3	4	5	6	7	8	9
1		<b>R2</b>	<b>R2 Adj.</b>	<b>Q2</b>	<b>SDY</b>	<b>RSD</b>	<b>N</b>	<b>Model Validity</b>	<b>Reproducibility</b>
2	<b>Fuel</b>	0.982412	0.974417	0.960539	12.2615	1.9612	17	0.939404	0.959283
3									
4	N = 17	Cond. no. = 2.979							
5	DF = 11								
6									
7	<b>NOx</b>	0.997018	0.994699	0.986781	6.35174	0.462435	17	0.566042	0.998711
8									
9									

## PLS summary list

To display the **PLS Summary List**,

1. On the **Analyze** tab, click the arrow under **Summary of fit**.
2. Click **PLS summary**.

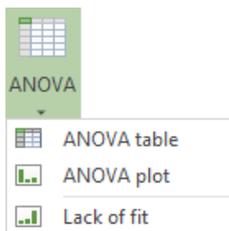
For each PLS component, the list displays R2, R2 adjusted, and Q2. The PLS Summary List also shows the total per component for all responses.

PLS Summary List				
	1	2	3	4
1			<b>Comp 1</b>	<b>Comp 2</b>
2	<b>Fuel</b>	R2	0.980282	0.982409
3		R2adj	0.97132	0.974414
4		Q2	0.849388	0.849388
5				
6	N = 17	Cond. no. =	2.846	
7	Comp. = 2			
8				
9	<b>NOx</b>	R2	0.962511	0.996492
10		R2adj	0.933353	0.993764
11		Q2	0.708571	0.956343
12				
13	N = 17	Cond. no. =	4.193	
14	Comp. = 2			
15				
16	<b>Soot~</b>	R2	0.944519	0.944557
17		R2adj	0.931716	0.931763
18		Q2	0.902275	0.902275
19				
20	N = 17	Cond. no. =	1.037	
21	Comp. = 2			
22				

## ANOVA

The analysis of variance (ANOVA) partitions the total variation of the response (Sum of Squares, SS, corrected for the mean) into a component due to the regression model and a component due to the residuals.

The **ANOVA** menu on the **Analyze** tab provides access to the **ANOVA table**, **ANOVA plot**, and the **Lack of fit** plot.



## ANOVA table

To display the **ANOVA Table**, on the **Analyze** tab, in the **Residual analysis** group, click **ANOVA**.

	1	2	3	4	5	6	7
	Fuel	DF	SS	MS (variance)	F	p	SD
2	Total	17	985053	57944.3			
3	Constant	1	982648	982648			
4							
5	Total corrected	16	2405.51	150.344			12.2615
6	Regression	5	2363.2	472.64	122.882	<b>0.000</b>	21.7403
7	Residual	11	42.3093	3.8463			1.9612
8							
9	Lack of Fit (Model error)	9	30.0661	3.34068	0.54572	<b>0.785</b>	1.82775
10	Pure error	2	12.2432	6.1216			2.47419
11	(Replicate error)						
12							
13							
14		N = 17	Q2 = 0.961		Cond. no. = 2.979		
15		DF = 11	R2 = 0.982		RSD = 1.961		
16			R2 adj. = 0.974				
17							

The selected response 'Fuel'

The analysis of variance (ANOVA) table is displayed for the selected responses.

If there are replicated observations, the residual sum of squares is further partitioned into Pure Error and Lack of Fit. A goodness of fit test is performed by comparing the MS (mean square) lack of fit to the MS (mean square) pure error.

### Replicated observations

MODDE checks the rows of the factors in the worksheet for replicates. Rows in the worksheet with the same factor values plus or minus a 10% tolerance interval are considered replicates and used for the computation of the pure error. The replicate tolerance can be changed in **MODDE options**.

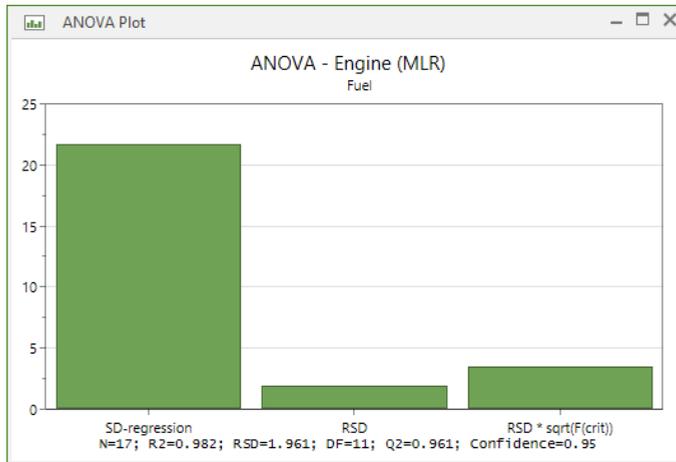
---

*Note: The red coloring of the p-values always refer to the 95% resp. 5% levels.*

---

## ANOVA plot

To display the **ANOVA Plot**, on the **Analyze** tab, in the **Residual analysis** group, click the arrow under **ANOVA**, then click **ANOVA plot**.



In the ANOVA plot the regression component is compared with the residual component and 3 bars are displayed.

- **SD Regression:** Shows the variation of the response explained by the model, adjusted for degrees of freedom and in the same units as Y. This is the square root of MS (mean square) regression.
- **RSD:** Shows the variation of the response *not* explained by the model, adjusted for degrees of freedom and in the same units as Y. This is the residuals standard deviation.
- **RSD\*sqrt(F(crit)):** Shows RSD (second bar) multiplied by the square root of the critical F.

The critical F is the value of the F-distribution over which SD regression is statistically significant at the 95% confidence level.

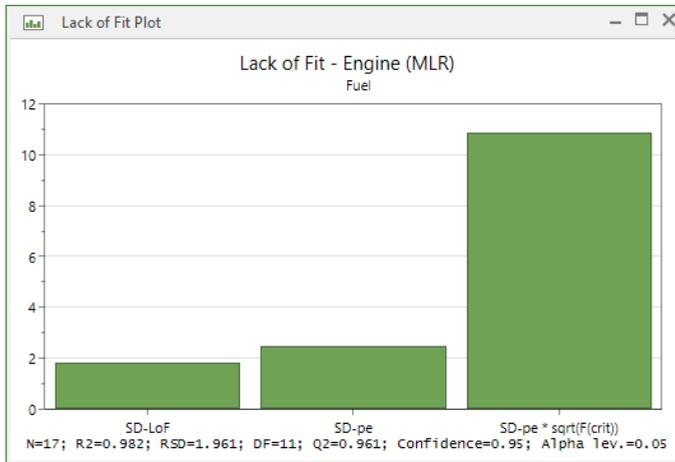
**Hence, when the third bar is smaller than the first, the model is significant at the 5% level.** For more details see the Statistical appendix.

---

## Lack of fit

To display the **Lack of Fit Plot**,

1. On the **Analyze** tab, in the **Residual analysis** group, click the arrow under **ANOVA**.
2. Click **Lack of fit**.



The alpha level for the **Lack of Fit Plot** can be set to 1%, 5%, or 10% by right-clicking the plot and clicking **Properties**.

The lack of fit plot compares the Lack of Fit (LoF) component to the pure error component and displays a graph with 3 bars.

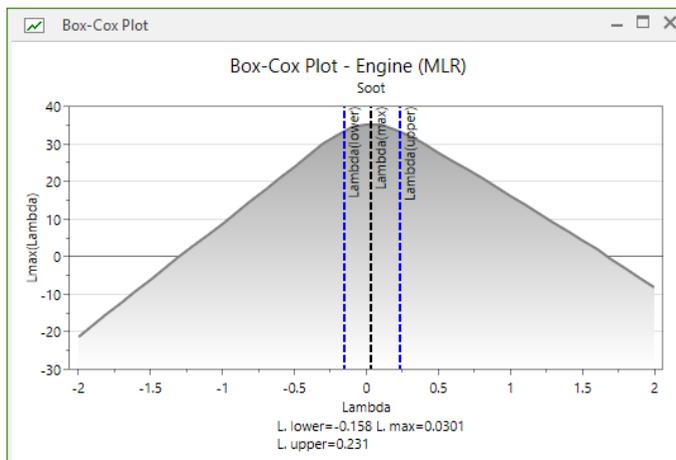
- **SD-LoF**: Shows the variation of the response due to the lack of fit of the model (i.e. the model error) adjusted for degrees of freedom and in the same units as Y. This is the square root of MS (mean square) lack of fit.
- **SD-pe (Pure error)**: Shows the variation **due to** the replicated experiments (observations) adjusted for degrees of freedom and in the same units as Y. This is the square root of MS (mean square) pure error.
- **SD-pe\*sqrt(F(crit))**: Shows SD pure error (second bar) multiplied by the square root of the critical F.

The critical F is the value of the F-distribution over which SD LoF is statistically significant at the 95% confidence level.

**Hence, when the third bar is smaller than the first, the lack of fit is significant at the 5% level.**

## Box-Cox plot (MLR only)

To open the **Box-Cox** plot, on the **Analyze** tab, in the **Residual analysis** group, click **Box-Cox**.



**Note:** The Box-Cox plot is only available when the model is fitted with MLR.

The **Box-Cox Plot** displays the maximum likelihood as a function of the power of the transformation by plotting values of lambda,  $\lambda$ , vs. the maximum likelihood.

If the response values vary more than a magnitude of ten in the experimental domain, a transformation is often recommended.

The maximum point on the Box-Cox plot gives the value of lambda,  $\lambda$ , for the response transformation  $Y^{\lambda}$  that gives the best fit of the model. This is the maximum likelihood estimator for  $\lambda$ .

### Box-Cox transform

MODDE displays  $\lambda_{max}$  and its 95% confidence interval as  $\lambda_{lower}$  and  $\lambda_{upper}$  in the footer and on the plot as 3 vertical lines.

If  $\lambda=1$  is included in that interval, then a transformation is not recommended.

If  $\lambda=1$  is not included in the interval then  $Y^{\lambda_{max}}$  is the recommended transformation. You do not have to use the precise value of  $\lambda_{max}$  but a near convenient value.

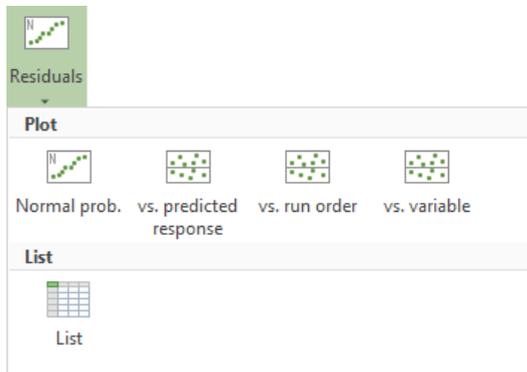
Common transformations are:

- $\lambda_{max} = -1$ , use the Power transformation and C3 = -1.
- $\lambda_{max} = 0$ , use the Logarithmic (10Log) transformation.
- $\lambda_{max} = 0.5$ , use the Power transformation and C3 = 0.5.

## Residuals

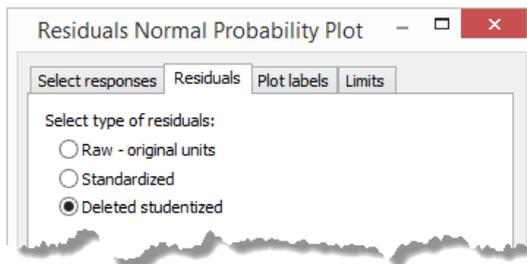
On the **Analyze** tab, in the **Residual analysis** group, click the arrow under **Residuals**. This provides access to the following forms of displaying residuals,

- Normal Probability Plot
- vs. Predicted Response Plot
- vs. Run Order Plot
- vs. Variable Plot
- List.



### Residual type

Residuals can be displayed as **Raw**, **Standardized**, or **Deleted studentized**. The residual type can be changed by right-clicking the plot and clicking **Properties**.



#### **Raw residuals**

The raw residual is the difference between the observed and the fitted (predicted) value.

#### **Standardized residuals**

The standardized residual is the raw residual divided by the residual standard deviation (RSD).

### Deleted studentized residuals

The deleted studentized residual is the raw residual ( $e_i$ ) divided by its standard deviation ( $s_i$ ) where the standard deviation ( $s_i$ ) is computed with observation ( $i$ ) left out of the analysis, and corrected for leverage. Deleted studentized residuals require at least three degrees of freedom.

---

*Note: Deleted studentized residuals are not available for PLS.*

---

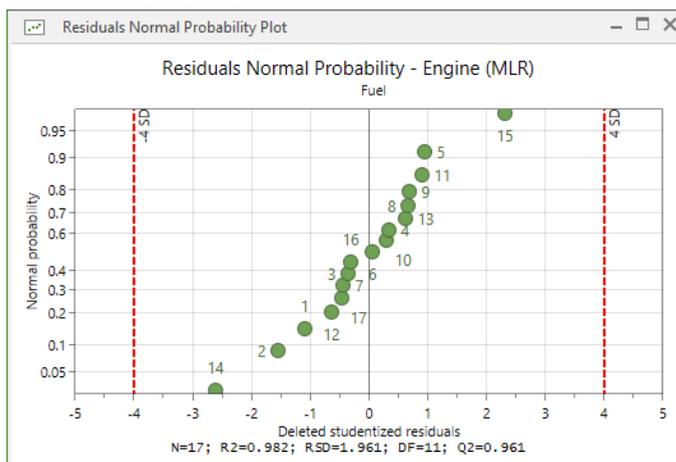
### Default

With MLR and three or more real degrees of freedom (replicate degrees of freedom are not counted), deleted studentized residuals are the MODDE default when plotting residuals.

With PLS, MODDE uses the standardized residuals as default instead.

### Residuals normal probability plot

To display the **Residuals Normal Probability Plot**, on the **Analyze** tab, in the **Residual analysis** group, click **Residuals**.



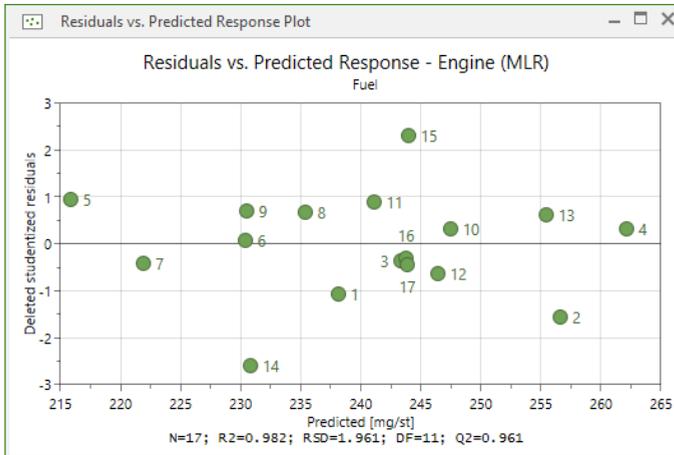
The residuals are plotted on a cumulative normal probability scale. This plot makes it easy to detect:

- Normality of the residuals. If the residuals are normally distributed, the points on the probability plot follow close to a straight line.
- Outliers. These are points deviating from the normal probability line, and having large absolute values of studentized residuals i.e. larger than 4 standard deviation indicated by red lines on the plot.

## Residuals vs. predicted response

To display the **Residuals vs. Predicted Response Plot**,

1. On the **Analyze** tab, in the **Residual analysis** group, click the arrow under **Residuals**.
2. Click **vs. predicted response**.

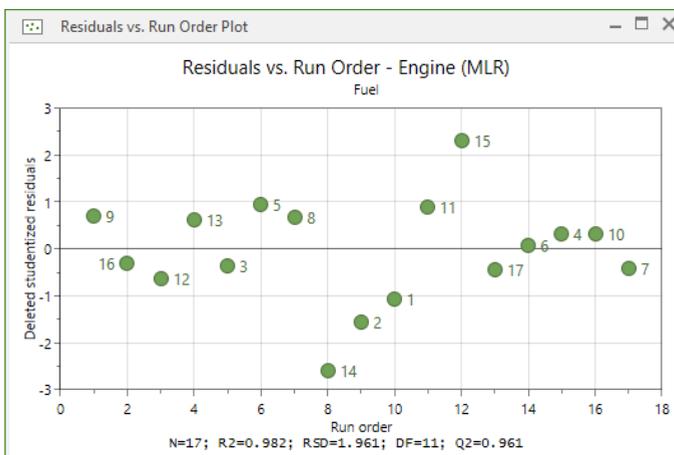


The plot shows the residuals vs. the fitted values. This plot is particularly useful to detect non-constant variance of the residuals. If the spread of the residuals increases with the fitted values, you may need to transform your response by taking its logarithm or its square root.

## Residuals vs. run order

To display the **Residuals vs. Run Order Plot**,

1. On the **Analyze** tab, in the **Residual analysis** group, click the arrow under **Residuals**.
2. Click **vs. run order**.

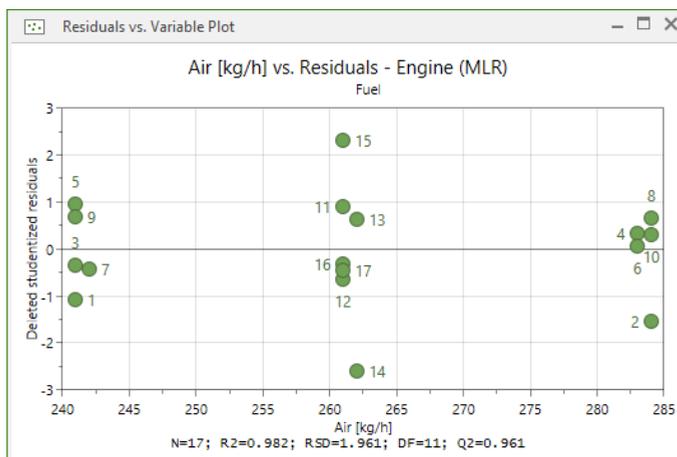


This plot shows the residuals vs. run order (the order in which you performed the experiments) and helps you detect any dependency of the residuals on time.

### Residuals vs. variable

To display the **Residuals vs. Variable Plot**,

1. On the **Analyze** tab, click the arrow under **Residuals**.
2. Click **vs. variable**.



Right-click the plot and click **Properties** to select which variable to display on the x-axis.

### Residuals list

To display the **Residuals List**,

1. On the **Analyze** tab, in the **Residual analysis** group, click the arrow under **Residuals**.
2. Click **List**.

	1	2	3	4	5
	<b>Fuel</b>	<b>Observed</b>	<b>Predicted</b>	<b>Obs - Pred</b>	<b>Conf. int(±)</b>
1	1	236.637	238.106	-1.46893	3.09744
2	2	254.722	256.664	-1.94249	3.17229
3	3	242.841	243.351	-0.510117	3.09228
4	4	262.626	262.148	0.478394	3.05352
5	5	217.069	215.807	1.26245	3.17091
6	6	230.527	230.438	0.088974	3.07892
7	7	221.274	221.884	-0.610245	3.109

The number under the response name is the experiment number.

**Observed:** The value of the response as listed in the worksheet.

**Predicted:** The predicted value for that observation.

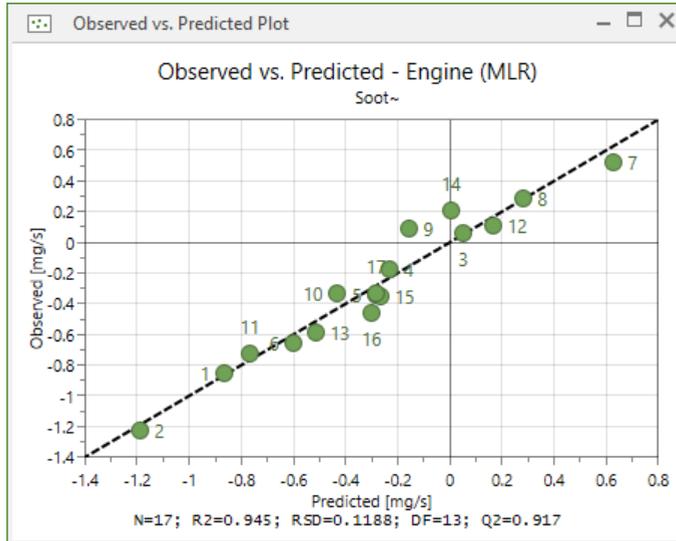
**Obs - Pred:** The residual for that observation.

**Conf. int:** The 95% confidence interval on the predicted value. Right-click the list and click **Properties** to change the confidence level.

## Observed vs. predicted

To open the plot, on the **Analyze** tab, in the **Residual analysis** group, click **Observed vs. predicted**.

The **Observed vs. Predicted Plot** displays observed values vs. predicted values.

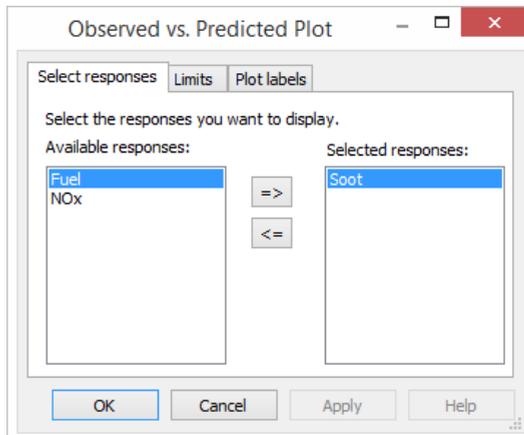


### Understanding the plot

Plots with points close to a straight line indicate good models. If Degrees of Freedom is under 3 ( $DF < 3$ ), the plot will implicitly give a perfect fit.

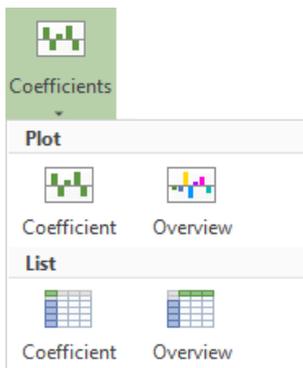
### Properties

The properties dialog box of the **Observed vs. Predicted Plot** allows for the responses to be selected, limits to be shown or not, and what plot labels to display.



## Coefficients

There are two coefficient plots and two coefficient lists available in MODDE. To see the coefficient options, on the **Analyze** tab, in the **Model interpretation** group, click the arrow under **Coefficients**.



The coefficient plots provide graphical presentation of the significance of the model terms.

A significant term is one with a large distance from  $y=0$  as well as having an uncertainty level that does not extend across  $y=0$ . A non significant model term is a model term close to  $y=0$  and with an uncertainty level that crosses  $y=0$ . The trigger for the **Auto tune** feature in the Coefficient plot is that Q2 increases if the smallest non significant terms are excluded.

When confoundings are present, the coefficient plots and lists display a bracket # after the term. Hover your mouse pointer over the column to view the confounded terms.

### *Understanding the plot*

The **Coefficient plot** presents a graphical representation of the model terms in order to determine their significance.

#### **Significant model terms:**



- Far away from  $y=0$  (either positive or negative)
- Uncertainty range does not cross  $y=0$ .

#### **Non significant model terms:**



- Close to  $y=0$
- Uncertainty range crosses  $y=0$ .

### Remove non significant model terms from the plot

1. Click **Exclude**.
2. Find the least significant model term.
3. Check that no significant model term depends on the least significant model term.
4. Click the non significant model term to remove it.
5. Repeat steps 2-4 for as many model terms as required.
6. Click **Exclude** again.

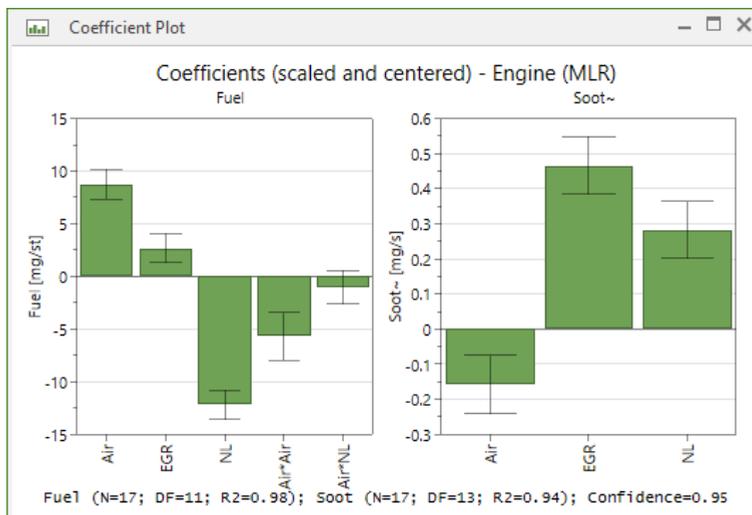
---

**Hint:** Exclude only one model term at a time as it may make other model terms more significant. Q2 and DF (Degrees of Freedom) will increase as non significant model terms are removed, so note the before and after values as each term is excluded. A higher DF and Q2 is better.

---

### Coefficient plot

The **Coefficient Plot** displays the coefficients, when changing from 0 to high, for the selected response with the confidence interval as error bars. By default, the coefficients refer to the data scaled and centered.



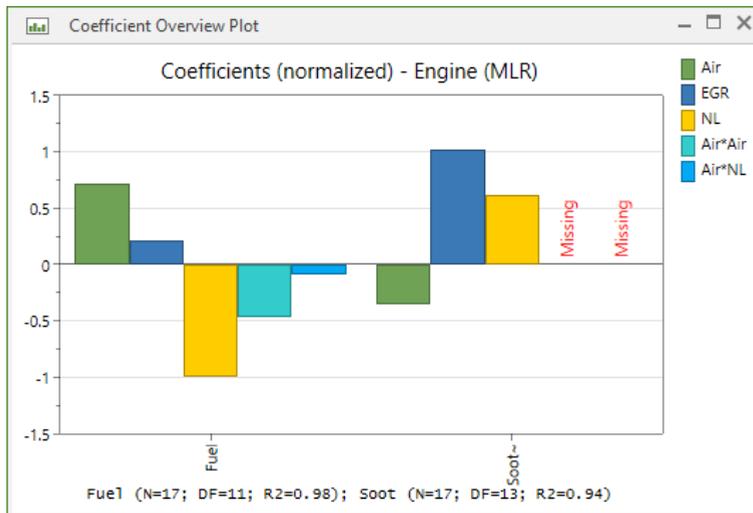
You can select which type of coefficient to display from **Properties**. Click **Scaled and centered** (default), **Normalized**, **Unscaled**, or **PLS orthogonal** (only available for PLS).

When you have confounded terms in your investigation these terms are marked with a bracket (#). Hover the mouse pointer over the column in the plot and you will get information about which terms it is confounded with. This information is available also in the coefficient overview plot.

### Coefficient overview plot

To make the coefficients comparable when responses (Y's) have different ranges, the coefficient overview plot displays the coefficients in **Normalized** form (the coefficients are divided by the standard deviation of their respective response). The **Normalized** mode is the default for the overview plot. You can change it to regular scale and centered coefficients by right-clicking and then clicking **Properties**.

The color for each factor is the same over all responses in the coefficient overview plot.



### Coefficient list

The coefficient list displays the scaled and centered coefficients for the selected responses. P-values signaling non significant coefficients at the selected confidence level are colored red.

	1	2	3	4	5
1	Soot~	Coeff. SC	Std. Err.	P	Conf. int(±)
2	Constant	-0.279945	0.0288614	2.56007e-07	0.0623513
3	Air	-0.157893	0.0389473	0.00136603	0.0841406
4	EGR%	0.46588	0.0380323	1.6319e-08	0.0821639
5	Needlelift	0.283157	0.0375012	4.18648e-06	0.0810165
6					
7					
8	N = 17	Q2 = 0.917		Cond. no. = 1.355	
9	DF = 13	R2 = 0.945		RSD = 0.1188	
10		R2 adj. = 0.932			
11				Confidence = 0.95	

Coefficients (scaled and centered).

For the selected responses, this list includes:

**Terms:** Under the response name the name of the terms included in the model are listed. To display the list for another response, or more than one response, use the **Response** box or the list's **Properties**.

**Coefficients:** Value of the coefficient.

**Standard error:** Standard error of the coefficient.

**P:** Probability to get the displayed value for the coefficient assuming that the coefficient has a true value of zero.

**Confidence interval:** The 95% confidence interval on the coefficient value. To select a different level for the confidence interval, right-click the list and then click **Properties**.

## Coefficient overview list

The coefficient overview list displays the normalized coefficients for all the responses. Non significant coefficients at the selected confidence level are shown in red.

	1	2	3	4
1		<b>Coeff. N: Fuel</b>	<b>Coeff. N: NOx</b>	<b>Coeff. N: Soot</b>
2	<b>Constant</b>	19.8207	3.16581	-0.67221
3	<b>Air</b>	0.71402	0.123965	-0.35213
4	<b>EGR%</b>	0.217511	-1.21232	1.02574
5	<b>Needlelift</b>	-0.994247	0.303375	0.621264
6	<b>Air*Air</b>	-0.464109	0.0749983	0.103057
7	<b>Air*NL</b>	-0.0851657	0.0339591	0.0341566
8				
9				
10	N = 17			
11	Cond. no. = 2.979			
12	DF = 11			
13				
14	Confidence = 0.95			

Coefficients (normalized).

*Note: The Coefficient overview list is only available when all responses are fitted to the same model, that is, when the same model terms are included for all responses.*

## Effects

In MODDE you can create three different effect plots:

- **Effect Plot:** Displays the effect calculated as twice the MLR coefficient and sorted descending in absolute value.
- **Effects Normal Probability Plot:** Displays the effects on a normal probability scale.
- **Main Effect Plot:** Displays predicted values of the selected responses, when the factor varies from its low to its high level.

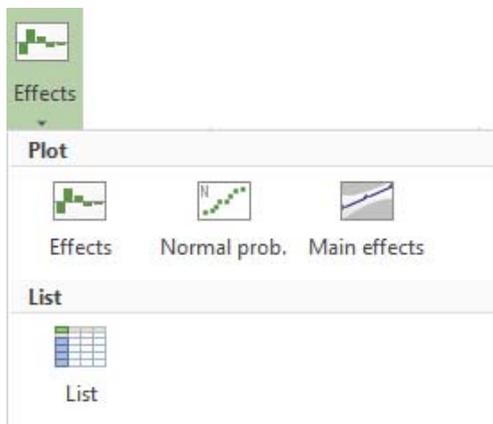
The **Effect List** displays the effects and their confidence intervals.

*Note: When an expanded list of the qualitative levels is desirable, use the coefficient plots and lists to display the coefficients for every level of a qualitative variable instead of the effect plots and lists. Note that the effects for linear and interaction models are twice the corresponding coefficients.*

## Accessing effects plots and list

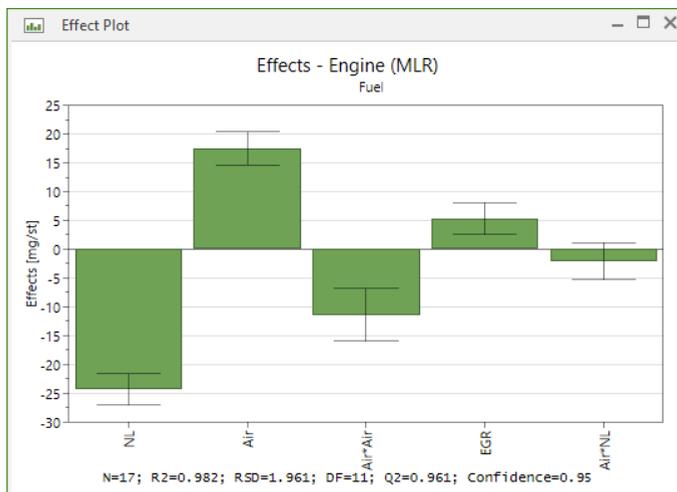
To open an effects list or plot,

1. On the **Analyze** tab, in the **Model interpretation** group, click the arrow under **Effects**.
2. Click the desired plot or list.



## Effect plot

For process factors the values of the effects (computed as twice the MLR coefficients) are plotted sorted (in absolute value) in descending order. The  $\pm$  95% confidence interval is shown as error bars. The confidence level can be changed in the **Effect Plot** dialog box (right-click the plot and click **Properties**).



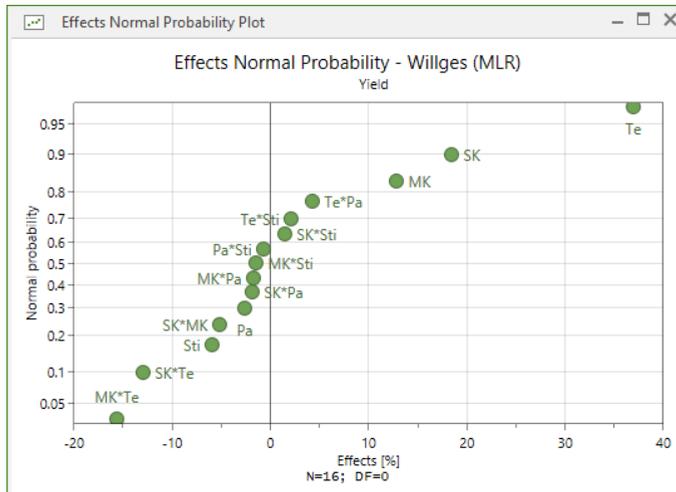
For mixture factors this plot displays the adjusted Cox effects (unavailable for Scheffé models). This effect represents the change in the response values when component  $k$  varies over its range, all other mixture factors kept in the same proportion as in the reference mixture.

For details on how the mixture effects are calculated, see the Statistical appendix.

## Effects normal probability plot

The plot displays the effects plotted on a cumulative normal probability scale.

This plot should only be used when the model is saturated (a model with as many terms as experimental runs). For saturated models MODDE cannot compute standard errors, p values and confidence intervals. Use the plot to help determine the important effects.



This plot, proposed by Daniel in 1954, is based on the fact that if all estimated effects were noise, they would have a normal distribution and when plotted on a normal cumulative plot, would fall on a straight line. Hence effects significantly different from zero (noise) will fall outside the normal line.

This plot *assumes independent effects*, and that all estimable effects are plotted. Hence, it is only relevant for screening designs with saturated models  $DF = 0$ . Also for this plot to be meaningful, models need to have at least 10 effects. If these conditions are not met MODDE will warn that this plot may not be statistically correct.

---

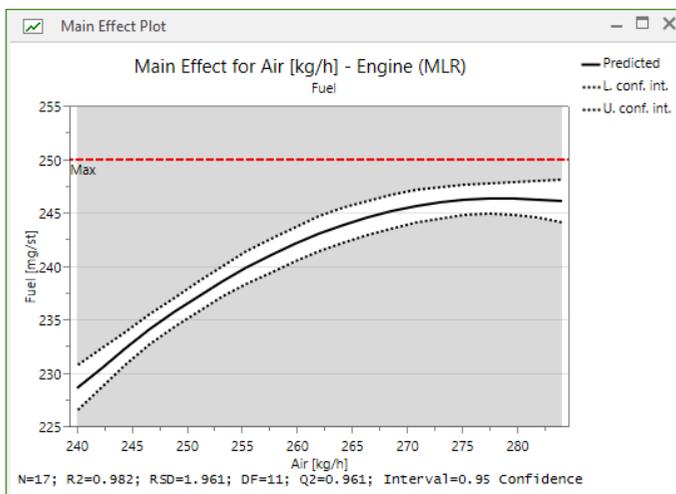
**Note:** *The normal probability of effects plot is not available with mixture factors.*

---

## Main effect plot

For process factors the plot displays the predicted values of the selected response, when the factor varies over its range, all other factors in the design held constant at their averages. Qualitative factors held constant are set at their numerical average and do not influence the predicted values.

If the design has replicated center points for the displayed factor, where the other factors are also at their averages, the observed values of these points are colored differently from other worksheet points.



You can select to display the worksheet data in this plot, current factor vs. response, not taking into account the other factor settings. For Stability testing designs the worksheet data are by default displayed.

To display/hide the worksheet data, right-click the plot, click **Properties**, and on the **Options** tab select/clear the **Worksheet data** check box.

### Main effect plot for mixture factors

When you select to display a mixture factor  $X_i$ , this plot displays the predicted change in the response when  $X_i$  varies from its **low** to its **high** level, the relative amounts of all other mixture factors are kept in the same proportion as in the standard reference mixture.

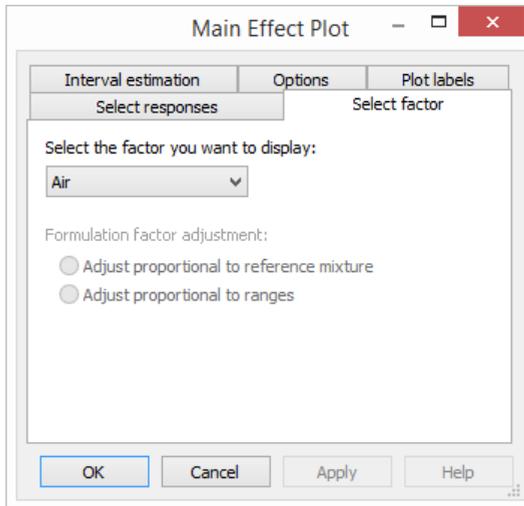
For details, see the Statistical appendix.

### Switching displayed factors

To switch factors,

1. Right-click the plot,
2. Click **Properties**,
3. Click the **Select factor** tab,
4. Select the desired factor,
5. Click **OK**.

Formulation factors can be adjusted proportional to the reference mixture or ranges on this property page.



## Effect list

The values of the effects (twice the coefficients) are listed with their 95% confidence interval sorted (in absolute value) in descending order. The confidence level can be changed in the **Properties** dialog box for the **Effect List** (right-click the plot and click **Properties**).

	1	2	3	4	5
1	<b>Fuel</b>	<b>Effect</b>	<b>Conf. int(±)</b>		
2	Needlelift	-24.3819	2.72595		
3	Air	17.5099	2.84415		
4	Air*Air	-11.3813	4.5893		
5	EGR%	5.33402	2.7645		
6	Air*NL	-2.08852	3.1763		
7					
8	N = 17	Q2 =	0.961	Cond. no. =	2.979
9	DF = 11	R2 =	0.982	RSD =	1.961
10		R2 adj. =	0.974		
11				Confidence =	0.95

## Interactions

To open the interaction plot, on the **Analyze** tab, in the **Model interpretation** group, click **Interactions**.

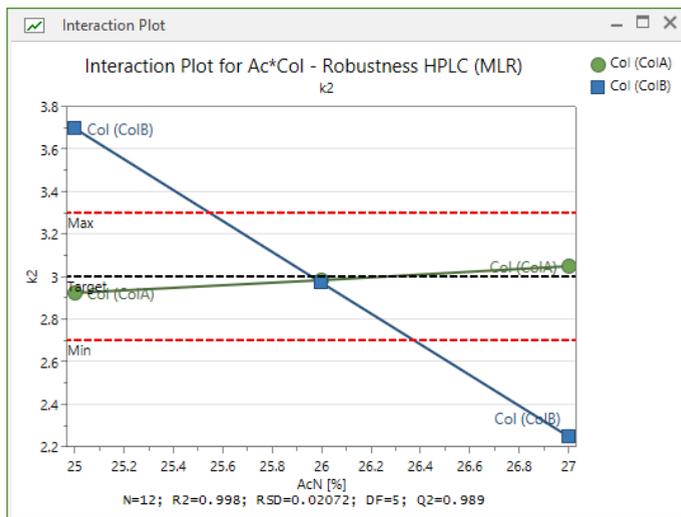
To switch to another interaction and/or switch the factor displayed on the x-axis:

1. Right-click the plot and click **Properties**.
2. In the **Interaction effects** tab, click the term in the **Interaction term** box and/or select the factor in the **Factor on X-axis** box.

*Note: This plot is only available for process factor. No interaction plots are available for mixture factors.*

When you select a 2 factor interaction, the predicted values of the response, when one factor varies from its low to its high level, are plotted for both levels of the other factor, all remaining factors in the design being set on their average.

If you have mixture factors in the model, these are all set at the standard reference mixture.



## PLS

The score and loading plots complement each other. The position of an observation in a given direction in a score plot is influenced by variables lying in the same direction in the loading plot.

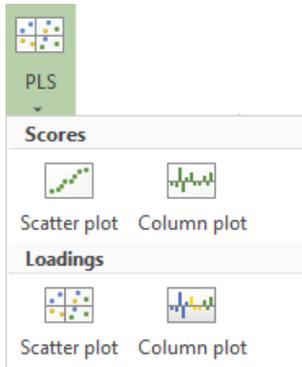
There are four types of PLS plots available:

- **Score Scatter Plot** – displays the score vectors  $t$  and/or  $u$  in a scatter plot. Use the score plots to reveal groups, trends, outliers, and similarities. The plot marks represent your experiments.
- **Score Column Plot** – displays the score vectors  $t$  or  $u$  in a column plot.
- **Loading Scatter Plot** – displays the loading vectors  $p$ ,  $c$ ,  $w$  and/or  $wc$  in a scatter plot. Use the loading plots to investigate the correlation between terms in your model. The plot marks represent the terms currently in the model and/or the response-variables. The response variables are displayed when plotting the  $c$ -vectors.
- **Loading Column Plot** – displays the loading vectors  $p$ ,  $c$ ,  $w$  or  $wc$  in a column plot.

### Accessing PLS plots

To access the PLS plots, on the **Analyze** tab, in the **PLS diagnostics** group, click the arrow under **PLS**.

This shows the available PLS plots.



## Score plots

The score plots reveal groups, trends, outliers, and similarities.

You can create three different score scatter plots:

- $T$  scores (for ex.:  $t_1$  vs.  $t_2$ ):  $t$  scores are windows in the  $X$  space displaying the objects as situated on the projection plane or hyper plane.
- $U$  scores (for ex.:  $u_1$  vs.  $u_2$ ):  $u$  scores are windows in the  $Y$  space, displaying the objects as situated on the projection plane or hyper plane.
- $T$  vs.  $U$  scores (for ex.:  $t_1$  vs.  $u_1$ ): Displays the objects in the projected  $X$  ( $T$ ) and  $Y$  ( $U$ ) space, and how well the  $Y$  space coordinate ( $u$ ) correlates to the  $X$  space coordinate ( $t$ ). This is the default.

Score column plots can only be created for one vector at a time.

---

*Note: Score plots are only available for PLS fitted models.*

---

## Loading plots

The loading plots display the correlation between the  $X$  variables  $T(X)$  and the  $Y$  variables  $U(Y)$ .

---

*Note: Loading plots are only available for PLS fitted models.*

---

You can create four different loading plots:

- $p$  loadings (for ex.:  $p_1$  vs.  $p_2$ ): These plots show the importance of the  $X$  variables in the approximation of the  $X$  matrix.
- $w$  loadings (for ex.:  $w_1$  vs.  $w_2$ ): The  $w$ 's are the weights that combine the  $X$  variables (first dimension) or the residuals of the  $X$  variables (subsequent dimensions) to form the scores  $t$ . These weights are selected so as to maximize the correlation between  $T$  and  $U$ , thereby indirectly  $Y$ .  $X$  variables with large  $w$ 's (positive or negative) are highly correlated with  $U$  ( $Y$ ). Variables with large  $w$ 's are situated far away from the origin (on the positive or negative side) on the plot.
- $c$  loadings (for ex.:  $c_1$  vs.  $c_2$ ): These plots display the correlation between the  $Y$  variables and the  $X$  scores  $T(X)$ . The  $c$ 's are the weights that combine the  $Y$  variables with the scores  $u$ ; so as to maximize their correlation with  $X$ .  $Y$  variables with **large  $c$ 's** are highly correlated with  $T(X)$ .

- wc loadings (for ex.: wc1 vs. wc2): This is the default option. These plots show **both** the X-weights (w) and Y-weights (c), and thereby the correlation structure between X and Y. One sees how the X and Y variables combine in the projections, and how the X variables relate to Y.

To change the currently showing loading plot,

1. Right-click the loading plot
2. Click **Properties**
3. On the **Loading Column Plot** or **Loading Scatter Plot** tab, click the desired option
4. Click **OK**.

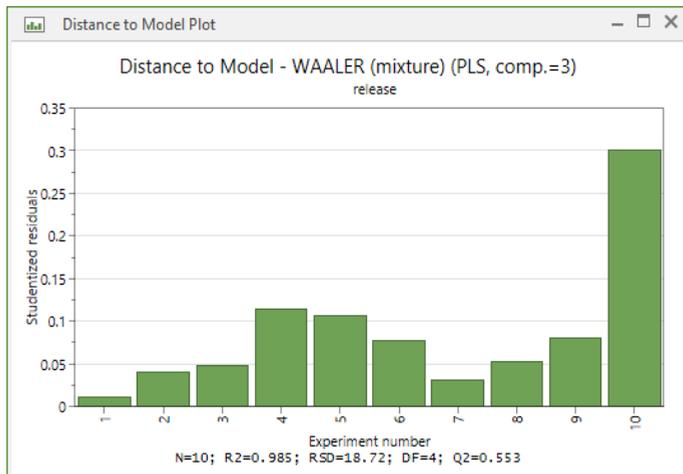
Loading column plots can only be created for one vector at a time.

## Distance to model (Y)

To open the distance to model plot, on the **Analyze** tab, in the **PLS diagnostics** group, click **Distance to model**.

The RSD of an object in the Y space is proportional to the object distance to the hyper plane of the PLS model in the Y space. MODDE computes the object distance to the PLS model (DModY) in the Y space and displays them as columns in a plot.

A large DModY value indicates that the experiment may be an outlier.



## VIP - Variable importance in the projection

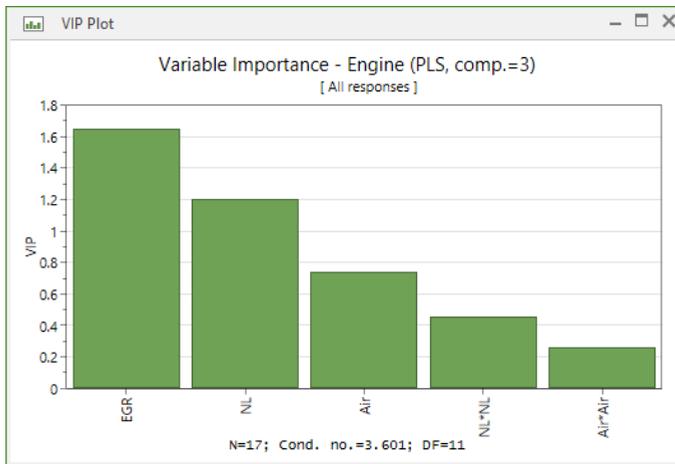
The VIP values reflect the importance of terms in the model both with respect to Y, i.e. its correlation to all the responses, and with respect to X (the projection). With designed data, i.e. close to orthogonal X the VIP values mainly reflect the correlation of the terms to all the responses.

*Note: The VIP plot and VIP list are only available for PLS fitted models.*

### VIP plot

The VIP plot displays the VIP values as a column plot sorted in descending order.

To open the **VIP Plot**, on the **Analyze** tab, in the **PLS diagnostics** group, click **VIP**.



### VIP list

The VIP list displays the sorted VIP values and the scaled and centered coefficients for all responses in the investigation.

To open the **VIP List**,

1. On the **Analyze** tab, in the **PLS diagnostics** group, click the arrow under **VIP**.
2. Click **VIP list**.

	1	2	3	4	5
1		<b>VIP</b>	<b>Coeff</b>	<b>Coeff</b>	<b>Coeff</b>
2		(cum)	<b>Fuel</b>	<b>NOx</b>	<b>Soot~</b>
3	<b>EGR%</b>	1.65193	2.07838	-6.05782	0.359391
4	<b>Needlelift</b>	1.20328	-9.78619	1.06713	0.191255
5	<b>Air</b>	0.739169	6.13271	0.0708223	-0.163738
6	<b>NL*NL</b>	0.454716	-1.81407	0.659173	0.0225876
7	<b>Air*Air</b>	0.264822	-1.3318	0.479865	0.0123445
8					
9	N = 17	Cond. no. = 3.601			
10	Comp. = 3				
11					

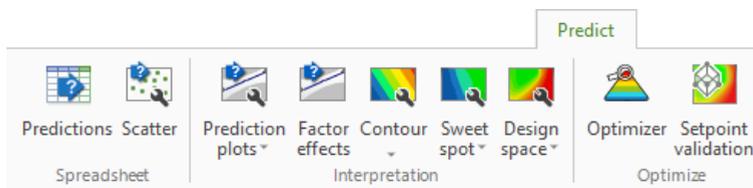


# 09-Predict

---

## Introduction

The **Predict** tab has a collection of functions that involve predictions and interpretation, as well as the Optimizer.



The functions on the **Predict** tab are organized into groups,

### Spreadsheet group

- **Predictions** - opens the **Prediction Spreadsheet**.
- **Scatter** - opens the **Prediction Scatter Plot** dialog box for creating 2D or 3D scatter plots using the content of the Prediction Spreadsheet.

### Interpretation group

- **Prediction plots** - creates prediction and overlay prediction plots.
- **Factor effects** - shows the **Factor Effect Plot** for the currently selected responses.
- **Contour** - opens the **Contour Plot** dialog box to create **Contour**, **4D contour**, and **Surface** plots.
- **Sweet spot** - creates **Sweet spot**, **4D sweet spot**, and **Surface** plots.
- **Design space** - opens the **Design Space Plot** dialog box for creating **Contour** and **4D contour** plots.

### Optimize group

- **Optimizer** - opens the optimizer window and provides access to the **Optimizer** contextual tab.
- **Setpoint validation** - opens the **Setpoint Validation** window.

---

## Prediction spreadsheet

The **Prediction Spreadsheet** allows you to type or paste factor settings to make predictions. The lower and upper level 95% confidence intervals for these predictions are also calculated.

To open the **Prediction Spreadsheet**, on the **Predict** tab, click **Predictions**.

To insert rows press the **Down** arrow key when on the last row.

If the model has been fitted the predictions are calculated automatically after entering settings for all factors. If you do not want the predictions to be automatically updated, right-click the list and clear **Auto update**. To make predictions with auto update cleared, right-click the list and click **Update predictions** or press the F5 key.

To change the interval estimation properties, right-click the list and click **Properties** and click the **Interval estimation** tab.

---

*Note: With PLS and an X matrix with large condition number, the standard error of predictions is computed and displayed instead of the interval estimate.*

---

	1	2	3	4	5	6	7	8
	Air	EGR%	Needlelift	Fuel	Lower	Upper	NOx	Lo
1	250.001	7.3628	-2.2009	232.338	230.671	234.004	24.1012	23.6
2	250.002	7.4144	-2.3588	233.019	231.386	234.652	23.7878	23.3
3	250.001	6.9531	-1.3295	228.469	226.522	230.416	26.3806	25.8
4	253	7.4782	-1.3624	230.927	229.082	232.772	24.8086	24.3
5	250.017	7.3245	-2.0433	231.68	229.98	233.381	24.3807	23.9
6	250	6.9652	-1.347	228.55	226.61	230.489	26.3191	25.8
7	250	6.9487	-1.326	228.451	226.502	230.399	26.3995	25.
8	250	6.9487	-1.326	228.451	226.502	230.399	26.3995	25.

Interval=0.95 Confidence

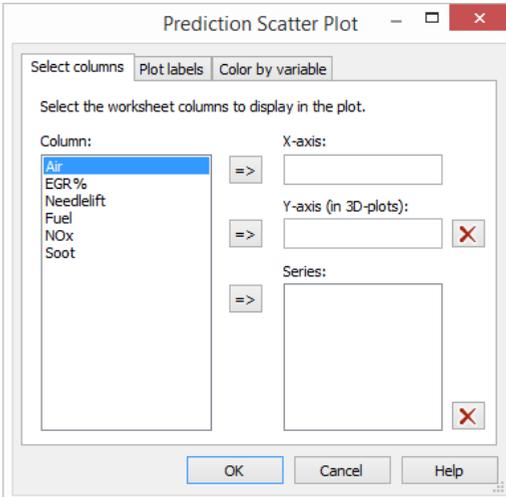
## Prediction scatter plot

In the prediction scatter plot you can view your factors and predicted responses as 2D and 3D scatter plots.

*Note: The points displayed in the scatter plots are the points in the current prediction spreadsheet. The scatter plot is updated with new points when you enter them in the prediction spreadsheet.*

---

To create a prediction scatter plot, on the **Predict** tab, click **Scatter Plot**. This will open up the **Prediction Scatter Plot** dialog box.



To add a variable to an axis or series, click the variable in the **Column** field, then click the relevant arrow . Click  to remove factors or responses from the **Y-axis** or from **Series**. To change what to display on the **X-axis**, mark a new variable and click .

For **2D** plots, select the variable to be plotted on the **X** and **Series (Z)** axes.

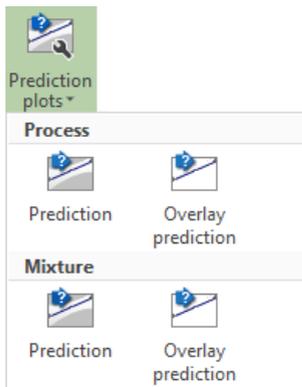
For **3D** plots, select the variables to be plotted on the **X**, **Y**, and **Series (Z)** axes.

Click the **Color by variable** tab to color by a factor or a response.

## Prediction plots

There are two prediction plots available for process and two prediction plots available for mixture.

On the **Predict** tab, in the **Interpretation** group, click the arrow under **Prediction plots**.



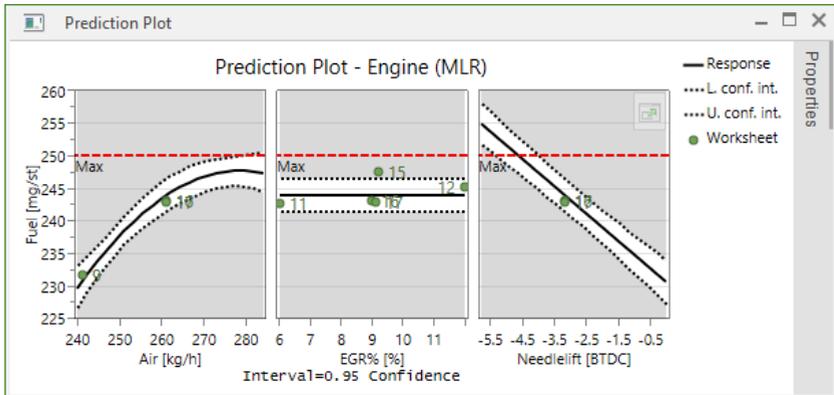
Right-click any prediction plot and click **Properties** to:

- Select the responses on the **Select responses** tab. You can also use the **Select response** box.
- Select which factor to display on the x-axes on the **Axes and constants** tab.
- Select interval estimation settings - Interval type, Confidence level, Tolerance proportion on the **Interval estimation** tab.
- Select if limits should be shown in the plot or not on the **Options** tab.
- Select if worksheet raw data points should be show or not on the **Options** tab.

Use the **Properties** pane factor sliders to interactively change the settings of the constant factors.

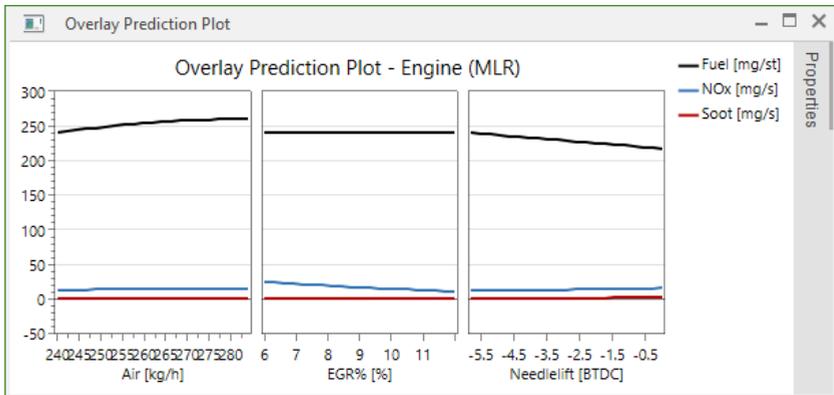
### Prediction plot

The prediction plot displays the predicted values of the selected responses, when the factors vary over their respective ranges. The plot supports up to three factors and an unlimited number of responses with confidence intervals when the **Show interval estimates** check box is selected. The plot also displays any worksheet points that are valid considering the constant settings of the factors not displayed on the axes in the subplot but in the **Properties** pane. The tolerance for the constant factors is specified by the replicate tolerance in MODDE options.



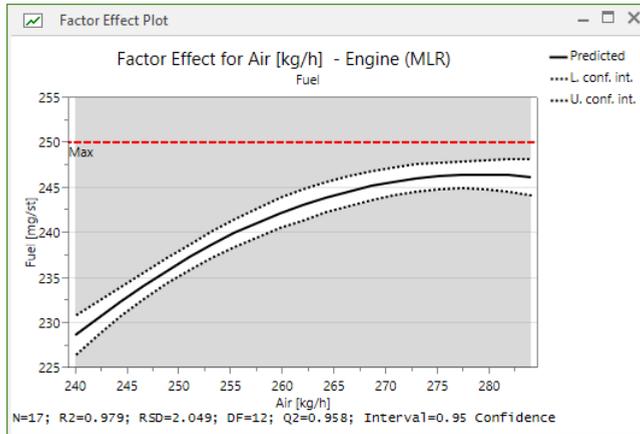
### Overlay prediction plot

The overlay prediction plot displays the predicted values for up to three factors and as many responses as desired. No confidence intervals can be displayed for the overlay plot.



## Factor effects

To access the Factor Effect Plot, on the **Predict** tab, in the **Interpretation** group, click **Factor effects**.



For process factors the plot displays the predicted values of the selected responses, when the factor varies over its range, all other factors in the design held constant at their averages. Qualitative factors held constant are set at their numerical average and do not influence the predicted values.

You can select to display the worksheet data in this plot, current factor vs. response, not taking into account the other factor settings. For Stability testing designs the worksheet data are by default displayed.

To display/hide the worksheet data, right-click the plot, click **Properties**, and on the **Options** tab select/clear the **Worksheet data** check box.

***Note:** The **Factor Effect Plot** by default shows both factors and the responses back transformed (in original units) while the **Main Effect Plot** shows the responses in their transformed units.*

### Factor effect plot for mixture factors

When you select to display a mixture factor  $X_i$ , this plot displays the predicted change in the response when  $X_i$  varies from its **low** to its **high** level, the relative amounts of all other mixture factors are kept in the same proportion as in the standard reference mixture.

For details, see the Statistical appendix.

### Switching displayed factors

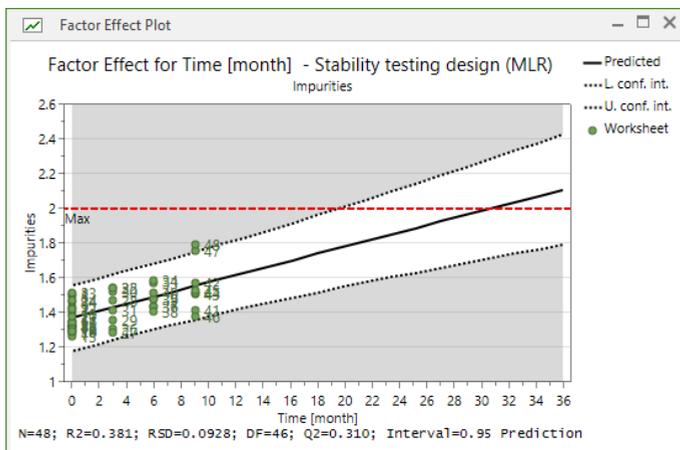
To switch factors,

1. Right-click the plot,
2. Click **Properties**,
3. Click the **Select factor** tab,
4. Select the desired factor,
5. Click **OK**.

Formulation factors can be adjusted proportional to the reference mixture or ranges on this property page.

## Factor effects, Stability testing

To access the Factor Effect Plot, on the **Predict** tab, in the **Interpretation** group, click **Factor effects**.



For stability testing designs the plot by default displays the predicted values of the selected response, when the *Time* factor varies from first to last time. All other factors in the design are held constant at their averages. Qualitative factors held constant are set at their numerical average and do not influence the predicted values.

You can select to display the worksheet data in this plot, current factor vs. response, not taking into account the other factor settings. For Stability testing designs the worksheet data are by default displayed.

To display/hide the worksheet data, right-click the plot, click **Properties**, and on the **Options** tab select/clear the **Worksheet data** check box.

**Note:** The **Factor Effect Plot** by default shows both factors and the responses back transformed (in original units) while the **Main Effect Plot** shows the responses in their transformed units.

### Switching displayed factors

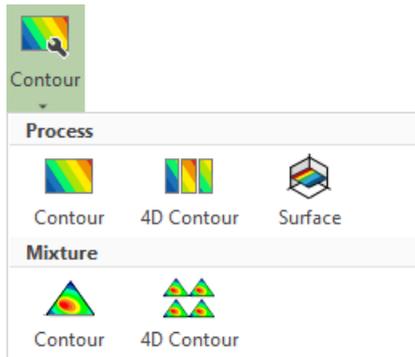
To switch factors,

1. Right-click the plot,
2. Click **Properties**,
3. Click the **Select factor** tab,
4. Select the desired factor,
5. Click **OK**.

For details about stability testing designs, see the Stability testing design section in the Generalized subset designs appendix.

## Contour

To open a contour plot, on the **Predict** tab, in the **Interpretation** group, click the arrow under **Contour**. Then click the desired plot.



The available contour plots for **Process** and **Mixture** are **Contour** and **4D Contour**. For **Process**, **Surface** is also available.

To open the **Contour Plot** wizard, click **Contour**.

### 2D contour

The 2D contour plot displays the predicted response values for the selected responses, spanned by two factors, in a response surface contour plot. For mixture the plot is spanned by three factors.

When Min, Target, or Max-values have been specified for the response, these limits are by default displayed in the plot.

#### **Factors kept constant**

By default, all constant factors are set at their mid-range values; change these values as desired in the Properties dialog or pane. After the plot has been created the constant factors can be seen (and changed) by clicking the **Properties** tag to the right of the plot to open the pane. When you change the value of the constant factors, the contour plots are updated.



## 4D contour

The **4D Contour** plot displays the predicted response values for the selected response, spanned by two factors, in 9 response surface contour plots in a 3x3 grid and spanned by another two factors. For mixture, the plots are spanned by three factors.

When Min, Target, or Max-values have been specified for the response, these limits are by default displayed in the plot.

*Note: The 4D plot is available for investigations with only process factors, and with both mixture and process factors, but not with mixture factors only.*

### 4D contour plot axes

When the model includes both process and mixture factors, the mixture factors can only be selected as inner axes factors in the **Contour Plot** wizard.

#### Inner axes factors

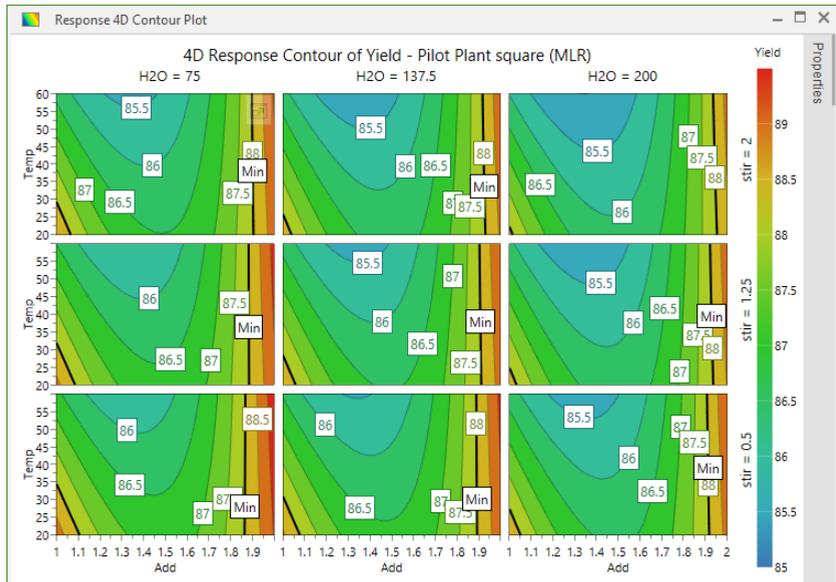
Select two process factors for the axes under **Factors at the plot axes**, or three mixture factors.

#### Outer axes factors

Select two process factors for the outer axes. The response contours are plotted for the low, middle, and high levels of these factors.

You can also select qualitative factors for the outer axes and select for which settings to display the contour plots.

When there are more than 4 factors in the model, the remaining factors are held constant.



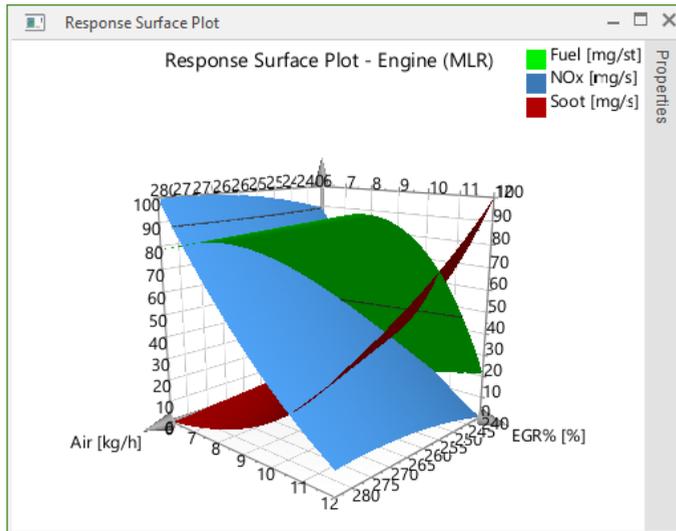
By default all plots are equally scaled, that is, the color coding is the same for all plots.

## Response surface

The **Response Surface Plot** displays the predicted response values, spanned by two factors, in a response surface plot. This plot is only available for models with two or more process factors and can display many responses.

### Factors kept constant

By default, all constant factors are held constant at their mid-range values; change these values as desired. After the plot has been created the constant factors are found to the right in the plot. When you change the value of the constant factors the contour surface plot is updated.



### Surface plot options

The plot options for the response surface plot are, as the Contour plot options, accessed by clicking **Plot options** in the **Contour Plot** wizard, or in tab **Surface plot options** in the properties dialog after creating the plot.

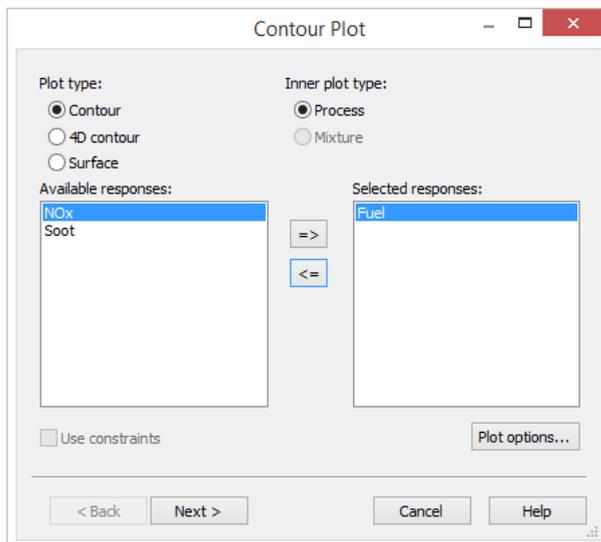
The Surface plot options include two options:

- Resolution** - The precision used when creating the contour plot. Selecting a higher resolution requires more calculations and is therefore more time consuming.  
To select the resolution of the plot, click one of the predefined resolutions or type a resolution value here.
- Scale surfaces 0-100** - How to scale the response axis. When selected (as in the example here), each response is scaled so that the smallest predicted value equals 0 and the largest predicted value equals 100. Note that with Min, Target and/or Max values in the response definition, these values are included in range used to rescale, possibly increasing that range. As a consequence the scale of the actual surface is less than 0-100 when a Min/Target/Max-value increases the range.

## Contour plot wizard

The contour plot wizard guides you through the selection and specification of 2D contour plots and 4D contour plots for mixture and process factors and 2D Surface plots for process factors.

To open the contour plot wizard, on the **Predict** or **Home** tab, click **Contour**.



### Inner plot type

Under **Inner plot type**, the default plot type for the inner axes is selected. If you have both process and formulation factors in the investigation, click the type of factor to display on the inner axes: **Process** or **Mixture**.

On the outer axes you can only vary process factors.

### Selecting responses

You can select to display all responses for the 2D contour plots but with more than 9 responses, the plots become very small. The contours are not overlaid, but displayed next to each other.

Modify which responses to display by marking in **Available responses** and clicking the arrow  and/or marking in **Selected responses** and clicking the back arrow .

### Using constraints

Clear the **Use constraints** check box if you do not want the available constraints to be displayed in the plot. MODDE's default is for constraints to be used and this should not be changed without reason.

### Plot options

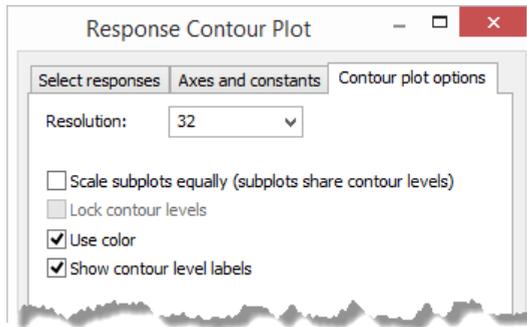
To select the resolution of the plot, that is, the grid calculated to create the contour plot, click **Plot Options**. Note that you can type a resolution value here or click one of the predefined resolutions.

You can also scale the subplots equally, to lock the contour levels, to produce the plot with/without color and with/without contour level labels.

For more, see the Contour plot options subsection.

## Contour plot options

Right-click the contour plot and click **Properties** to open the property page. The Contour plot options can also be accessed by clicking **Plot options** in the Contour plot wizard.



### **Resolution**

Resolution is the precision used when creating the contour plot. Selecting a higher resolution requires more calculations and is therefore more time consuming.

To select the resolution of the plot, click one of the predefined resolutions or type a resolution value here.

### **Scale subplots equally**

When the **Scale subplots equally** check box is selected the colors of all contour subplots represent the same values.

The plot has subplots when you display the 4D contour or you have selected to display more than one response in the 2D contour plot.

### **Lock contour levels**

You can select the **Lock contour levels** check box to keep the current contour level colors and limits.

This option is available when the **Scale subplots equally** check box has been selected.

### **Use color**

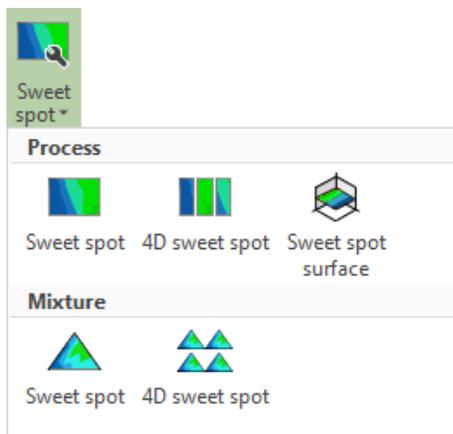
By clearing the **Use color** check box you can choose to display the contour and surface plots without colors.

### **Show contour level labels**

Clear the **Show contour levels labels** check box to display the contour plots without the level labels.

## Sweet spot

The **Sweet Spot** plot highlights the areas where the responses are within the user specified ranges but does not add the probability assessment that is done by the **Design Space** plot. The sweet spot plot can be displayed as 2D or 4D for process or mixture factors. For process factors the plot can also be displayed as a surface where the response intersections are viewed in 3D.

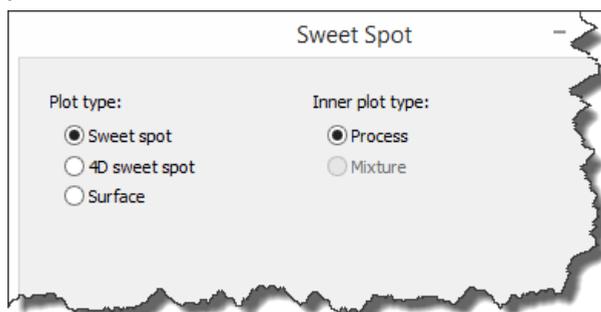


### Creating a sweet spot plot

To open the sweet spot plot wizard, on the **Predict** tab, click **Sweet spot**.

#### **Inner plot type**

If the investigation contains both mixture and process factors, select the type of factor you want on the axes for the 2D and for the inner axes for the 4D.



#### **Plot type**

To select to create the 2D sweet spot plot select **Sweet spot** and to select the 4D with up to four factors (5 with mixture) click **4D sweet spot**. To create a sweet spot surface, click **Surface**. Click **Next**.

## Response selection

In the **Responses** page, enter the settings for the relevant responses.

***Note:** If you have entered **Min** and/or **Max** values in the response definition, this page is automatically filled with those values. To update with updated values from the response definition, click **Get limits**.*

For each response select:

1. To **Include** or **Exclude** each response under **Incl/Excl**.
2. Type the values that are of interest under **Min** and **Max**. If you have entered the Min and Max yourself, MODDE uses those values. If you have not specified a Min and Max, MODDE enters the smallest and the largest values found in the worksheet and these should be changed based on your desired sweet spot.

You may also:

- Select or clear the **Use constraints** check box.
- Change the resolution in the **Resolution** box.
- Update to current limits in the response definition by clicking **Get limits**.

Responses

Set the thresholds for the responses that you want to display in the Sweet Spot plot.

Response	Incl/Excl	Min	Max
Fuel	Include		250
NOx	Include		25
Soot	Include		1

Use constraints    Get limits    Resolution: 256

< Back    Next >    Cancel    Help

### Plot axes

Click **Next** to open the **Axes and constants** page, or for 4D sweet spot, the **4D axes** page.

Mixture factors can only be selected as inner factors in a 4D plot.

Axes and constants

Factors at the plot axes:

	Low:	High:	
1st axis:	Air	240	284
2nd axis:	EGR%	6	12

Constant

Factor	Value	Setting
Needlelift	-2.89	<>

< Back   Next >   Finish   Cancel   Help

Select two process factors for the axes under **Factors at the plot axes**, or three mixture factors.

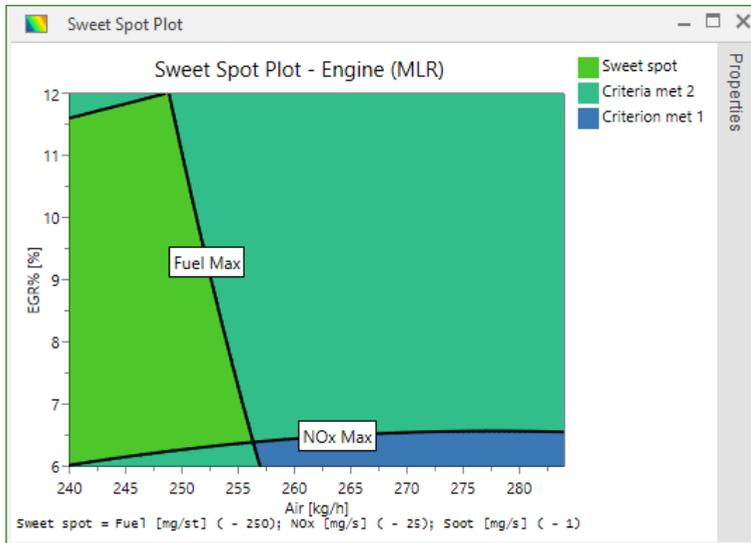
For the 4D, select two process factors for the outer axes. The sweet spot contours are plotted for the low, middle, and high levels of these factors.

### Colors in the sweet spot plot

The sweet spot plot uses the color scale from green to blue with:

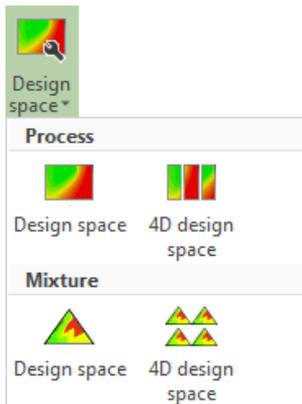
- Green for the 'sweet spot', that is the areas where all responses are within the selected range.
- Blue for areas where one of the responses is within its selected range.
- White for areas where none of the responses are within their selected ranges.
- Other colors for areas with more than one response within its range but not all.

The above are the default settings and you can define specific colors in the **Format Plot** dialog box.



## Design space

On the **Predict** tab, in the **Interpretation** group, click **Design space** to open the **Design Space Plot** dialog box with settings for creating the plot. Click the arrow under **Design space** to open a specific design space plot directly.



Available options are:

### Process

- Design space
- 4D design space

**Mixture** (only available when mixture factors are part of the design)

- Design space
- 4D design space

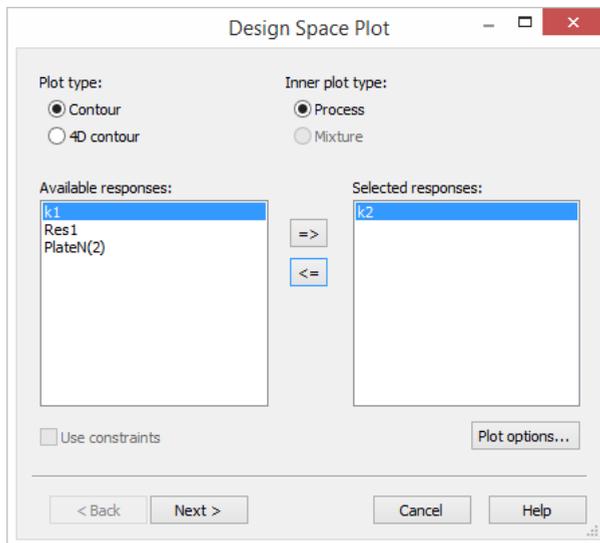
## Design space wizard

The **Design Space Plot** wizard opens automatically when clicking **Design space** on the **Home** or **Predict** tabs.

### First page

The first page of the **Design Space Plot** wizard lets you choose;

- **Plot type** - either Contour or 4D contour.
- **Inner plot type** - Process or Mixture.
- **Selected responses** - As many as desired.
- **Plot options** - see the Design space plot options subsection next.

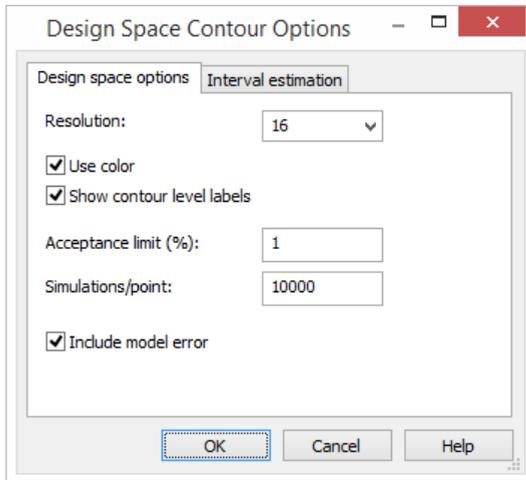


### *Design space plot options*

Click **Plot options...** on the first page in the wizard to open the plot options, alternatively right-click the created plot and click **Properties**. The **Design Space Contour Options** lets you specify,

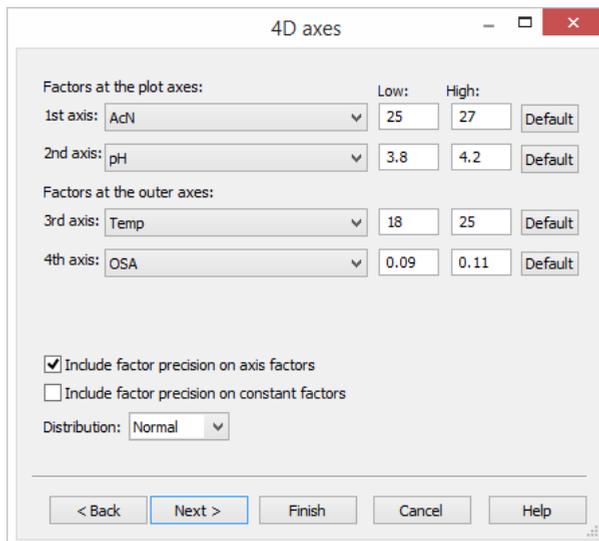
- **Resolution** - A higher resolution means a more detailed plot, but it can also slow down your computer.
- **Use color** - This determines if the plot should be colored or merely display the contour lines.
- **Show contour level labels** - Select to display contour level labels.
- **Acceptance limit** - The acceptance criteria aimed for either in percentage or in the DPMO space.
- **Simulations/point** - The number of simulations per point; the higher the more time consuming but also more accurate.

- **Include model error** - This determines if model error should be included in the plot. Note that when this check box is cleared, the plot merely presents the same information as the Sweet spot plot.
- Interval type, Confidence level and Tolerance proportion on the **Interval estimation** tab. For more, see the Interval type and probability levels subsection in chapter 4, File.



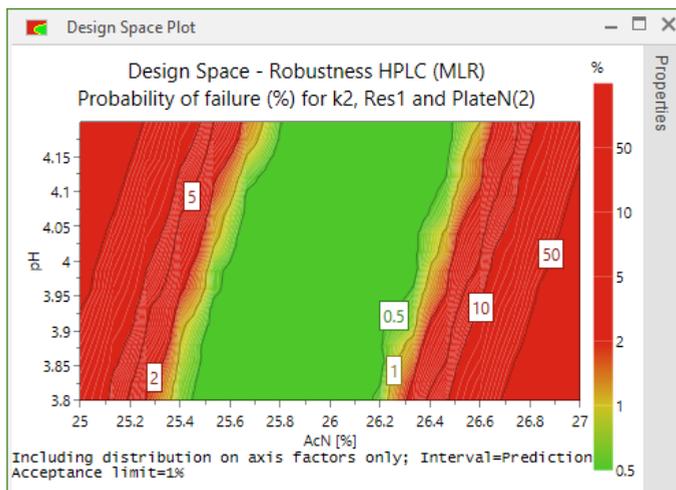
## 4D axes

The second page of the wizard lets you select what factors to display on each of the plot axes. In **Distribution**, the type of distribution can be selected; *Normal* (default), *Uniform* or *Triangular*.



**Note:** When creating the plot from the *Setpoint analysis* group on the **Optimizer** tab, you can also select to include distribution from the setpoint analysis on the axis factors and if you want to include distribution from setpoint analysis on constant factors.

Click **Finish** to see the **Design Space Plot** showing the probability of failure (%) for the shown factor combinations. Alternatively, **Next** to specify the settings for the constant factors.




---

## Optimizer

On the **Predict** tab, in the **Optimize** group, click **Optimizer** to open the **Optimizer** window and provide access to the **Optimizer** contextual tab.

More information about the optimizer is available in Chapter 12, Optimizer.

---

## Setpoint validation

On the **Predict** tab, click **Setpoint validation** to open the **Setpoint Validation** window.

Setpoint validation tests the robustness by making a large number of random disturbances (Monte Carlo simulation) in the specified region. The result is shown as a distribution of random samples including model prediction errors. The result can be expressed in general statistics as well as capability indices Cpk or Probability of failure.

Setpoint validation is a way to test if the investigated system is robust against disturbances in the investigated region.

The aim of robustness testing is to evaluate if a process, or a system, performs satisfactory even when some influential factors are allowed to vary. In other words, we want to investigate the system's sensitivity (or preferably lack of sensitivity) to changes in certain critical factors. The advantages of a robust process or system include simpler process control, a known range of applicability and an ensured quality of the product or process.

A robustness test is usually carried out before the release of an almost finished product, or analytical system, as a test to ensure quality. Umetrics recommends the use of DoE for robustness testing and such a design is usually centered on the factor combination, which is currently used for running the analytical system, or the process. We call this the setpoint. The setpoint may have been found through a screening design, an optimization design, or some other identification principle, such as written quality documentation. The aim of robustness testing is, therefore, to explore robustness close to the chosen setpoint.

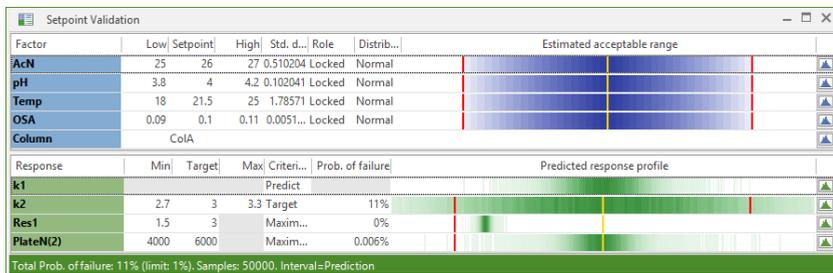
The regression model originates from a low resolution design supporting linear models since we assume that small disturbances have mainly linear effects. Fractional factorial resolution III and Plackett Burman designs are recommended.

## Setpoint validation plots and lists

The Setpoint validation features available for working with setpoint diagnostics are;

- Factor distribution
- Response distribution
- Setpoint validation list.

For more information on these functions, see Chapter 13, Setpoint.



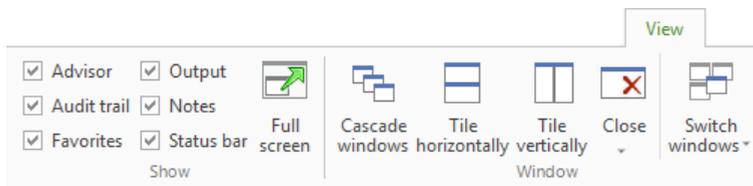


# 10-View

---

## Introduction

The **View** tab offers a selection of options for changing what windows are visible as well as choosing how to display open windows.



The available functions include,

- **Show** - Toggles the visibility of sections of MODDE such as the **Status bar** and **Advisor**.
- **Full screen** - Enter full screen mode.
- **Window** - Rearrange, close, or swap between MODDE's windows.

---

## Show

On the **View** tab in the **Show** group it is possible to choose to show or hide the following parts of the window:

- **Advisor** - The **Advisor** pane is automatically activated after fitting an investigation. The advisor explains a selection of the analysis plots and results available in MODDE.
- **Audit trail** - for details, see the Audit trail subsection later in this chapter.
- **Favorites** - for details, see the Favorites section later in this chapter.
- **Output** - The **Output** pane is a log book of the session. All MODDE messages and actions are recorded in the logbook.
- **Notes** - In the **Notes** pane you can record your own notes concerning the investigation. You can paste MODDE plots and lists. Note the commands available by right-clicking the window. This file can then be saved as .rft (Rich Text Format) and read directly by a word processor with all plots.
- **Status bar** - Displays details about the current investigation. Note that clicking the information opens a list/spreadsheet or dialog box. For instance, clicking **Factors** opens the **Factors** spreadsheet, clicking **Objective** opens the **Design Wizard** on the **Select objective** page.

All of the items above (with the exception of the status bar) can be moved around to other sections of the viewable area by clicking and dragging them. To make the windows smaller, pull the top/side and it will remember its size. Double-click the window caption to make it floating or make it dock.

---

**Hint:** While dragging the window hold down the **CTRL** key. It won't dock if you are close to the frame of the main frame window.

---



---

## Audit trail

When the **Audit Trail** is turned on each investigation in MODDE has a separate audit trail. Each audit trail consists of one or more sessions that in turn consist of events. A new session is started and appended to the audit trail when an investigation is opened, and ends when the investigation is saved.

In addition to logging events, MODDE logs information about the user, and date and time of the events.

To view the audit trail, click the **Audit Trail** tab in the **Output / Notes / Audit trail** pane. If this tab is not shown, display it by selecting the **Audit Trail** check box on the **View** tab.

### **Enable and disable the Audit Trail**

By default the audit trail is disabled.

To turn it on for the current investigation, in **File | Options**, on the **Investigation options** page, under Audit trail, select *Yes* in **Enable the audit trail**.

To turn on the audit trail for new investigations, in **File | Options**, on the **MODDE options** page, under Audit trail, select *Yes* in **Enable the audit trail for new investigations**.

Administrators can disable the turning on and off of the audit trail, i.e., always have it on or off. For instructions on how to disable the audit trail options, see the knowledge base at [www.umetrics.com](http://www.umetrics.com).

### **Save or clear the Audit Trail**

To empty the audit trail, in the Audit Trail pane, right-click and click **Clear Audit Trail**.

To save the current version of the audit trail, separate from the investigation, in XML format, in the Audit Trail pane, right-click and click **Save as**.

### **Audit trail and internet explorer**

MODDE uses Internet Explorer functionality to display the audit trail.

### **Logged in the audit trail**

Specific actions the MODDE audit trail logs are:

- Factors (adding, modifying, deleting), displaying all details about the factor after the change
- Responses (adding, modifying, deleting), displaying all details about the response after the change
- Constraints (modifying, deleting)

- Candidate set
- Inclusions
- Reference mixture
- Objective
- Generators
- Design
- Complement design
- Model
- Worksheet, every change of every cell
- Fit method
- Activation and deactivation of the Audit Trail.
- The audit trail also registers when a digital signature in the Audit trail is incorrect.

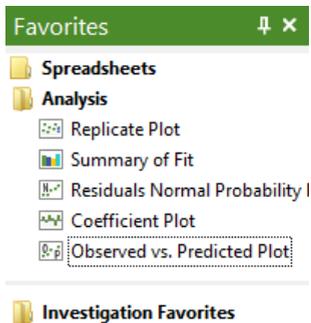
---

## Favorites

### Introduction

To open or close the Favorites pane, on the **View** tab, select the **Favorites** check box.

The **Favorites** pane by default includes a few plots, lists and spreadsheets and can be extended and modified as desired.



Open a plot or list by clicking it. When settings must be selected, a dialog box opens; select the appropriate settings and click **OK**. Plots and lists are displayed with the selected settings.

Almost all plots in MODDE can be added to favorites.

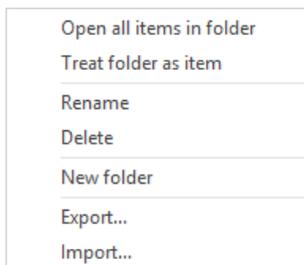
### **Adding favorites**

To add an open plot or list to favorites, on the **Tools** tab, click **Add to favorites** or right-click and click **Add to favorites**. Note that this adds the plot with the selected settings to the **Favorites** pane.

## ***Opening favorites***

To open a favorite, double-click it or mark it, right-click and click **Open**.

Right-clicking the **Favorites** pane opens the menu below. A description of the menu items follows.



### ***Open all items in folder – Executing a folder***

To execute all commands in a folder, right click the folder and select **Open all items in folder**.

For example with the default Analysis folder, opening all items in the folder displays the replicate, summary of fit, residuals normal probability, coefficient and observed vs. predicted plots for the first response tiled.

### ***Treat folder as item***

Click **Treat folder as item** to treat the folder as an item. The folder is then displayed as an item:

 **Analysis**

Click it to open all the items. This gives the same result as **Open all items in folder** but with a single click.

### ***Rename***

All folders, plots, spreadsheets, and lists can be renamed according to your wishes by right-clicking it and clicking **Rename**, or marking it and pressing the F2-key on your keyboard.

### ***Delete***

All items in **Favorites** can be deleted by right-clicking the item and clicking **Delete**, or marking it and pressing the DELETE-key on your keyboard.

### ***Create a new folder***

It is convenient to group commands in folders, and automatically execute all the commands in the folder in sequence. To create a folder in **Favorites**, right-click and click **New folder**.

The created folder is default named 'New folder'; change the name as desired, for example to 'Residuals Plots'.

To move favorites to the Residuals Plots folder, use the drag-and-drop feature.

### **Importing and exporting Favorites configuration**

The Favorites window configuration can be saved as an .xml-file.

To save the current favorites configuration to file:

1. Right-click the **Favorites** window.
2. Click **Export**.
3. Enter the name and location in the **Save As** dialog,
4. Click **Save**.

---

*Note: When importing a favorites file, the new favorites will replace the current favorites. If you want to keep your current configuration and switch back to it later, export to file before importing a new one.*

---

To import favorites from .xml-file:

1. Right-click the **Favorites** window.
2. Click **Import**.
3. Browse for the file in the **Open** dialog.
4. Click **Open**.

### **Restoring favorites**

To restore **Favorites** to the MODDE default, on the **File** tab, click **Options** and on the **Restore** page click **Restore** in the Favorites section.

---

## **Full screen**

You can use the **Full screen** command to maximize the plot area. **Full screen** can be toggled on and off using **F11** and is available on the **View** tab. Use **F11** to leave full screen mode.

---

## **Window**

On the **View** tab in the **Window** group there are icons to choose how open windows are arranged;

- Cascade windows,
- Tile horizontally,
- Tile vertically.

### **Close**

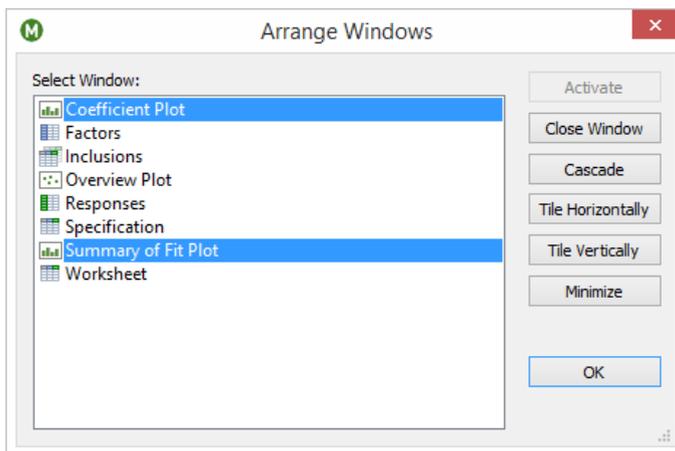
Clicking **Close** closes the active window. Clicking the arrow under **Close** offers other close options;

- Close all windows,
- Close all but active,
- Close all plots,
- Close all lists.

## Switch windows

**Switch windows** offers a list of open windows. The selected window will be brought to the front.

The **Switch windows** function also offers an option to **Arrange windows**. Selecting this brings up the **Arrange Windows** dialog box where you can select a specific window to bring to the front (**Activate**) or to close or minimize selected windows. If you select more than one window (using the normal windows method of Shift or Ctrl to select multiple items) the **Cascade**, **Tile Horizontally** and **Tile Vertically** options become available.

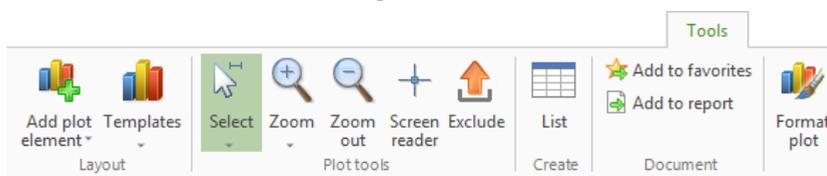


# 11-Tools

---

## Introduction

The **Tools** tab provides functions to interact with the currently selected plot or list. If a function is not available for the active plot or list, it will also be unavailable on the tab.



The functions are organized into groups as follows,

### Layout group

- **Add plot element** - add or remove elements such as headers and footers to the currently selected plot.
- **Templates** - save or load a template.

### Plot tools group

- **Select** - selection tool, such as free form, or x-axis.
- **Zoom** - Zoom in, zoom X, zoom Y, zoom subplot.
- **Zoom out** - zooms out if zoomed in.
- **Screen reader** - get the coordinates from a specific point on a plot.
- **Exclude** - turn on/off exclude mode or exclude selected data.

### Create group

- **List** - creates a list for the active plot.

### Document group

- **Add to favorites** - add the currently active plot or list to the favorites list.
- **Add to report** - add the currently selected plot or list to a report.

**Format plot** - opens the **Format Plot** dialog box.

---

## Add plot element

Add or remove plot elements such as headers, footers, and legend.

To add new plot elements to a plot:

1. On the **Tools** tab, in the **Layout** group, click **Add plot element**.
2. Select the desired plot elements.

Alternatively:

1. Click the plot window to make the mini toolbar appear.
2. Click the add plot element icon .

## Available plot elements

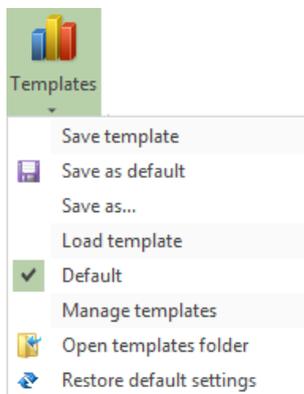
- Header
- Footer
- Legend
- Axis titles
- Axes
- Regression line
- Timestamp.

---

## Templates

On the **Tools** tab, in the **Layout** group, click **Templates** to display options to save, load, and manage plot templates.

The plot template specifies the visualization of the plots (fonts, colors and sizes for headers, background color, legend alignment, etc.).



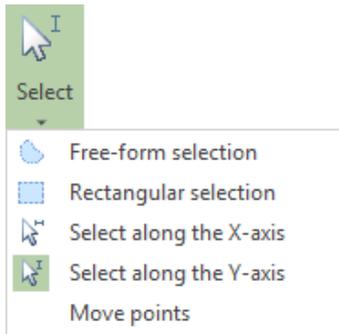
**Save template** - You can choose to save the current template under the current name, as default, or as anything else. Note that the *Default* template is marked, even if change have been introduced, until you save the template under a different name.

**Load template** - Choose between previously saved templates or the default template. The current template is marked.

**Manage templates** - Open the templates folder for browsing or **Restore default settings**. **Restore default settings** restores and switches to the settings that MODDE had when installed, not changing any custom templates.

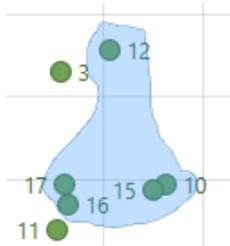
## Select

On the **Tools** tab, in the **Plot tools** group, click the arrow under **Select** to view available selection methods.



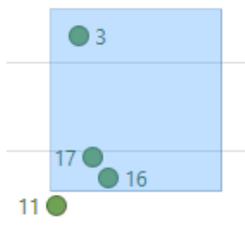
### Free-form selection

Click a plot area with the left mouse button and drag to create a free form selection area. Release the mouse button to select everything within that area.



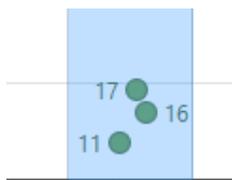
### Rectangular selection

Click a plot area with the left mouse button and drag to create a rectangular selection area. Release the mouse button to select everything within that area.



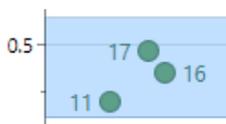
### Select along the X-axis

Click a plot area with the left mouse button and drag left or right, along the x-axis. Release the mouse button to select everything within that area.



### Select along the Y-axis

Click a plot area with the left mouse button and drag up or down, along the y-axis. Release the mouse button to select everything within that area.



### Move points

When the active plot is the **Dynamic profile**, the **Move points** option is available. Click a point with your left mouse button and drag the point to the desired location. Release the mouse button and the **Dynamic profile** plot and the **Optimizer** window will update with the new **Setpoint** location.

---

## Zoom and zoom out

To zoom in or out in a plot, **Zoom** and **Zoom out** are available on the **Tools** tab in the **Plot tools** group. Click the arrow under **Zoom** to access the different zooming options.

To undo the zoom and return to the previous zoom level, click **Zoom out**.

Zoom subplot by activating the zooming tool and double-clicking the subplot.



### Zoom and rotate 3D plots

Turn the 3D scatter plot or response surface plot by holding down the left mouse-button and moving the mouse in the direction you want to turn the plot.

Zoom in or out in the plot by using the mouse wheel.

To reset the rotation to default, right-click and click **Reset rotation**.

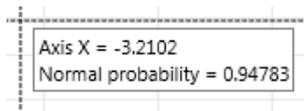
---

## Screen reader

On the **Tools** tab, in the **Plot tools** group, click **Screen reader** to change the select tool to the screen reader tool.



This presents information relevant for that specific plot at the point the mouse pointer is located.




---

## Exclude

**Exclude** is located on the **Tools** tab, in the **Plot tools** group and in the **Editing** group on the **Home** tab.



Use **Exclude** to remove unwanted model terms or experiments from plots and charts. **Exclude** can either be clicked before selecting points or after selecting points. When clicked before selecting points, **Exclude** remains active allowing for the repeated removal of model terms.

Click **Undo** to read newly removed model terms or experiments.

---

## Format plot

Format plot allows for you to change plot details such as the legend, error bars, columns, styles, gridlines, background, etc. The options available when using format plot depend on the plot that is currently active. The **Format Plot** dialog box provides options for completely changing the look of your plot. Everything from the footer, timestamp, headers, subheaders, shape of the points, etc can be changed. MODDE allows you to completely customize your plots so that they will best fit into your report. Any element in a plot that you can click, can likely be customized.

### Accessing format plot

To access the format plot options for the currently active plot,

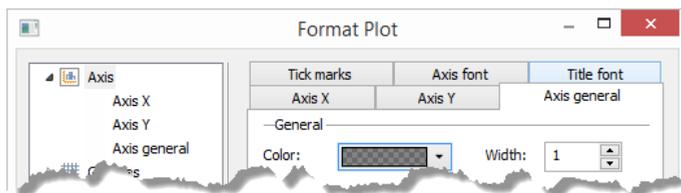
- On the **Tools** tab, click **Format plot**.
- Right-click a plot and select **Format plot** from the menu.
- Click a plot element and click the **Format plot** on the mini toolbar. This will take you directly to the section of the **Format Plot** dialog box that lets you specify how the element looks.

## Mini toolbar

Some elements in plots can be customized individually without using the **Format Plot** dialog box. This is useful if you would like one specific data point to stand out with a different shape, size, glow, or color. To see if an element can be individually customized, you can click the element, then look at the mini toolbar to see the available options. See Chapter 1, Getting started for more information on the mini toolbar options.

## Axis

Click **Axis** in the **Format Plot** dialog box window to open up the options for the axes.

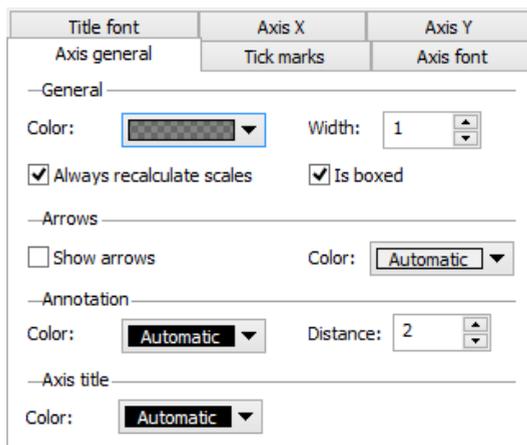


The available tabs on **Axis** are

- Axis general
- Axis X & Axis Y
- Tick marks
- Title font & Axis font.

### *Axis general*

The **Axis general** page under **Axis** in the **Format Plot** dialog box presents general axis options for such things as color and width of the axes. There are check boxes for **Always recalculate scales** and **Is boxed**. You can select if arrows should be shown and their color. For axes titles, the color for them can be selected as well.



### Axis X and Axis Y

On the **Axis X** and **Axis Y** tabs under **Axis** in the **Format Plot** dialog box you can choose to show the axis as well as various settings for each axis such as the title, rotation, step size, and if the axis should be reversed.

—Axis—

Show axis

Minimum:  Maximum:

Auto adjust scales for suitable limits and step size

Step size:

Reverse axis

—Annotation—

Rotation:

—Title—

Show title

Title:

### Tick marks

The **Tick marks** tab under **Axis** in the **Format Plot** dialog box lets you choose the tick mark type and size for both minor and major tick marks. Available tick mark types are **None**, **Outside**, **Inside**, and **Cross**.

—Major tick marks—

Tick mark type:

Tick mark size:

—Minor tick marks—

Tick mark type:

Tick mark size:

## Axis Font and Title Font

The **Axis Font** and **Title Font** tabs under **Axis** in the **Format Plot** dialog box let you choose the **Font**, **Font style**, and **Size** of the axis font or title font.

Font:	Font style:	Size:
<ul style="list-style-type: none"> <li>Segoe UI</li> <li>Rage Italic</li> <li>Ravie</li> <li>Rockwell</li> <li>Rockwell Condensed</li> <li>Rockwell Extra Bold</li> <li>Rod</li> <li>Sakkal Majalla</li> <li>Script MT Bold</li> <li>Segoe Print</li> <li>Segoe Script</li> <li>Segoe UI</li> <li>Segoe UI Emoji</li> </ul>	<ul style="list-style-type: none"> <li>Regular</li> <li>Regular</li> <li>Bold</li> <li>Italic</li> <li>Bold Italic</li> </ul>	<ul style="list-style-type: none"> <li>8</li> <li>7</li> <li>8</li> <li>9</li> <li>10</li> <li>11</li> <li>12</li> <li>14</li> <li>16</li> <li>18</li> <li>20</li> <li>22</li> <li>24</li> </ul>
Preview:		
<div style="border: 1px solid gray; padding: 5px; display: inline-block;">AaBbYyZz</div>		
<input type="checkbox"/> Scalable <input type="checkbox"/> Rotatable <input checked="" type="checkbox"/> Anti aliased <input checked="" type="checkbox"/> Filled		

## Gridlines

Gridlines can be controlled either for the entire graph using the main category gridlines, or specifically by using vertical and horizontal. The **Gridlines** page in the **Format Plot** dialog box lets you choose to turn on or off gridlines, the type of gridline, if grid stripes should be used and if the gridlines should be drawn behind the plot or not. Available gridline types are **Solid**, **Dashed**, **Dotted**, **Dash dot**, and **Dash dot dot**.

—Gridlines—

No gridlines  
 Gridline
 Solid

Color: [Color swatch]
 Width: 1

—Grid stripes—

No grid stripes  
 Solid fill  
 Gradient fill
 [Dropdown]

Color: [Color swatch]
 Color 2: [Color swatch]

—General—

Draw behind plot

## Background

The **Background** page of the **Format Plot** dialog box lets you change the fill and border of the window and plot areas of the current plot.

—Fill—

No fill

Solid fill

Gradient fill Vertical

Color:   Color 2:  

—Border—

No border

Solid line

Color:   Width: 1

## Titles and Font

The **Title** section of the **Format Plot** dialog box includes the titles and font pages. These are available for either the entire graph or sections of the graph; **Header**, **Footer**, **Timestamp**, and **Subheader**. The visibility, background color, text color, border, location, and the text itself can be changed.

Titles Font

—Title—

Is visible Anchor: Top

Summary of Fit - Engine (MLR)

≡ ≡ ≡

—Color—

Text color: Automatic

Background color: Automatic

—Border—

Is visible Color: Automatic

Margin: 5 Width: 2

## Legend and Font

The **Legend** page of the **Format Plot** dialog box lets you choose whether or not to show the legend and the position, orientation, text alignment, font, text and background color, and border of the plot's legend.

Legend Font

—Placement—

Show legend (using default rules)

Position: Top right

Orientation: Vertical

Text alignment: Left

—Color—

Text color: Automatic

Background color: Automatic

—Border—

Is visible Color: [Color Picker]

Margin: 2 Width: 1

## Limits and regions

The **Limits and regions** section of the **Format Plot** dialog box lets you change the options associated with standard deviation lines, min and max lines, target lines, etc. The fill can be specified for over or under a line and a gradient with two colors selected.

Region style Font Label style

—Style—

Line: Style: Dashed Width: 2

Color: [Red]

Fill: Type: No fill

Style: Solid

Gradient: Vertical

Color 1: [Color Picker]

Color 2: [Color Picker]

## Labels

If the plot has any labels, the **Labels** option will be available in the **Format Plot** dialog box. In the **General** tab, the option **Avoid overlapping labels** is by default selected and limited to 100 points.

The image shows a dialog box titled 'Label style' with two tabs: 'Font' and 'General'. The 'General' tab is active. The dialog is organized into sections separated by horizontal lines:

- Label**
  - Position: Right (dropdown menu)
  - Rotation: 0 (spin box)
  - Offset: 4 (spin box)
- Color**
  - Text color: Automatic (dropdown menu)
  - Background color: Automatic (dropdown menu)
- Border**
  - Is visible
  - Color: (color selection box)
  - Margin: 1 (spin box)
  - Width: 1 (spin box)
  - Draw connection line

## Contour

If the currently active plot is a contour plot, then the **Contour** page is available in the **Format Plot** dialog box. Here it is possible to select the colors to be used in the selected contour plot.

---

**Hint:** You can create a custom level for the contour plot by typing a number in the field below **Add** and then clicking **Add**.

---

The default plot displays a number of contour levels. To display fewer or more contour levels, type a new value in the **Contour levels** field. In **Individual level colors** the new number of levels will be displayed.

---

**Note:** To for instance display the contour levels for 10, 20, 30 ...100 you need to type **Min=0 and Max=110 and in the Contour levels field type 10**.

---

The **Color Begin** and **End** colors define the color scale used in the plot.

To change from the default color range, click the respective *Begin* and *End* colors, and click the new color to use.

To change the color of one of the available levels under **Individual level colors**, mark the level and click the colored button to the right (above **Remove**), and then click the new color.

To remove one of the available levels, mark it and click **Remove**.

Click the **Contour level line style** tab to customize the contour line color, width, and pattern.

The screenshot shows the 'Contour' dialog box with two tabs: 'Contour levels' (selected) and 'Contour level line style'. The 'Contour levels' tab contains the following controls:

- Contour levels:** A numeric input field set to '8' with up and down arrow buttons.
- Range:** Two input fields for 'Min' (set to 215) and 'Max' (set to 260).
- Color:** Two color selection buttons, one blue and one red.
- Individual level color:** A list of color swatches for levels 220, 225, 230, 235, 240, 245, 250, 255, and Above. To the right of this list are three buttons: a color selection button, a 'Remove' button, and an 'Add' button. Below the 'Add' button is an empty input field for entering a new level value.

## Error bars

When the active plot uses error bars, these can be customized in the **Error bars** section of the **Format Plot** dialog box. The error bars can have their visibility set, the color, the line width, and the error bar width as a percent of the column width.

The screenshot shows the 'Error bars' section of a dialog box. It contains the following elements:

- A checkbox labeled 'Is visible' which is checked.
- A 'Color:' label followed by a dropdown menu currently showing 'Automatic'.
- A 'Line width:' label followed by a spinner control set to the value '1'.
- An 'Error bar width:' label followed by a spinner control set to the value '60', with the text '(% of column width)' to its right.

## Column

The **Column** option in the **Format Plot** dialog box becomes available when the currently selected plot uses columns. You can change the column width.

## Styles

The **Styles** section of the **Format Plot** dialog box provides specific styling of elements in the currently selected plot. The available styling varies greatly based on the plot selected. The Histogram plot, for example, allows the Coefficient columns to be styled.

---

## List

To create a list of the currently selected plot, on the **Tools** tab, in the **Create** group, click **List**.

*Note: This option is available for most plots.*

---



---

## Add to favorites

To add the active plot or list to **Favorites**, on the **Tools** tab, in the **Document** group, click **Add to favorites**.

---

## Add to report

To add the active plot or list to your report, on the **Tools** tab, in the **Document** group, click **Add to report**. If no report exists, the **Add to report** icon will start the **Report** and guide you through that process.

---

## Plots and lists

### Properties dialog box

To open the property page of a plot or list, right-click the plot or list and click **Properties**.

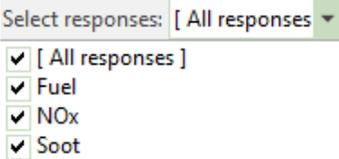
With **Properties** open, change as desired and the plot or list is updated when you click **OK** or **Apply**. The default of most properties available from the **Properties** page can be changed in Investigation options (**File | Options**).

### Automatic update of plots and lists

If you, after fitting the model, make changes to the responses (add, delete or transform), the worksheet (include, exclude runs or values), the model (add or remove terms), or change fit method, MODDE refits the model and all open plots or lists are updated.

### Generating multiple plots or lists

When there are many responses in your investigation you can select to display a multiple plot with all or selected responses. Click the desired responses or *All responses* in the **Select responses** box.



Select responses: [ All responses ]

- [ All responses ]
- Fuel
- NOx
- Soot

# 12-Optimizer

---

## Introduction

The **Optimizer** helps with finding the optimal conditions or best compromise as a setpoint and alternatively the most robust setpoint.

### Starting the optimizer

To open the **Optimizer**, on either the **Predict** tab or the **Home** tab, click **Optimizer**. MODDE then opens the **Optimizer** window and tab for the optimization analysis.

### Optimizer theory

The optimizer works according to a given set of specifications and the specification of the factors and responses are selected according to the desired result. Therefore, if the response specifications are unrealistic, it might be impossible for the optimizer to reach the best possible solution. With a good strategy and by using complementary tools, such as contour plots, probability of failure estimates, sweet spot plots, Design Space estimates, robust setpoint, setpoint analysis and the predicted min/predicted max values listed in the Optimizer, a good understanding of the possible specifications can be obtained. Note here that the most important consideration in order to obtain optimal optimization results is to certify that the *Pred. min* and *Pred. max* range and the response specifications at least partly overlap.

The optimizer is used to find an experimental *setpoint* that fulfills various criteria. The optimizer uses a search function to find the best possible solution to an equation that depends on a number of operating criteria.

The following section describes the possibilities and limitations of the optimizer function. The first part is a description of how the optimizer works and the second part discusses how different objectives can be reached by selecting different start criteria for the optimization.

### Search function

The optimizer uses desirability functions,  $d_k$ , for each response,  $k=1, \dots, m$ , and searches for the combination of factor settings that predicts a result inside the response specifications and as close as possible to the targets for all responses. When searching for a solution with many criteria, the result will be a compromise between those criteria. This compromise is expressed as the *overall desirability function*,  $f(\mathbf{ds})$ , a sum of all  $d_k$ . This compromise is also expressed as a normalized distance to target ( $D$ ) for all responses.

The success of the desirability function depends on the optimizer specification (Min, Target, Max) and the selected **Desirability** objective (Limit/Target/Custom/etc.). It must be possible to reach the optimizer objective for the current data in order for the desirability function to succeed.

The search for a robust setpoint, **Find robust setpoint**, is based on Monte Carlo simulations and is available if a setpoint can be found that predicts all responses within their limits.

## Optimizer objectives

The optimizer can be set up for different objectives:

1. *Limit optimization* – where the objective is to reach a solution in which the response is within the specification limits (Min and Max limits). This is the default approach in MODDE.
2. *Target optimization* – where the objective is to reach a solution in which the response is as close to target as possible. For the target optimization to work properly, it is necessary that the response can be optimized close to or on target; otherwise the search may end up with an unacceptable solution.
3. *Custom optimization* – user defined customization of the Target optimization.
4. *Focus optimization* – where the objective is to favor one or several responses over the others; accomplished by manipulating the individual weights.
5. *Robust setpoint* – where the most robust setpoint is found. Depends on the existence of a solution based on objectives 1-4.

All of the above optimization types are described in the Optimizer appendix, while the Robust setpoint is described in the Find robust setpoint subsection later in this chapter.

To control the optimization, the overall desirability function,  $f(ds)$ , plays a key role, as well as reasonable limits and targets for the responses. The optimizer will strive to reach the lowest possible value of the overall desirability function,  $f(ds)$ , and will strive to reach the lowest possible value. The shape of the function is controlled by the settings of the criteria (Min, Target, Max) for each response and the choice of the  $d_k$  functions.

With weight 1, the lowest possible value of the individual desirability functions,  $d_k$ , is -100. Note that in the plots, the individual desirability function,  $d_k$ , is translated to [-1, 0], so that a weight of 1 has a lowest possible value in the plot of -1, and so on. The goal is to reach a minimum in the overall desirability function,  $f(ds)$ .

---

## Optimizer window

To open the **Optimizer** window, on either the **Predict** tab or the **Home** tab, click **Optimizer**.

MODDE then opens the **Optimizer** window and tab for the optimization analysis.

The start specifications are from the initial factor and response definitions. If no response specification for Min, Target, and/or Max exists the default criteria is 'Predict'.

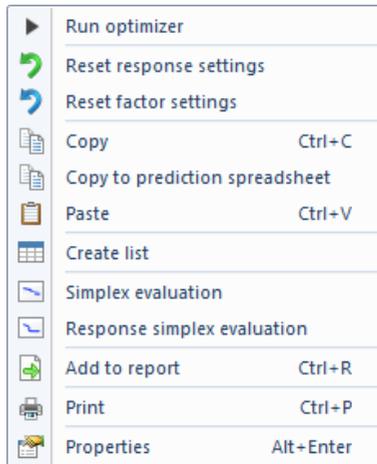
Reaching an optimal result is in many cases an iterative process. If the response specifications are impossible to reach the criteria will probably have to be reevaluated. With the help of some raw data analysis and some initial model analysis, you can get a reasonable understanding of the possibilities.

The Optimizer window is made up of three tabs,

- Objective,
- Setpoint, and
- Alternative setpoints.

### **Optimizer options**

Right-click the optimizer window to access some options for the Optimizer.



**Reset response settings** will reset the settings to the current Min, Target and Max of the response definition.

**Reset factor setting** will reset the factor settings to the low and high values in the factor definition. For CCC and CCO designs, the star point values are used as low and high.

**Copy to prediction spreadsheet** will add the list of factor settings from the **Alternative setpoint** spreadsheet to the **Prediction Spreadsheet**.

**Create list** opens the Optimizer list summarizing the results.

**Simplex evaluation** displays the Log(D) plotted vs. iterations for all runs in the alternative setpoint list.

**Response simplex evaluation** focuses on one response and displays the Log(D) plotted vs. iterations for all runs in the alternative setpoint list.

**Add to report** will add the current optimizer setup and selected setpoint to MODDE's Report.

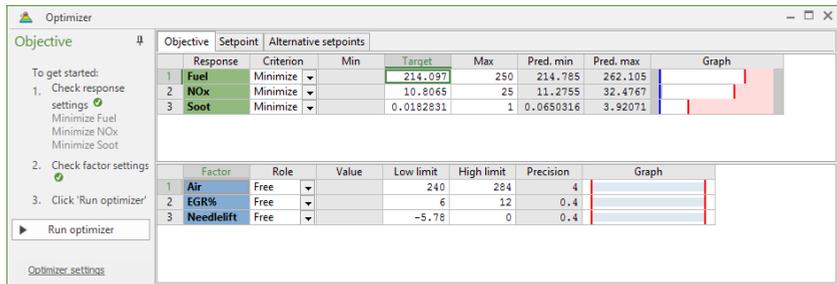
**Properties** opens the Optimizer Properties dialog box.

## Objective tab

The first tab of the **Optimizer** window is the **Objective** tab. The top half of the tab is occupied by a response spreadsheet while the bottom half has a factor spreadsheet.

Change the weights, targets, min, max, etc as desired then click **Run optimizer**.

If you for instance want to minimize a response, but only have the Max-value, MODDE will suggest a Target-value for you, slightly smaller than the calculated predicted minimum value, to make the Optimization run smoothly.

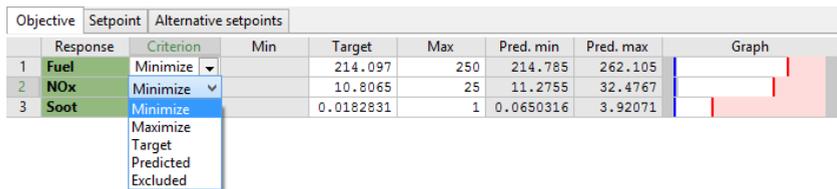


## Response spreadsheet

In the response spreadsheet all the responses used in your model are available. Before starting the optimization you must select the criteria, weights and limits for your responses.

If you have specified **Min**, **Target**, and/or **Max** in the response definition, these specifications are copied to the **Optimizer** response section. If no **Min**, **Target**, or **Max** settings are defined in the response definition, the response will be by default *Predicted*.

*Note: You can fetch updated limits from the response definition by right-clicking and clicking **Reset response settings**.*



## Criteria and limits

You can choose:

- **Minimize** the response. Type the highest value you can accept under Max and your target value under Target.
- **Maximize** the response. Type the smallest value you can accept under Min and your target value under Target.
- **Target** the response. Type the smallest, largest, resp. desired value under Min, Max, resp. Target.
- **Predicted** response. The response will not participate in the optimization but the prediction will be displayed.

- **Excluded** response. The response will not be a part of the optimization nor displayed in the run list.

When you have not entered **Min**, **Target**, or **Max** values and you choose to **Minimize** or **Maximize**, MODDE makes an educated guess of your limits, that you should modify to suit your desires. It is imperative that the limits are well chosen and reflect the data at hand.

When the difference between the minimum/maximum and target values is too small the optimizer will not run. When this happens you should increase the range.

**Pred. min** and **Pred. max** columns show what MODDE can find as the highest and lowest predicted value for the specific response using the final model. This is a help function to compare what is possible to reach in the investigated region compared with the desire for the specific response. If the goal for the response is outside the **Pred. min** and **Pred. max** the optimization will not work.

The column **Graph** presents the possibilities and the desire for each response. The white area is the union of possible range and specification. Gray is out of predicted range and pink is possible but outside specification. The red line is the min/max specification and the blue line is the target.

The **Weight** and **Desirability** columns are hidden by default and can be shown by changing the **Optimizer settings**.

For information about the **Desirability** column. See the Optimizer desirability section below.

## Optimizer desirability

The optimizer can be set up for different objectives using the **Desirability** column in the responses spreadsheet in the **Optimizer** window:

The optimizer can be set up for different objectives:

1. *Limit optimization* – where the objective is to reach a solution in which the response is within the specification limits (Min and Max limits). This is the default approach in MODDE.
2. *Target optimization* – where the objective is to reach a solution in which the response is as close to target as possible. For the target optimization to work properly, it is necessary that the response can be optimized close to or on target; otherwise the search may end up with an unacceptable solution.
3. *Custom optimization* – user defined customization of the Target optimization.

---

**Note:** The **Desirability** column is hidden by default. To show the **Desirability** column, change the **Optimizer settings**.

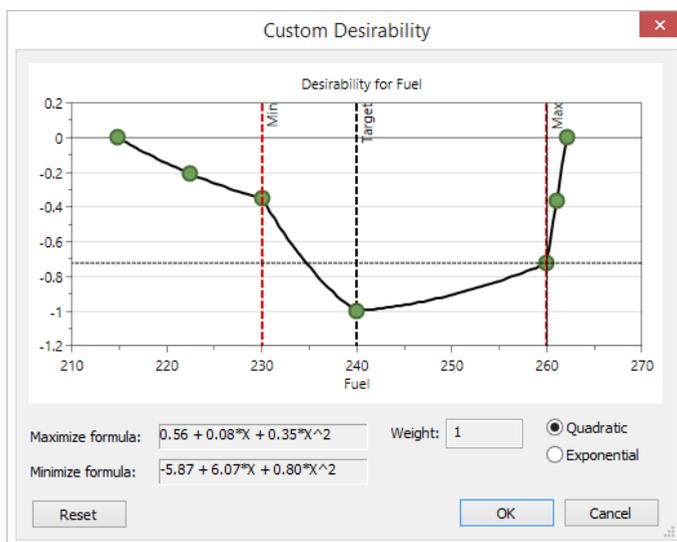
---

The optimizer will use a function that is called *Desirability*; this function can be set for different objectives or changed by the user by selecting *Custom* in the **Desirability** column in the **Objective** tab of the **Optimizer** window.



The exponential function corresponds to Limit optimization and is optimal for a compromise when all responses shall be within the specification limit. The quadratic function corresponds to the Target optimization meaning that all responses are expected to reach target or close to target with the optimal recipe.

The blue desirability setpoints can be moved by the user for tuning of the desirability function.



For more information about the optimizer objectives see the Optimizer section in the Statistical appendix.

## Factor spreadsheet

In the factor spreadsheet all factors are available with their current roles and their ranges according to the factor definition.

### Role of the factor

The **Role** is set to define whether the factor can vary or not during the optimization. If the factor can vary within an interval, you should set the role *Free*, if the factor should not vary you should set the role *Constant*. By default, all factors included in the model are *Free*.

When a factor is *Free* it is varied during the optimization within its specified range as defined by **Low limit** and **High limit**. By default, these low and high limits are taken from the factor specification when the design is generated. If there is a CCx design, the limits are the defined star points. You can change the range to widen or narrow the search region.

When a factor is *Constant*, it will be held at the selected constant value during the optimization. The default constant value is the center point. Change this value by typing another under **Value**.

The **Precision** is the disturbance added to the factor settings proposed by the optimizer in the run list. 0.15 means +/-0.15 as 95% confidence interval. It is used to calculate the probability of getting a prediction outside the response specifications taking the given precision into account. The result is shown in the run list as Probability of failure, that is, probability that points will fall outside the accepted response area.

If no precision is given at the initial factor specification for the design setup a default precision of  $\pm 5\%$  of the factor range will be added for the sensitivity analysis.

The column **Graph** presents the factor range used in the optimization. By default it will be the factor range used in the design framed by the red vertical lines and the colored bar is the used range relative the design specification.

	Factor	Role	Value	Low limit	High limit	Precision	Graph
1	Air	Free		240	284	4	
2	EGR%	Free		6	12	0.4	
3	Needlelift	Free		-5.78	0	0.4	
		Constant					

## CCC constraints

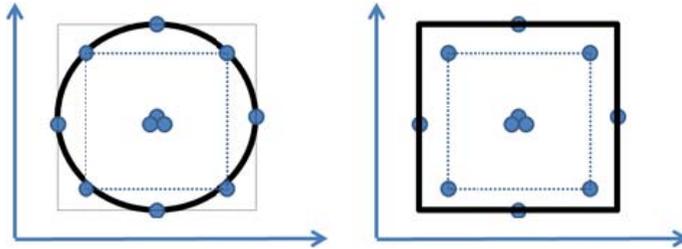
With CCC and CCO designs the optimizer search area is the CCC/CCO spherical space. The star point settings are used as optimizer low and high limits and the CCC constraints define the spherical space cutting off the corners (left graph).

### CCC constraints not used

If the defined optimizer low and high limits are not *exactly* the star point settings, or there are transformed factors in the design, the CCC constraints are not used and the optimizer search area becomes a hypercube. The hypercube is then defined by the optimizer low and high limits. This means that the corners of the search area will be outside the investigated region (right graph). In this case, it is preferable to enter the factor definition low and high settings in the optimizer objective.

To reset to default optimizer low and high limits (star distances at design creation), on the **Objective** tab of the **Optimizer**, right-click and click **Reset factor settings**.

Graphs below: The left graph illustrates the optimizer search area when the CCC constraints are used. The graph to the right illustrates the optimizer search area when the CCC constraints are not used due to that one of optimizer low/high settings was rounded.



### Setpoint tab

The **Setpoint** tab in the **Optimizer** shows the settings of the selected setpoint. You can change the selected setpoint by clicking a different point in the alternative setpoints list, either on the **Alternative setpoints** tab or in the list to the left of the response and factor spreadsheets on the **Setpoint** tab.

**Setpoint**

Selected setpoint: #13  
 The run with lowest log(D) is selected.

Alternative setpoints:

#	log(D)	Prob. of ...
9	-0.383	3%
10	-0.258	10%
11	-0.384	2%
12	-0.258	13%
13	-0.384	1.7%
14	-0.258	11%

Select best run  
 Find robust setpoint  
 Optimizer settings

Objective	Response	Criterion	Value	Graph	log(D)	Prob. of failure	Cpk
1	Fuel	Minimize	234.519		-0.490064	0%	1.59442
2	NOx	Minimize	17.8029		-0.614429	0%	2.84569
3	Soot	Minimize	0.486195		-0.172591	1.7%	0.891786

Factor	Role	Value	Graph	Factor contribution	
1	Air	Free	240.002		38.2935
2	EGR%	Free	8.77252		36.5484
3	Needlelift	Free	-4.4801		25.1581

In the **Setpoint** pane, the alternative setpoints are displayed and by default the run with the lowest log(D) is selected. The *D* in Log(*D*) is a normalized distance to the target.

In **Select best run** you can make the automatic selection on the lowest probability of failure instead. Note that if alternative solutions with log(D) close to the best are found, another run than the marked one may be optimal when looking at practical aspects for the factor settings or probability of failure.

Minimum for Log(*D*) = -10 (on target). A Log(*D*) ≤ 0 means that all results are within specification limits or very close.

**Probability of failure** gives additional information about how robust the proposed run will be to disturbances in the factor settings.

### Response prediction with selected setpoint

The result of the optimization is presented with the predicted result from the selected run, the **Value** column. The **Graph** presentation is the same as in the **Objective** tab complemented with black dots presenting the predicted result. Additional information such as **log(D)**, **Probability of failure** and **Cpk** (capability estimate) is calculated for all responses. More information is available in the Statistical appendix.

### Factor setpoint

The setpoint for the factors is presented with the selected settings in numerical (Value) and graphical presentation (Graph). The **Graph** presentation is the same as in **Objective** complemented with black dots presenting the factor values selected for prediction. **Factor contribution** is a normalized ranking of the factor influence around the selected setpoint with range 0-100. Moving the factors +/- 5% of the factor range around the setpoint will give a relative influence on all results. The factor with the largest contribution has the biggest impact in moving the predictions. For additional information see the Statistical appendix.

### Find robust setpoint

After running the **Optimizer** you can search for the robust setpoint by clicking **Find robust setpoint**, positioned under **Select best run**.

### Optimizer settings

At the bottom of the **Setpoint** pane, the Optimizer Properties dialog box can be opened by clicking **Optimizer settings**.

### Alternative setpoints tab

The **Alternative setpoints** tab in the **Optimizer** shows all calculated setpoints.

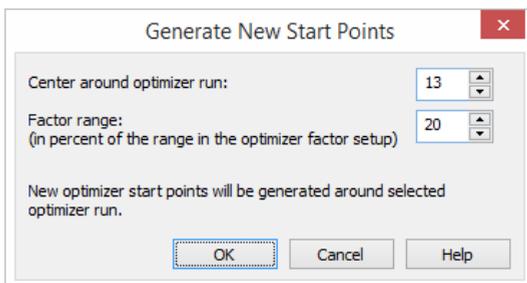
The optimization search procedure starts from various points in the investigated design region. The setup for the start is a design itself. From those settings a simplex algorithm will search for a solution that gives the lowest log(D) over all responses. Starting the search procedure at several positions minimizes the probability to get trapped in a non optimal solution; note the variation in the resulting log(D).

	Objective	Setpoint (#13)	Alternative setpoints								
	1	2	3	4	5	6	7	8	9	10	11
	Air	EGR%	Needlelift	Fuel	NOx	Soot	Iterations	log(D)	Prob. of failure	Cpk (Fuel)	Cpk (NO)
7	240.004	8.54796	-4.55291	234.569	17.5593	0.491333	132	-0.353051	1.94	1.54955	2.195
8	283.985	8.47447	-1.203	238.972	19.3016	0.479721	115	-0.258424	124	1.3652	2.292
9	240.008	8.73562	-3.99504	232.421	18.1077	0.535261	66	-0.352708	34	1.81154	2.474
10	283.994	9.40732	-1.23322	239.091	19.4476	0.688899	91	-0.258037	104	1.34158	2.168
11	240.002	8.87056	-4.52436	234.777	17.5441	0.498541	91	-0.353884	24	1.57094	2.942
12	283.998	9.62881	-1.11097	238.508	19.5066	0.682545	90	-0.257997	134	1.43407	2.111
13	240.002	8.77282	-4.4801	234.519	17.8029	0.486195	64	-0.354134	1.74	1.59442	2.848
14	283.856	9.65137	-1.21262	239.019	19.3538	0.673908	232	-0.257807	114	1.37296	2.232
15	240.004	8.54796	-4.55291	234.569	17.5593	0.491333	132	-0.353051	1.94	1.57652	2.908

Minimum for Log(D) = -10 (on target). A Log(D) <=0 means that all results are within specification limits or very close.

**Probability of failure** gives additional information about how robust the proposed run will be to disturbances in the factor settings. In the example above, alternatives to the lowest log(D) give more robust solutions.

After convergence, you can always click **Run optimizer** (play) to restart the optimizer. It restarts from the displayed resulting runs of the previous search. If you do not want to continue with the resulting runs click **New start points** or **New from selected** (dialog box below displayed) to specify other start runs.

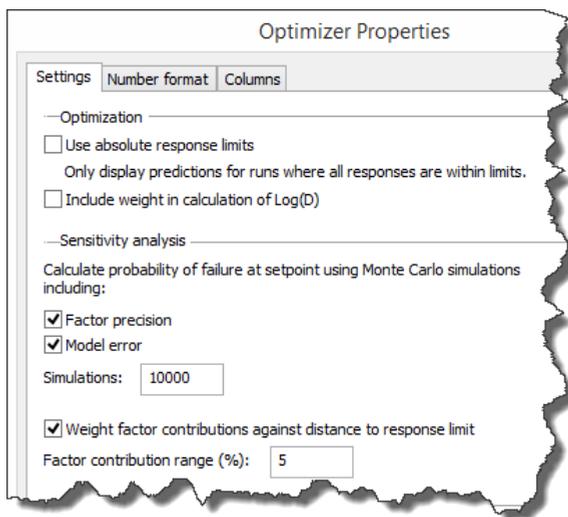


In the dialog box displayed, the run you marked is the default run to optimize around. To switch to another run, enter another run in the **Center around optimizer run** box.

Enter the percent of the factor range in the **Factor range** box. The percentage entered here is used to calculate the new high and low limits for the start runs.

### Optimizer Properties

In the **Optimizer** window, click **Optimizer settings** to open the **Optimizer Properties** dialog box.



#### Settings tab

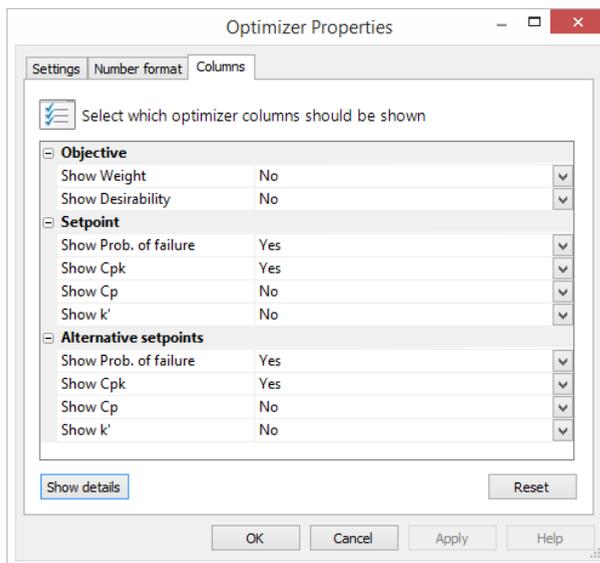
Feature	Description	Default
Use absolute response limits	When selected, only the runs where all responses are predicted inside the specified limits are displayed in the list.	Not selected.
Include weight in calculation of Log(D)	When selected, the weight specified in the response spreadsheet is included when calculating Log(D).	Selected.

Feature	Description	Default
Factor precision	When selected, the DPMO is calculated for runs where all responses are predicted inside the specified limits. The <b>Precision</b> value for the factors specifies the range used in the DPMO calculations. This is a sensitivity analysis that indicates if a solution is sensitive or insensitive to small changes in the factor settings.	Selected.
Model error	When selected, the probability of failure analysis includes the model error.	Selected.
Simulations	The number of simulations in the calculation of the Probability of failure. The result is scaled up to the unit 'dots per million operations' when using DPMO.	1%
Weight factor contribution	Will give a higher contribution weight to a factor that has a setpoint value which predicts a response close to a defined limit by dividing the factor contribution with the distance to the limit.	Not selected.
Factor contribution range	The part of the factor range used as variation around the setpoint to calculate factor contribution.	5%

*Note: If the Factor precision and Model error check boxes are both cleared, no simulations are done. The results then lack probability estimation.*

### Optimizer Properties, Columns tab

Presents additional columns in the **Objective response** settings section for the different pages in the Optimizer.



### Objective page

In the **Weight** column you can enter a number between 0.1 and 1 reflecting the importance of the response. Default is 1 indicating that all responses are of equal importance. Individual weights for the responses results in an optimization where the responses with higher weights are favored in the quest for all responses to reach inside the limits.

---

*Note: For the weight change to have the expected result, the selected desirability needs to be **Target**.*

---

To control the optimization, the overall desirability function,  $f(ds)$ , plays a key role, as well as reasonable limits and targets for the responses. The optimizer will strive to reach the lowest possible value of the overall desirability function,  $f(ds)$ , and will strive to reach the lowest possible value. The shape of the function is controlled by the settings of the criteria (Min, Target, Max) for each response and the choice of the  $d_k$  functions.

With weight 1, the lowest possible value of the individual desirability functions,  $d_k$ , is -100. Note that in the plots, the individual desirability function,  $d_k$ , is translated to  $[-1, 0]$ , so that a weight of 1 has a lowest possible value in the plot of -1, and so on. The goal is to reach a minimum in the overall desirability function,  $f(ds)$ .

For information about the **Desirability** column, see the Optimizer desirability section earlier in this chapter.

### Setpoint page

The Probability of failure, Cpk, Cp and k' columns that can be selected for the page all display calculated values that you may want to display.

### Alternative setpoints page

The Probability of failure, Cpk, Cp and k' columns that can be selected for the page all display calculated values that you may want to display.

## Optimizer list

The Optimizer list shows the setup used by the optimizer and the final results including robust setpoint and hypercube information when available.

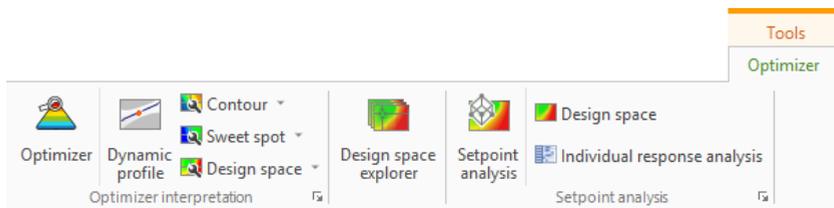
1	2	3	4	5	6	7	8	9	10	
1	Setup									
2	Factor	Role	Value	Low limit	High limit	Precision	Unit			
3	Air	Free		240	284	4	kg/h			
4	EGR%	Free		6	12	0.4	%			
5	Needlelift	Free		-5.78	0	0.4	BTDC			
6										
7	Response	Criterion	Weight	Min	Target	Max	Pred. min	Pred. max	Unit	Desirability
8	Fuel	Minimize	1		214.097	250	214.785	262.105	mg/st	Limit
9	NOx	Minimize	1		10.8065	25	11.2755	32.4767	mg/s	Limit
10	Soot	Minimize	1		0.0182831	1	0.0650316	3.92071	mg/s	Limit
11										
12	Selected setpoint									
13	Factor	Role	Value	Unit	Factor contribution					
14	Air	Free	245.867	kg/h	35.2164					
15	EGR%	Free	8	%	38.4595					
16	Needlelift	Free	-4.624	BTDC	26.3241					
17										
18	Response	Criterion	Value	Unit	log(D)	Prob. of failure	Cpk	Cp	k'	
19	Fuel	Minimize	239.518	mg/st	-0.299888	0.05%	1.1685			
20	NOx	Minimize	20.2253	mg/s	-0.356191	0%	1.94035			
21	Soot	Minimize	0.32408	mg/s	-0.287226	0.12%	1.81719			
22										
23	Robust setpoint (R)									
24	Factor	Role	Value	Robust low edge	Robust high edge	Unit	Factor contri	Robust resol	Hypercube k	Hypercube h
25	Air	Free	245.867	240.367	251.367	kg/h	35.2164	2 <-> 2	240	248.8
26	EGR%	Free	8	7.25	8.75	%	38.4595	2 <-> 2	7.2	8.4
27	Needlelift	Free	-4.624	-5.3465	-3.54025	BTDC	26.3241	2 <-> 3	-5.00933	-4.23867
28										
29	Response	Criterion	Value	Unit	log(D)	Prob. of failure	Cpk	Cp	k'	
30	Fuel	Minimize	239.518	mg/st	-0.313415		1.1685		0.706392	
31	NOx	Minimize	20.2253	mg/s	-0.313415		1.94035		0.664071	
32	Soot	Minimize	0.32408	mg/s	-0.313415		1.81719		0.332081	
33										
34	Design space summary									
35	Resolution	Iterations	DS count	Relative volume %						
36		16	50000	231	5.63965					

To open the Optimizer list, with the **Optimizer** as the active window, on the **Tools** tab, click **List**.

*Note: The Design space summary details are not saved with the investigation. This means that if you close and reopen the investigation, the Design space summary section will be empty.*

## Optimizer tab

After clicking **Optimizer** on the **Predict** tab or the **Home** tab, the **Optimizer** contextual tab becomes available.



The Optimizer tab contains functions that are related to the Optimizer. These functions, described in this section, are,

- Optimizer - to reactivate the Optimizer window,
- Dynamic profile,
- Contour,
- Sweet spot,

- Design space,
- Design space explorer - Opens the **Design Space Explorer** dialog box if Find robust setpoint has not been run, otherwise the plot,
- Setpoint analysis,
- Design space in the Setpoint group, and
- Individual response analysis in the Setpoint group.

The Optimizer List is opened by clicking **List** on the **Tools** tab.

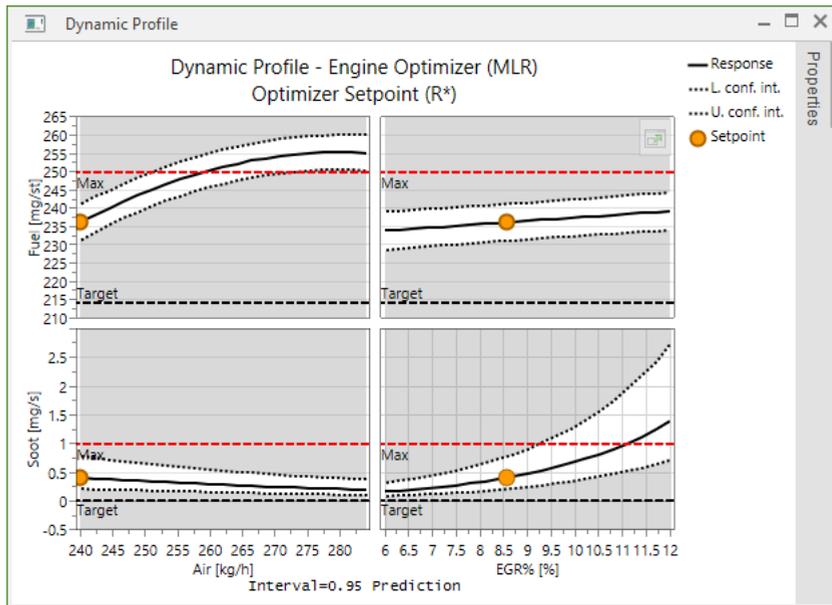
---

*Note: The Optimizer tab is available only after the optimizer has been run.*

---

## Dynamic profile

To access the **Dynamic profile** window, on the **Optimizer** contextual tab, click **Dynamic profile**. It will show the effect of the factor over the investigated range given that all the other factors are at a specific setpoint. The default for this setpoint is the one calculated by the optimizer.



The **Properties** section of the window to the right allows for graphically and numerically changing the settings of the factors. The effect curve of one factors vs one response depends on the settings on the other factors. Sliding to different factor settings will give a graphical view of the alternative solutions and update the optimizer window.

## Creating contour plots from the optimizer

After convergence of the optimizer you can create contour plots using the results.

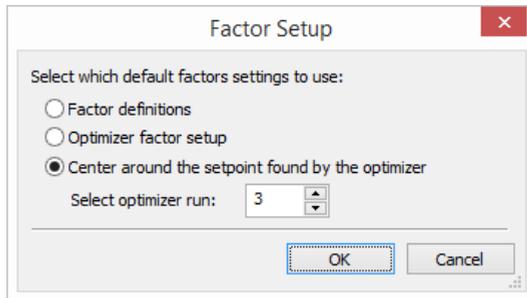
On the **Optimizer** contextual tab, click **Contour** and click one of the plots in the gallery. The displayed plot is centered around the selected setpoint.

In the created plot the predicted parts that are outside the original factor settings are displayed shaded when the plot is extrapolated.

The selected run is displayed in the plot as arrows originating from the axes pointing toward the position of the selected run.

## Contour plot wizard from the optimizer

After convergence of the optimizer you can create contour plots using the results. On the **Optimizer** contextual tab, click **Contour** to open the **Factor Setup** dialog box.



### Select factor settings

In the **Factor Setup** dialog:

- **Center around the setpoint found by the optimizer** is by default marked, and the run to use in the center is the one you marked. If you want to select another run enter the new run number.
- Selecting **Factor definitions** results in displaying same factor settings as when you enter the contour plot/sweet spot/design space wizard from the **Predict** tab
- Selecting **Optimizer factor setup** results in displaying the current factor settings in the **Optimizer**.

### Center around the setpoint found by the optimizer

With **Center around the setpoint ...** selected, clicking **OK** opens the plot wizard dialog box.

If you create a 2D plot the first two factors are varied around their setpoint settings (20% range) and the other factors are set constant to their setpoint values.

If you create a 4D plot the first two factors are varied around their setpoint settings (20% range), the 3<sup>rd</sup> and 4<sup>th</sup> are set at their low, high, and center values using the 20% setpoint range, and other factors are set constant to their setpoint values.

The resulting plot displays the area outside the design factor range shaded.

The selected run is displayed in the plot as arrows originating from the axes pointing toward the position of the selected run.

This option is unavailable for mixture (formulation) factors.

## Creating a sweet spot plot from the optimizer

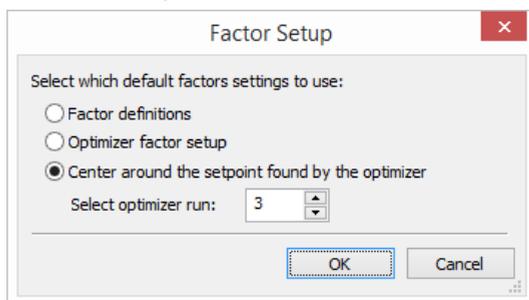
When you choose to create the sweet spot plot from the optimizer, MODDE uses the acceptance range of the responses as specified in the optimizer response spreadsheet fields **Min**, **Target**, and **Max**.

To create the sweet spot plot click **Sweet spot** on the **Optimizer** contextual tab, and click one of the plots in the gallery. The displayed plot is centered around the selected setpoint.

In the created plot the predicted parts that are outside the original factor settings are displayed shaded when the plot is extrapolated. The selected run is displayed in the plot as arrows originating from the axes pointing toward the position of the selected run.

## Sweet spot plot wizard from the optimizer

When you choose to create the sweet spot plot from the optimizer, MODDE uses the acceptance range of the responses as specified in the optimizer response spreadsheet fields **Min**, **Target**, and **Max**.



### Select factor settings

In the **Factor Setup** dialog:

- **Center around the setpoint found by the optimizer** is by default marked, and the run to use in the center is the one you marked. If you want to select another run enter the new run number.
- Selecting **Factor definitions** results in displaying same factor settings as when you enter the contour plot/sweet spot/design space wizard from the **Predict** tab
- Selecting **Optimizer factor setup** results in displaying the current factor settings in the **Optimizer**.

### Center around the setpoint found by the optimizer

With **Center around the setpoint ...** selected, clicking **OK** opens the plot wizard dialog box.

If you create a 2D plot the first two factors are varied around their setpoint settings (20% range) and the other factors are set constant to their setpoint values.

If you create a 4D plot the first two factors are varied around their setpoint settings (20% range), the 3<sup>rd</sup> and 4<sup>th</sup> are set at their low, high, and center values using the 20% setpoint range, and other factors are set constant to their setpoint values.

The resulting plot displays the area outside the design factor range shaded.

The selected run is displayed in the plot as arrows originating from the axes pointing toward the position of the selected run.

This option is unavailable for mixture (formulation) factors.

## Creating a design space plot from the optimizer

After convergence of the optimizer you can create design space plots using the results. The **Design Space Plot** is based on the result of Monte Carlo simulations done on the models for the responses that are active in the optimizer. The important properties that separate the **Design Space Plot** from a Sweet spot plot can be found in the **Design Space Plot Properties** described later in this section.

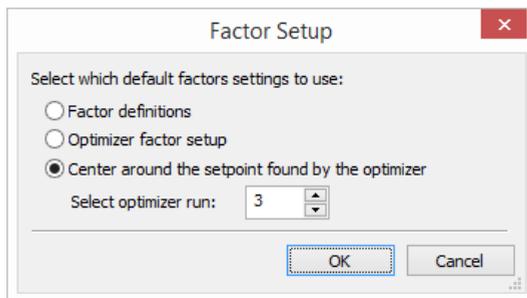
On the **Optimizer** contextual tab, click **Design space** and click one of the plots in the gallery. The displayed plot is centered around the selected setpoint.

In the created plot the predicted parts that are outside the original factor settings are displayed shaded when the plot is extrapolated.

The selected run is displayed in the plot as arrows originating from the axes pointing toward the position of the selected run.

## Design space wizard from the optimizer

When you choose to create the design space plot from the **Optimizer** contextual tab, MODDE uses the acceptance range of the responses as specified in the optimizer response spreadsheet fields **Min**, **Target**, and **Max**.



### Select factor settings

In the **Factor Setup** dialog:

- **Center around the setpoint found by the optimizer** is by default marked, and the run to use in the center is the one you marked. If you want to select another run enter the new run number.
- Selecting **Factor definitions** results in displaying same factor settings as when you enter the contour plot/sweet spot/design space wizard from the **Predict** tab
- Selecting **Optimizer factor setup** results in displaying the current factor settings in the **Optimizer**.

### Center around the setpoint found by the optimizer

With **Center around the setpoint ...** selected, clicking **OK** opens the plot wizard dialog box.

If you create a 2D plot the first two factors are varied around their setpoint settings (20% range) and the other factors are set constant to their setpoint values.

If you create a 4D plot the first two factors are varied around their setpoint settings (20% range), the 3<sup>rd</sup> and 4<sup>th</sup> are set at their low, high, and center values using the 20% setpoint range, and other factors are set constant to their setpoint values.

The resulting plot displays the area outside the design factor range shaded.

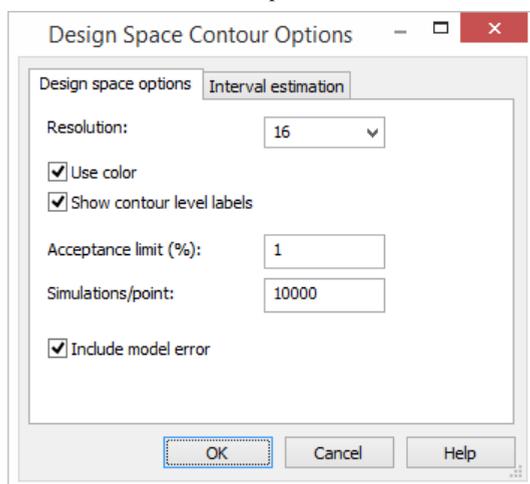
The selected run is displayed in the plot as arrows originating from the axes pointing toward the position of the selected run.

This option is unavailable for mixture (formulation) factors.

### ***Design space plot options***

Click **Plot options...** on the first page in the wizard to open the plot options, alternatively right-click the created plot and click **Properties**. The **Design Space Contour Options** lets you specify,

- **Resolution** - A higher resolution means a more detailed plot, but it can also slow down your computer.
- **Use color** - This determines if the plot should be colored or merely display the contour lines.
- **Show contour level labels** - Select to display contour level labels.
- **Acceptance limit** - The acceptance criteria aimed for either in percentage or in the DPMO space.
- **Simulations/point** - The number of simulations per point; the higher the more time consuming but also more accurate.
- **Include model error** - This determines if model error should be included in the plot. Note that when this check box is cleared, the plot merely presents the same information as the Sweet spot plot.
- Interval type, Confidence level and Tolerance proportion on the **Interval estimation** tab. For more, see the Interval type and probability levels subsection in chapter 4, File.



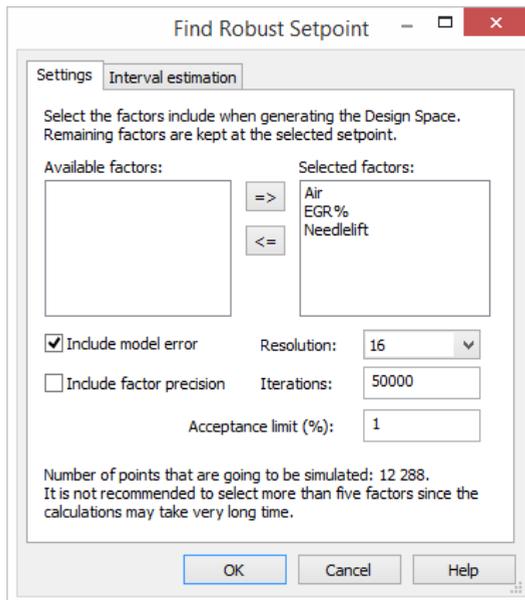
## Find robust setpoint

After running the **Optimizer** you can search for the robust setpoint. The robust setpoint function will maximize the distance from the acceptance boundaries in the Design Space. The procedure will be to first generate the design space with selected factors and a given specification. It can be very time-consuming, depending on the settings chosen. The resolution must be considered when you search for the robust setpoint; it is very easy to get an overly large data set that takes a long time to calculate. Note that the dialog box tells you how many points are going to be simulated based on the number of factors selected and your resolution. The greater the resolution, the tighter the grid of points that are calculated.

The number of points and calculations:

Number of points to be simulated = number of responses \* (resolution<sup>(number of selected factors)</sup>).

Number of calculations = number of responses \* (resolution<sup>(number of selected factors)</sup>) \* Iterations.

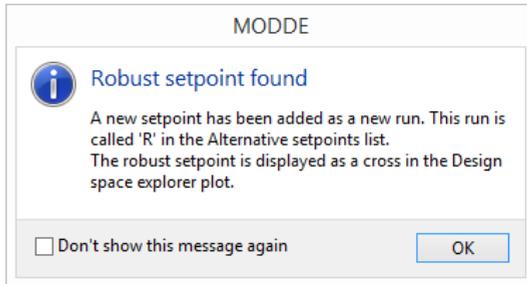


To search for the robust setpoint;

1. In the **Optimizer** window, click **Find robust setpoint** or on the **Optimizer** contextual tab, click **Design space explorer**. When you click **Design space explorer** the plot automatically opens when the calculations are complete.
2. Select which factors to use, the resolution, iterations, Acceptance limit, and if model error and factor precision should be included.
3. Optionally change Interval estimation settings on the **Interval estimation** tab.
4. When you are happy with your settings and the number of points to be simulated, click **OK**.

### Robust setpoint results

If a robust setpoint was found, the result is added to the optimizer **Alternative setpoints** list as a run with the notation *R*. The factor settings are shown in the **Setpoint** tab and the corresponding result is shown in the response spreadsheet. This point is also displayed as a cross with arrows pointing to the robust setpoint, in the **Design Space Explorer** plot automatically opened after completing the calculations.



Detailed information is presented in the **Optimizer List**.

12	Selected setpoint									
13	Factor	Role	Value	Unit	Factor contribution					
14	Air	Free	245.677	kg/h	35.2164					
15	EGR%	Free	7.93548	%	38.4595					
16	Needlelift	Free	-4.66129	BTDC	26.3241					
17										
18	Response	Criterion	Value	Unit	log(D)	Prob. of failure	Cpk	Cp	k'	
19	Fuel	Minimize	239.454	mg/st	-0.302063	0.06%	1.16714			
20	NOx	Minimize	20.3686	mg/s	-0.343069	0%	1.87974			
21	Soot	Minimize	0.315024	mg/s	-0.295832	0.12%	1.94981			
22										
23	Robust setpoint (R)									
24	Factor	Role	Value	Robust low edge	Robust high edge	Unit	Factor contri	Robust resol	Hypercube k	Hypercube
25	Air	Free	245.677	240.177	251.177	kg/h	35.2164	4 <-> 4	240	248.516
26	EGR%	Free	7.93548	6.99798	8.87298	%	38.4595	5 <-> 5	7.16129	8.32258
27	Needlelift	Free	-4.66129	-5.74504	-3.21629	BTDC	26.3241	6 <-> 8	-5.03419	-4.10194
28										
29	Response	Criterion	Value	Unit	log(D)	Prob. of failure	Cpk	Cp	k'	
30	Fuel	Minimize	239.454	mg/st	-0.313155		1.16714		0.704912	
31	NOx	Minimize	20.3686	mg/s	-0.313155		1.87974		0.673512	
32	Soot	Minimize	0.315024	mg/s	-0.313155		1.94981		0.318876	
33										
34	Design space summary									
35	Resolution	Iterations	DS count	Relative volume %						
36	32	50000	1875	5.72205						

**Robust low/high edge:** The setting of the factor that reaches the low/high boundary of the Design Space as a Manhattan distance (perpendicular distance in the factor space) from the robust point to the Design Space limit.

**Robust resolution distance:** The calculations of the robust point are done with a specific resolution. The resolution limits the position of the robust setpoint. If the robust distance is less than 4 the recommendation is to increase the resolution for the robust setpoint estimate. The Robust low edge and Robust high edge are calculated from those resolution values and define the robust range.

**Hypercube low edge and Hypercube high edge:** The factor settings that define the design space hypercube.

**Relative volume:** The relative volume in Design Space calculation is the resolution<sup>number of factors</sup>, in this case  $32^3 = 32768$ . The part of this volume that is considered as a *DS count* = 1875; the DS count corresponding to Relative volume = 5.72%.

## Design space explorer

The **Design Space Explorer** allows you to investigate the Design Space volume by sliding the constant factors to alternative settings.

To display the Design Space Explorer plot the calculations, specified in the **Find Robust Setpoint** dialog, have to be performed. This dialog opens when clicking **Design space explorer** on the **Optimizer** tab or **Find robust setpoint** in the **Optimizer** window **Setpoint** tab.

If a robust setpoint is found, that point is displayed in the Design Space Explorer plot as arrows forming a cross pointing to it. The setting for the factors held constant is also marked with an arrow along the slider in the Properties pane.

The dotted frame in the Design Space Explorer plot designates the largest possible *regular* hypercube that can be inserted into the *irregular* design space volume. How this regular hypercube extends into many dimensions is given by the green bars seen in the Hypercube range part in the picture above. The green bars mark the *mutual ranges* within which all factors can be changed at the same time and without further restrictions.

---

**Note:** When sliding a constant factor, the arrows at the end of the lines disappear but the lines are still displayed to show where the setpoint was. The arrows reappear when all constants are at their respective setpoint again.

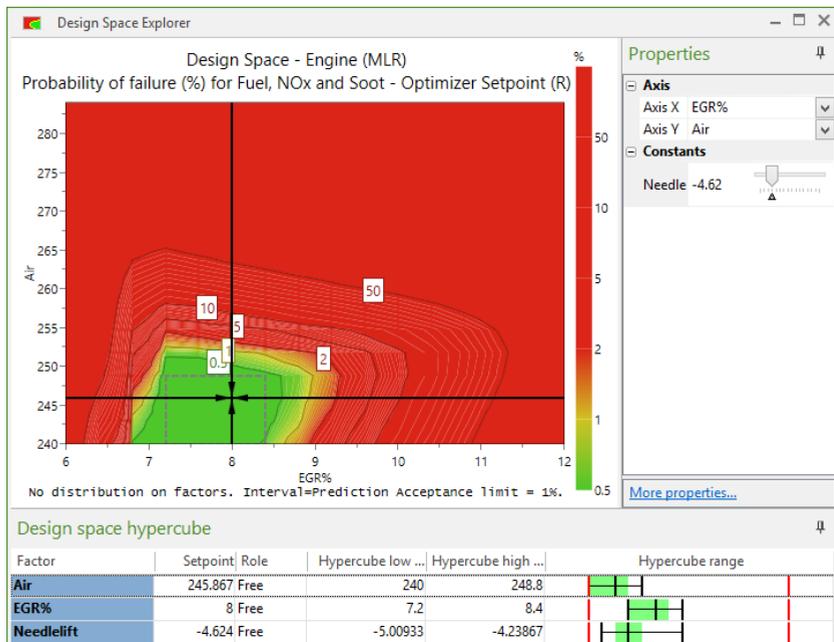
---



---

**Hint:** You can move the current setpoint by right-clicking the new position in the plot and clicking **Set as selected setpoint**. This new point can then be evaluated using **Setpoint analysis**. Note here that the original robust setpoint cannot be displayed after selecting a new setpoint in the plot without full recalculation.

---



To save the calculated results from search for the robust setpoint;

1. Create the **Design Space Explorer** plot.
2. On the **Tools** tab, in the **Create** group, click **List** and then save that list.

Displayed in the list are the coordinates for the factors and the overall probability of failure for the included responses.

## Setpoint analysis

Using setpoint analysis we can show how the factor settings can be varied around a selected setpoint and still fulfill the response criteria that correspond to the factor tolerance that can be used at the setpoint.

Limitations with a sweet spot or design space plot presentation are the number of dimensions and the lack of probability estimate in the predicted surface (Sweet spot plot only). Setpoint analysis uses the setpoint that was selected in the Optimizer (either the optimal point or the robust point). From there it increases the variation for the respective factors until it has reached the limit of predicted distribution for the responses. These calculations have to consider both the definition limits for the responses as well as the Acceptance limit.

The estimation is performed using Monte Carlo simulations on the factor settings. MODDE performs a search to identify the largest possible tolerance range for each factor that can be used and still meet all response requirements. The default limit is less than 1% hits outside the limits for each response. The results are displayed in the **Setpoint Analysis** window.

To open the **Setpoint Analysis** window,

1. On the **Predict** tab or **Home** tab, click **Optimizer**.
2. Run the Optimizer.
3. Then on the **Optimizer** contextual tab, click **Setpoint analysis**.

For more information about the **Setpoint Analysis** window, see Chapter 13, Setpoint.

# 13-Setpoint

---

## Introduction

Using setpoint analysis we can show how the factor settings can be varied around a selected setpoint and still fulfill the response criteria that correspond to the factor tolerance that can be used at the setpoint.

Limitations with a sweet spot or design space plot presentation are the number of dimensions and the lack of probability estimate in the predicted surface (Sweet spot plot only). Setpoint analysis uses the setpoint that was selected in the Optimizer (either the optimal point or the robust point). From there it increases the variation for the respective factors until it has reached the limit of predicted distribution for the responses. These calculations have to consider both the definition limits for the responses as well as the Acceptance limit.

The estimation is performed using Monte Carlo simulations on the factor settings. MODDE performs a search to identify the largest possible tolerance range for each factor that can be used and still meet all response requirements. The default limit is less than 1% hits outside the limits for each response. The results are displayed in the **Setpoint Analysis** window.

---

## Setpoint analysis plots and lists

To open the **Setpoint Analysis** window,

1. On the **Predict** tab or **Home** tab, click **Optimizer**.
2. Run the Optimizer.
3. Then on the **Optimizer** contextual tab, click **Setpoint analysis**.

After opening the Setpoint analysis window, the following Setpoint analysis features are available;

- Factor distribution
- Response distribution
- Individual response analysis
- Design space contour plot
- Setpoint analysis list.

## Setpoint window

The **Setpoint Validation** and **Setpoint Analysis** windows function in the same way. The windows consist of a factor and a response spreadsheet.

The **Setpoint Analysis** window appears when clicking **Setpoint analysis** on the **Optimizer** contextual tab.

The **Setpoint Validation** window appears when clicking **Setpoint validation** on the **Predict** tab. In the validation case, the preset conditions used are the factor average and the full range as input for the prediction.

For *factors*, the yellow Setpoint lines and the blue distribution range can be changed by using the mouse pointer to click and drag or by typing new values in the fields. The red lines are the normalized design specification lines.

For *responses*, the yellow Target lines and the red Max/Min lines can be moved either by typing a new value in the appropriate fields or by using the mouse pointer to click and drag the lines to the desired levels.

Factor	Low Setpoint	High	Std. d...	Role	Distri...	Estimated acceptable range
Air	234.706	240	245.294	2.7009	Free	Normal
EGR%	7.73143	8.57143	9.41143	0.428571	Free	Normal
Needlelift	-5.79329	-4.95429	-4.115...	0.428062	Free	Normal

Response	Min	Target	Max	Criterion	Prob. of failure	Predicted response profile
Fuel		214.097	250	Minimize	0.04%	
NOx		10.8065	25	Minimize	0%	
Soot		0.01828...	1	Minimize	1%	

Total Prob. of failure: 1.1% (limit: 1%). Samples: 50000. Interval=Prediction

Every change in settings is followed by a recalculation by default. This might be disturbing when tuning several settings, so it is possible to turn off automatic recalculation preventing the automatic update.

### Preventing auto update

To prevent the **Setpoint** window from automatically updating on every change,

1. Right-click the **Setpoint** window.
2. Click **Properties**.
3. Clear the **Automatic update** check box.

With automatic updating turned off, you will need to click **Click to resample** under **Predicted response profile**.

### Adding more columns

You can add more columns to the **Factor** and **Response** spreadsheets in the **Setpoint** window;

1. Right-click the header row in the Response/Factor spreadsheet.
2. Select and clear as desired.
3. To restore, click **Restore default layout**.

## Setpoint properties

Right-click the **Setpoint validation** or **Setpoint analysis** window to open the **Properties** dialog box.

### Parameters section

In the **Parameters** section, the settings used in the setpoint analysis/validation are displayed and can be adjusted.

Option	Description	Default
<b>Acceptable range simulations</b>	Number of simulations done in each step of the search for the acceptable range for the factors.	20 000
<b>Response profile simulations</b>	Number of simulations for the final predictions of the response profile.	50 000
<b>Acceptance limit</b>	Predictions outside the specification limits. The first response that exceeds this limit will stop the expansion of the accepted factor range.	1%
<b>Use total Probability of failure</b>	Stop criterion will be the total number of predictions outside the limit for all responses. When not selected the Acceptance limit is for each response.	Not selected.
<b>Include model error</b>	Includes model error in the predictions of the response distribution.	Selected.

Option	Description	Default
<b>Automatic update</b>	Any change in the settings results in a recalculation of the Setpoint analysis/validation. Clear this check box to do the calculations when all settings have been adjusted instead of after each adjustment.	Selected.
<b>Limits at 95 or 99 % confidence for normal distribution</b>	Defines the Low and High values in such a way that 95%, in the default case, of the Monte Carlo simulations are inside the limits. Selecting 99% widens the included area by lowering the Low limit and increasing the High limit in the factor spreadsheet.	95%

### Show and Color range sections

In the **Show** and **Color range** sections, to display limits and colors of the graphs can be specified.

Option	Description	Default
<b>Possible Min/Max lines</b>	Display the possible minimum and maximum for the factors, as listed on the Objective page in the Optimizer.	Not selected.
<b>Color range</b>	Displays the range of colors of the distribution bars.	Blue for factors, green for responses.

*Note: You can select which columns to display for the Factor and Response spreadsheets respectively by right-clicking the header row in the Setpoint window.*

### Factor spreadsheet

In the factor spreadsheet the following functionality is available for each factor;

Item	Description	Default
<b>Low, Setpoint, High.</b>	The acceptable factor range	The largest range that complies with the response criteria, estimated by MODDE.
<b>Std. dev.</b>	One standard deviation of the used factor range.	Calculated.
<b>Role</b>	<b>Free:</b> The largest possible range of variation for the specific factor is used, with respect to the settings for other factors and response specifications. <b>Locked:</b> The range settings of the factor are locked according to the set range.	For <b>Setpoint validation:</b> by default 'Locked'. For <b>Setpoint analysis:</b> by default 'Free'.

Item	Description	Default
<b>Distribution</b>	<p>The distribution to be used for each factor range.</p> <p><b>Uniform:</b> All factor settings within the specified range have the same probability to appear.</p> <p><b>Normal:</b> The simulations are normally distributed within the factor range.</p> <p><b>Triangular:</b> The distribution has the shape of a triangle; a good way to get a skewed distribution for a factor.</p> <p><b>Target:</b> The factor is set to a fixed value, the setpoint value.</p>	<p>For quantitative factors, by default 'Normal'.</p> <p>For formulation factors, by default 'Triangular'.</p>
<b>Possible min</b>	<p>The left side of the T bar in the <b>Estimated acceptable range</b> is the estimated minimum factor value for which the predictions still fall inside the specifications when all other factors are locked at the setpoint.</p>	<p>Calculated, by default not shown.</p>
<b>Possible max</b>	<p>The right side of the T bar in the <b>Estimated acceptable range</b> is the estimated maximum factor value for which the predictions still fall inside the specifications when all other factors are locked at the setpoint.</p>	<p>Calculated, by default not shown.</p>
<b>Estimated acceptable range</b>	<p>The <b>yellow lines</b> are the default factor settings of the selected run in the optimizer.</p> <p>The <b>red lines</b> are the Low and High factor settings of the experimental region.</p> <p>The <b>blue region</b> represents the 95%* part of random factor variability with normal distribution where all predictions are within the specifications.</p> <p>Not default displayed: The <b>black T bar</b> represents the region of acceptable variability valid for that factor when all other factors are locked at the setpoint. Valid means that all predictions of the responses are within the specifications. No model error is considered in this search.</p>	

\*95% is the default. In **Setpoint Properties** you can change it to 99%.

## Response spreadsheet

In the response spreadsheet the following functionality is available for each response:

Item	Description	Default
<b>Min, Target, Max</b>	The optimization range	Min, Target, and Max as specified in the Optimizer.
<b>Criterion</b>	What the algorithm is aiming for.	<i>Minimize, Maximize, Target, and Predicted</i> as specified in the optimizer. The criterion <i>Not met</i> signifies that the response was excluded from the calculations.
<b>Cpk</b>	A <i>Process Capability Index, Cpk</i> , which originates from the Six Sigma statistics, is estimated in this simulation. $Cpk = 1$ means that approximately 0.13%* of the predictions will fall outside the specifications.	Not displayed. Estimated by MODDE when selected.
<b>Cp</b>	The <i>process potential index, Cp</i> , estimates what a process is "capable" of producing if the target is centered between the specification limits under the "natural" tolerance $6\sigma$ . If $Cp = 1$ , the process is said to be "capable". Originates from the Six Sigma statistics.	Not displayed. Estimated by MODDE when selected.
<b>k'</b>	The <i>process deviation index, k</i> , estimates the amount of deviation of the process mean from the target. Originates from the Six Sigma statistics.	Not displayed. Estimated by MODDE when selected.
<b>Probability of failure</b>	Probability of failure is an estimate of how many predictions will be outside the specification in the setpoint analysis/validation with the selected distribution. Displayed in DPMO or percent.	Estimated by MODDE.
<b>Average</b>	The average of the predicted response values.	Not displayed. Calculated by MODDE when selected.
<b>Std. dev.</b>	The standard deviation of the predicted response values.	Not displayed. Calculated by MODDE when selected.
<b>% outside</b>	Percentage of predicted points found outside the Min and Max limits.	Not displayed. Estimated by MODDE when selected.
<b>Median</b>	The median of the predicted response values.	Not displayed. Calculated by MODDE when selected.

Item	Description	Default
<b>1st quartile</b>	25% of the predicted response values are found below this value.	Not displayed. Calculated by MODDE when selected.
<b>3rd quartile</b>	25% of the predicted response values are found above this value.	Not displayed. Calculated by MODDE when selected.
<b>Predicted response profile</b>	<p>The <b>yellow line</b> represents the Target value for the response as specified in the Optimizer.</p> <p>The <b>red lines</b> are the specification limits for each response as specified in Optimizer.</p> <p>The <b>faded green region</b> represents the predictions for a random distribution of factor settings in the given ranges (low-setpoint-high).</p>	

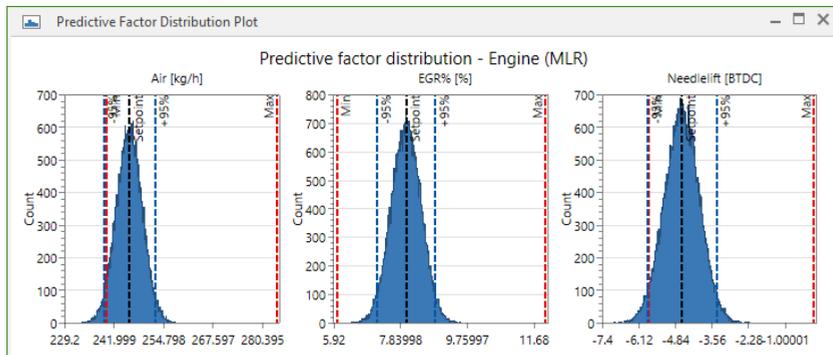
\* See the table in the Probability of failure and process capability indices subsection in the Statistical appendix.

## Factor distribution

Open the **Predictive Factor Distribution Plot** by clicking the histogram to the far right on each row. 

This histogram displays the distribution of the factor settings used for the simulation. The specified Min and Max values are displayed as lines.

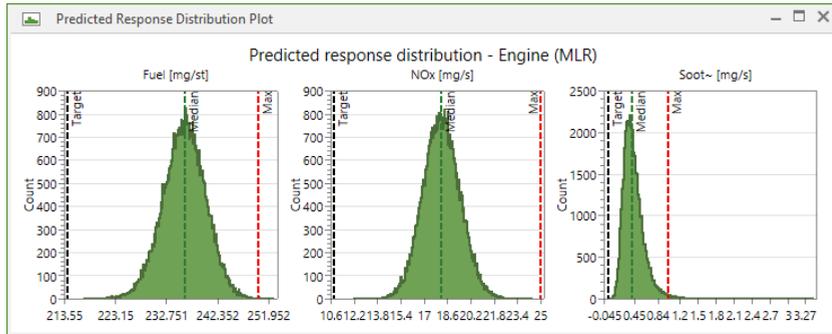
**Hint:** You can also display the Predictive Factor Distribution Plot by double-clicking Estimated acceptable range.



## Response distribution

To display the result of the Monte Carlo simulations as frequency histograms, on the **Setpoint analysis** or **Setpoint validation** window, click the histogram **Open predicted response distribution plot** to the far right. 

**Hint:** You can also display the Predicted Response Distribution Plot by double-clicking Predicted response profile.



## Setpoint list

The Setpoint Analysis and Setpoint Validation Lists show the information displayed in the Setpoint analysis or Setpoint validation window.

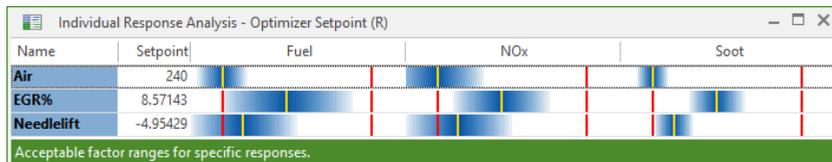
Factor	2	3	4	5	6	7	8	9	10	11	12	13
Air	234.706	240	245.294	Normal	231.2	258.7	240	284				
EGR%	7.73143	8.57143	9.41143	Normal	6.17143	10.9714	6	12				
Needlelift	-4.95429	-4.95429	-4.11529	Normal	-6.97729	-1.93279	-5.78	0				
Confidence for normal distribution:	95											
Response	Prob. of failure	Cpk estimate	Cp estimate	K estimate	% out of range	Average	Median	1st quartile	3rd quartile	Std. dev.	Skewness	DF
Fuel	0.04%	1.15524	0.41148	0.04	236.051	236.125	233.429	238.763	4.02476	-0.11725	11	
NOx	0%	1.77874	0.518235	0	18.1763	18.1615	17.3152	19.0205	1.27876	0.0706449	9	
Soot	1%	1.08176	0.424849	1.036	0.435365	0.404967	0.316688	0.51843	0.173986	1.65943	13	
Model error included in predictions:	Yes											
Number of samples:	50000											

To open the Setpoint analysis/validation list, with the **Setpoint analysis/validation** as the active window, on the **Tools** tab, click **List**.

## Individual response analysis

To open the **Individual Response Analysis** window, on the **Optimizer** contextual tab, click **Individual response analysis**.

The window shows the acceptable factor ranges for specific responses.



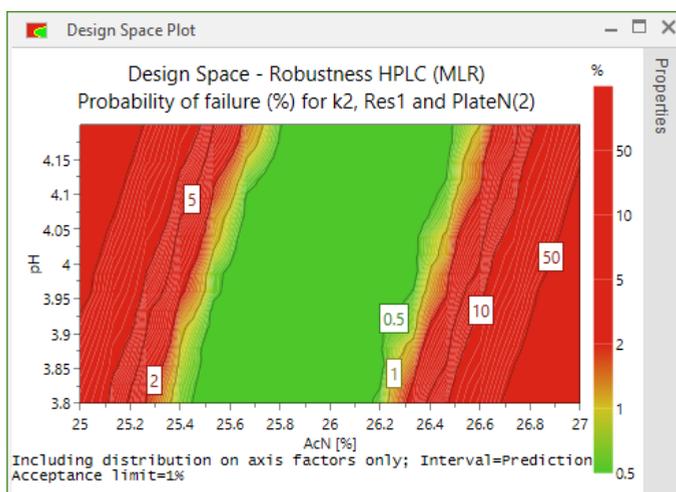
## Setpoint, design space

In the Setpoint analysis group on the **Optimizer** contextual tab, click **Design space** to open the **Design Space Plot** wizard.

This plot can be investigated from two angles:

1. Visualization of the 2D design space from any selected setpoint - here the **Include distribution from setpoint analysis on axis factors** check box should be cleared while the **Include distribution from setpoint analysis on constant factors** check box should be selected.
2. Analyze the size and shape of the design space at any given setpoint with a specific factor tolerance - here both check boxes should be selected.

Click **Finish** to see the **Design Space Plot** showing the probability of failure (%) for the shown factor combinations. Alternatively, **Next** to specify the settings for the constant factors.



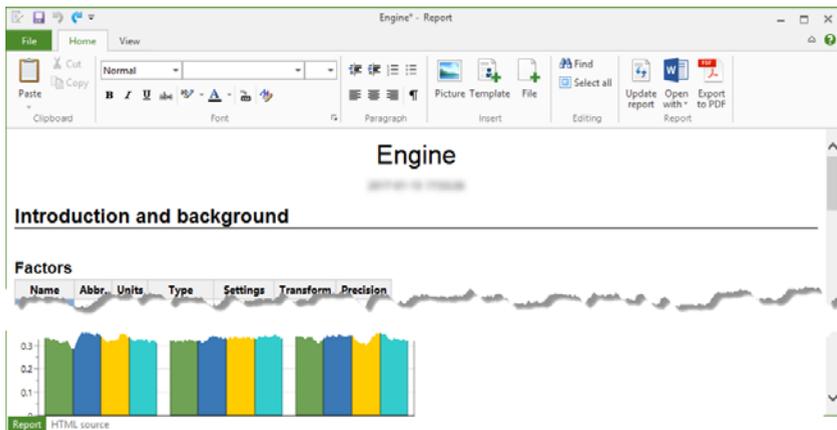


# 14-Report

---

## Introduction

MODDE has an automatic report generator. In the report, basic formatting functionality is available for writing text. Plots, lists, and model results of MODDE can be added to the report at any time. These items are added to the report as placeholders. A placeholder stands in the place of contents which MODDE will provide, let it be a plot, list or any text or number. The placeholders enable you to save the report as a template and use it for different investigations.



---

## Opening the Report

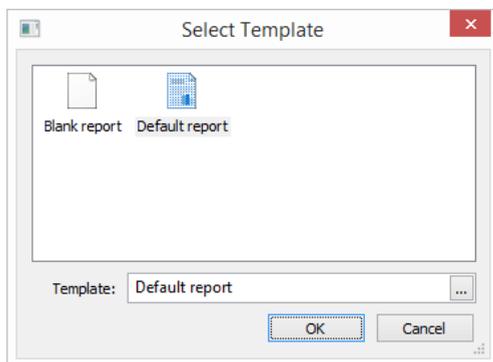
To create a new report, on the **File** tab, click **Info** and then click **Report**. A dialog box opens and you can select which template to open.

If you have saved a report with the investigation, MODDE automatically opens that report.

---

**Hint:** You can also open the report by pressing Ctrl+R, adding the current window to the saved/blank report.

---



When your investigation has not been fitted MODDE will ask to fit the investigation so it can automatically fill all the placeholders.

---

## Saving a report

To save a report created in this version of MODDE, save and close the investigation. The report is automatically saved with the .mip.

To save a report created in MODDE 10 or earlier, see the Opening and saving an old report subsection next.

To save the report as a separate file, on the **File** tab, click **Save as**, and then click **Save a copy**.

Note that this separate file is *NOT* saved with the investigation, but only as a separate copy of the report. However, if you then click **Save** in the report and also **Save** in the MODDE window, the current report is also saved with the .mip.

The report can also be exported to PDF. For details, see the Home tab topic in Chapter 14, Report.

---

## Opening and saving an old report

To open a report saved in MODDE 10 or earlier,

- On the **File** tab, click **Info** and then click **Report**. In the **Select Templates** dialog, browse for the old report.

Or,

- If a report was already saved with the investigation, in the report generator click **Open** on the **File** tab and browse for the old report.

To save the old report,

- With the investigation it was opened for, on the **File** tab, click **Save**.

Or,

- Separately to file, on the **File** tab, click **Save as** | **Save a copy**.

Note that this separate file is *NOT* saved with the investigation, but only as a separate copy of the report. However, if you then click **Save** in the report and also **Save** in the MODDE window, the current report is also saved with the .mip.

---

## Report window

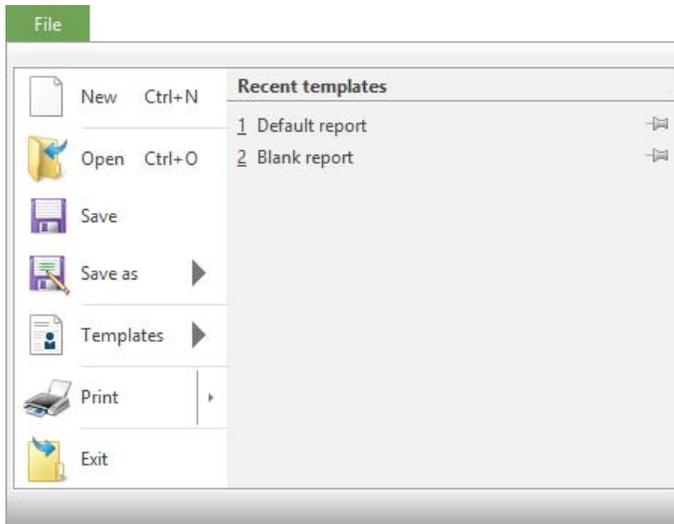
The **Report** window contains:

- **File** tab, containing overall commands such as **New** and **Open** but also **Template** commands.
- **Home** tab, containing all commonly used commands including clipboard, formatting, alignment etc.
- **View** tab, containing placeholder commands, views and panes.
- Quick Access Toolbar, by default containing **Save**, **Undo**, **Redo** and **Customize**. Customize allows you to select how to display the ribbon and which of the following commands to display: **New**, **Open**, **Save**, **Save as**, **Print**, **Print preview**, **Undo**, **Redo**.
- Main window, showing the report template.

---

## File tab

On the **File** tab you find the general Windows commands and additional MODDE specific commands; **New**, **Open**, **Save**, **Save as**, **Templates**, **Print**, **Exit**, and **Recent templates**.



### General Windows commands

Click **New** to create a new report from the selected template.

Click **Open** to open a report or template saved in MHT or HTML format.

Click **Save** to save the report with the current investigation.

Click **Print** to Print the report, open the **Page setup** dialog, or **Print preview** to preview the report.

Click **Exit** to close the report.

## Save as

Click **File** | **Save as** |

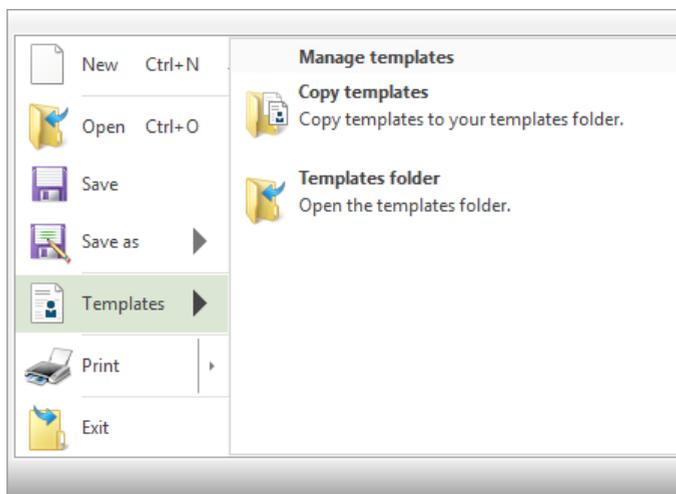
- **Save a copy** to save the report as a separate file in MHT or HTML format.  
Note that this separate file is *NOT* saved with the investigation, but only as a separate copy of the report. However, if you then click **Save** in the report and also **Save** in the MODDE window, the current report is also saved with the .mip.
- **Save as template** to save the current template when you have changed/created a template according to your wishes and want to save it to use it as template when you generate a report.

You can save and copy custom templates by clicking **Templates** on the **File** tab.

## Templates

Click **Templates** |

- **Copy templates** to search for templates to copy to the templates folder making them available in the **Select templates** dialog box when creating a new report.
- **Templates folder** to view the available custom templates, delete or copy.




---

*Note: The default templates **Default report** and **Blank report** cannot be modified nor deleted.*

---

## Recent templates

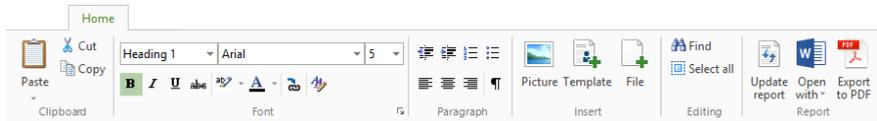
Under **Recent templates**, the templates used to create a new report are listed.

---

## Home

### Introduction

On the **Home** tab you find the most commonly used commands, organized in groups: **Clipboard**, **Font**, **Paragraph**, **Insert**, **Editing**, and **Report**.



## Clipboard

In the **Clipboard** group you find the standard commands

- **Paste** (Ctrl+V)
- **Paste | Paste unformatted** (Ctrl+Shift+V)
- **Cut** (Ctrl+X)
- **Copy** (Ctrl+C)

To clear the marked content, press the Del key.

**Undo** (Ctrl+Z) and **Redo** (Ctrl+Y) are available in the Quick Access Toolbar.

## Font and Paragraph

For details about the **Font** and **Paragraph** groups, see the **Formatting** topic.

## Insert

Use the **Insert** group to insert a **Picture**, **Template**, or **File** into the current report.

### *Picture*

Click **Picture** to insert a picture in the report. You can also right-click the report and then click **Insert picture**.

### *Template*

Click **Template** to open a dialog box where you can select a custom template to insert in the report. To be able to select the template it must be available in the templates folder. Templates can be added manually or by using the **File | Save as | Save as template** or **File | Templates | Copy template** commands.

### *File*

Click **File** to insert a web page file (\*.htm, \*.html), text file (\*.txt), or picture file (\*.jpg, \*.png, \*.gif, \*.bmp).

## *Editing*

In the **Editing** group you find the standard commands

- **Find** (Ctrl+F)
- **Select all** (Ctrl+A)

## Report

The **Report** group contains commands for updating all placeholders, viewing the report in a browser, editing the report in the editor of your choice, and exporting the report to PDF.

### Update report

Click **Update report** to update all placeholders to the default plots and lists of the current investigation.

Note that any customization, not saved in the currently used plot template, will be lost when clicking **Update report**. For instance, if you have enlarged/colored one point in a scatter plot and added it to the report, clicking **Update report** will result in the default plot, without that customization.

To keep the visualization of a plot, with customizations, remove the placeholder for that particular plot by clicking it so that it becomes selected and on the **View** tab clicking **Update placeholder**.

### Open with

Click **Open with** to view or edit the current report. You can continue to work with your report in the editor of your choice by selecting an editor. The applications listed here are the installed applications that have registered that they can view or edit HTML text with Windows.

### Export to PDF

Click Export to PDF to export the current report to PDF using `wk<html>topdf`, <http://wkhtmltopdf.org/>.

## Formatting

In the **Font** group the standard commands, such as text style, font, font size, bold, italic, underline, strikethrough, highlight color, and text color, are available. The two last commands are:

- **Convert to hyperlink** - Converts the current marking to a hyperlink after selecting hyperlink type and url.
- **Styles** - Opens the **Styles and Formatting** dialog for customization and preview of the text styles in the report.



In the **Paragraph** group, indenting, numbering, bullets, aligning and show formatting are available.




---

## View tab

On the **View** tab you can select to hide or show the following:

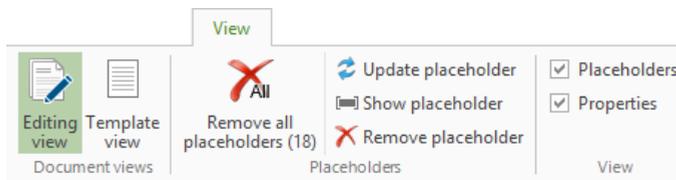
### Document views

- **Editing view** - regular report with placeholders filled with content.

- **Template view** - report displaying placeholders.

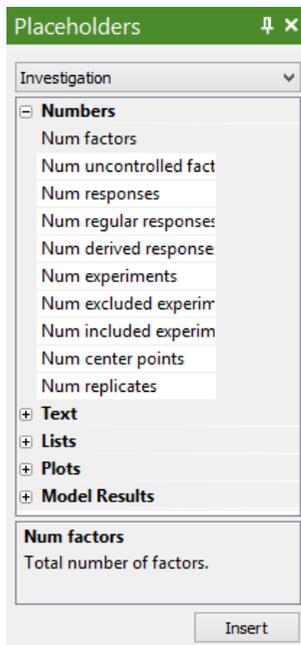
### Placeholders

- **Remove all placeholders** - removes all placeholders in the report leaving static plots, lists and details.
- **Update placeholder** - updates the selected placeholder.
- **Show placeholder** - shows the selected placeholder.
- **Remove placeholder** - removes the selected placeholder.
- **Placeholders** - displays/hides the **Placeholders** window holding all placeholders that can be added to the report.
- **Properties** - displays/hides the properties of plots and images that can be customized. The Properties window also displays the default settings for plots.



## Placeholders window

Open the **Placeholders** window by selecting the **Placeholders** check box on the **View** tab.



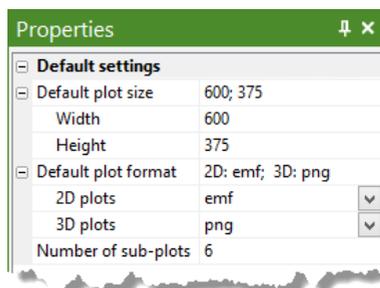
The combo box lists different categories of placeholders.

To insert a placeholder in the report, mark the placeholder and click **Insert**, alternatively double-click the placeholder. The newly inserted placeholder is automatically updated.

To show the underlying placeholders, on the **View** tab, click **Template view**.

## Properties window

Open the **Properties** window by selecting the **Properties** check box on the **View** tab.



### Default settings

This category contains the report generator's default settings. Plot sizes, etc are only updated when the placeholders in the report are updated ('Update report').

In the **Properties** window you can change the default plot size and to save a plot as .png, .bmp, .jpg or .emf (.emf for 2D only).

Properties for placeholders are displayed when you click the placeholder. You can change the properties of the current placeholder in the **Placeholder settings** section in the properties window. The **Default settings** section displays and allows changing of the default placeholder properties.

*Note: For changes in **Default settings** to take effect the report has to be updated (**Update report**), reverting all plots and lists to the default for the current investigation. See also the **Update report** subsection earlier in this chapter.*

## Adding plots and lists to the report

To add a plot or list to the report;

1. Click the desired position in the report so that the pointer is where the plot or list should be placed.
2. Right-click the plot or list in MODDE.
3. Click **Add to report**.

**Hint:** You can press Ctrl+R in MODDE to place the currently selected plot or list into the current location in the report.

# Statistical appendix

---

## Introduction

The statistical appendix gives a more detailed, mathematical and statistical look into the calculations used in MODDE.

---

## Fit methods

MODDE currently supports two regression methods for fitting a model to the data. These are multiple linear regression (MLR) and partial least squares regression (PLS or PLS regression). Both models predict one (or several) dependent variables,  $\mathbf{Y}$  ( $n \times m$ ), also called the responses, by a regression model from a set of independent variables,  $\mathbf{X}$  ( $n \times p$ ), also called the factors, such that

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E},$$

where  $\mathbf{B}$  is a  $p \times m$  vector of regression coefficients and  $\mathbf{E}$  is a matrix of residuals. The prediction of the dependent variables is

$$\hat{\mathbf{Y}} = \mathbf{XB}.$$

The column dimension,  $p$ , in the data matrix,  $\mathbf{X}$ , denotes the number of terms in the model, and includes the intercept.

### Multiple linear regression (MLR)

Multiple linear regression (MLR) is extensively described in the literature, and this section will only identify the numerical algorithm used to compute the regression coefficients. For additional information on MLR, see e.g. Draper and Smith (1981).

The MLR model relates a set of dependent variables,  $\mathbf{Y}$  ( $n \times m$ ), also called the responses, by a linear regression model to a set of independent variables,  $\mathbf{X}$  ( $n \times p$ ), also called the factors, such that

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E},$$

where  $\mathbf{B}$  is a  $p \times m$  vector of regression coefficients and  $\mathbf{E}$  is a matrix of residuals. The prediction of the dependent variables is

$$\hat{\mathbf{Y}} = \mathbf{XB}.$$

In MLR, the regression can be posed as an optimization problem

$$\text{maximise}_{\mathbf{B} \in \mathbb{R}^{p \times m}} \frac{1}{2} \|\mathbf{Y} - \mathbf{XB}\|_2^2.$$

MLR thus minimizes the squared difference between the dependent variables,  $\mathbf{Y}$ , and the linear prediction model,  $\mathbf{XB}$ . MODDE uses the singular value decomposition (SVD) to obtain the regression coefficients; see Golub and Van Loan (1983) for a description of the SVD and its use to obtain the regression coefficients.

---

*Note: In case of missing data in a row, the row is excluded for the relevant response before the MLR model is fit.*

---

### References

Draper, Norman and Smith, Harry, *Applied Regression Analysis*, Second Edition, Wiley, New York, 1981.

Golub, Gene H. and Van Loan, Charles F, *Matrix Computations*, *The Johns Hopkins University Press*, Baltimore, 1983.

### Partial least squares regression (PLS)

When the problem is poorly conditioned, the variables are correlated or  $p > n$ , regularization is needed in order to stabilize the model and to avoid overfitting. PLS regression is a latent variable method that offers regularization by representing the dependent variables by a small set of linear combinations (the latent variables) of the variables, and by that reducing the dimension of the space spanned by the independent variables. PLS also offers improved interpretation of the regression model.

PLS regression is described extensively in the existing literature, which is why only a brief description will be given here.

#### PLS1

In the case when  $m = 1$ , PLS solves the same loss function as MLR, but with a constraint on the regression coefficients. This problem is sometimes called PLS1. The optimization problem is posed as

$$\text{maximise}_{\beta \in \mathcal{K}_A(\mathbf{X}'\mathbf{X}, \mathbf{X}'\mathbf{y})} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\beta\|_2^2,$$

where

$$\mathcal{K}_A(\mathbf{X}'\mathbf{X}, \mathbf{X}'\mathbf{y})$$

is a Krylov subspace generated by  $\mathbf{X}'\mathbf{X}$  and  $\mathbf{X}'\mathbf{y}$ , and where  $A$  is the number of components of the PLS regression model.

PLS regression is a non-linear regression method, because the dependent variable,  $y$ , is included in the prediction model.

---

*Note: When the models for the responses are different, PLS regression fits each response separately.*

---

#### PLS2

The case when  $m > 1$  is sometimes called PLS2. The PLS regression model relates a set of dependent variables,  $\mathbf{Y}$  ( $n \times m$ ), also called the responses, by a regression model to a set of independent variables,  $\mathbf{X}$  ( $n \times p$ ), also called the factors, such that

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E},$$

where  $\mathbf{B}$  is a  $p \times m$  vector of regression coefficients and  $\mathbf{E}$  is a matrix of residuals. The prediction of the dependent variables is

$$\hat{\mathbf{Y}} = \mathbf{XB}.$$

The PLS regression algorithm creates new variables,  $t_a$ , called  $X$  scores, as linear combinations of the original variables, i.e., such that  $t_a = Xw_a$ , where  $w_a$  is a vector of weights. These  $X$  scores are few, often just two or three, and orthogonal. The  $X$  scores are then used to model the responses.

With more than one response, i.e., when  $m > 1$ , the  $Y$  variables are similarly combined to a few  $Y$  scores,  $u_a$ , by using weights,  $c_a$ , such that  $u_a = Yc_a$ .

The PLS regression model finds the score vectors by solving the optimization problem

$$\begin{aligned} & \text{maximize} && \text{Cov}(X_a w_a, Y_a c_a), \\ & \text{subject to} && \|w_a\|_2 = \|c_a\|_2 = 1, \end{aligned}$$

for  $a = 1, 2, \dots, A$ , such that all  $X_a w_a$  and  $X_b w_b$  for  $a \neq b$  are orthogonal. The orthogonality is enforced by deflating the found variation from the  $X$  and  $Y$  matrices after each component such that

$$\begin{aligned} X_{a+1} &= X_a - t_a p'_a, \\ Y_{a+1} &= Y_a - t_a c'_a, \end{aligned}$$

where  $p_a = X'_a t_a / (t'_a t_a)$  are called loadings and  $X_I = X$ .

The PLS2, the multi- $Y$  problem is less well understood compared to the single- $y$  problem, but nevertheless performs impressively when  $p > n$ .

One PLS regression component, indexed by  $a$ , consists of one vector of  $X$  scores,  $t_a$ , and one of  $Y$  scores,  $u_a$ , together with the  $X$  and  $Y$  weights,  $w_a$  and  $c_a$ . The scores are collected such that  $T = [t_1, t_2, \dots, t_A]$  and  $U = [u_1, u_2, \dots, u_A]$ , and the loadings  $P = [p_1, p_2, \dots, p_A]$ . The PLS regression model finally regresses  $U$  on  $T$ . The final PLS regression model is stated as

$$\begin{aligned} X &= TP' + E, \\ Y &= TC' + F. \end{aligned}$$

The regression coefficient vector is computed as

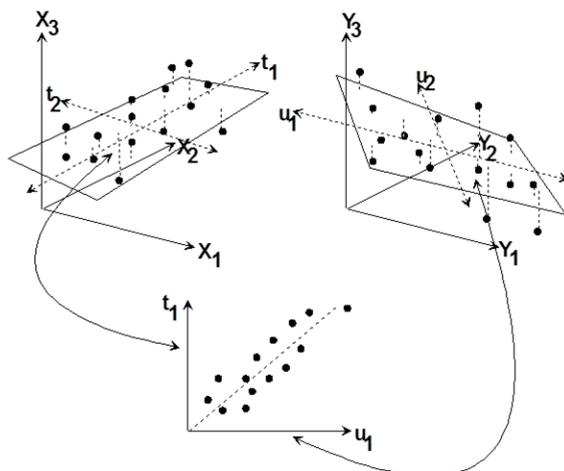
$$B = W(P'W)^{-1}C',$$

and predictions are made as

$$\hat{Y} = XB$$

A geometric interpretation of the PLS regression method is that it projects the  $X$  and  $Y$  matrices on lower-dimensional hyper planes, spanned by  $w_a$  and  $c_a$ , for  $a = 1, 2, \dots, A$ , respectively, and that the coordinates in this subspace are  $T$  and  $U$ , that summarize  $X$  and  $Y$ , respectively, and that the regression model relates  $T$  to  $U$ .

The matrices  $X$  and  $Y$  can be seen as  $n$  points in two Euclidean vector spaces (in  $\mathbb{R}^{n \times p}$  and  $\mathbb{R}^{n \times m}$ ), respectively (see the figure below) the  $X$  space with  $p$  coordinates, and the  $Y$  space with  $m$  coordinates,  $p$  and  $m$  being the number of columns in  $X$  (terms in the model) and in  $Y$  (responses).



The number of components,  $A$ , is determined by a cross-validation procedure, in which the prediction error sum of squares (PRESS) is minimized. When  $A = \text{rank}(X)$ , the regression coefficients of PLS regression are equivalent to those of MLR, and hence MLR could be considered as a special case of PLS regression. That is, if you with PLS regression extract all components, the PLS regression model is equivalent to the MLR model.

### Cross-validation significance rules

In order to determine if a component is significant, MODDE considers the following rules in order:

A component is considered *non-significant* by **Rule 3** if:

$$A + 1 > \min(n,p).$$

A component is considered *significant* by **Rule 1** if

$$Q^2 > CV_{lim}$$

or otherwise significant by **Rule 2** if

$$Q_k^2 > CV_{lim},$$

for any  $k = 1, \dots, m$ .

A component is considered *non-significant* by **Rule 4**, regardless of **Rule 1** and **Rule 2**, if

$$R_A^2 < 0.01$$

and

$$R_{A,k}^2 < 0.03$$

for all  $k = 1, \dots, m$ , and if was not deemed *non-significant* by **Rule 3**.

If no rule at all has come in effect yet, the component is considered *non-significant* by **Rule 5**.

The cross-validation limit for  $Q^2$ , i.e.,  $CV_{lim}$ , is defined as

$$CV_{lim} = \begin{cases} -0.2 & \text{if } A \leq \frac{\min(n, p)}{2} \\ 0.05 & \text{otherwise} \end{cases}$$

MODDE will try to compute at least two PLS regression components (if there exist two), regardless of their significance according to these rules.

### **Singular models**

A model is considered singular if the condition number of  $\mathbf{X}$  is greater than 3000. Such models may only be fit by PLS regression.

### **Model**

You may edit the model and add or delete terms for individual factors. You may add up to third order terms (i.e., cubic terms, and 3-factor interactions).

In order to regularize a design that is singular with respect to your model, MODDE will fit a PLS regression model to the data. MLR will not be available in this situation.

### **Hierarchy**

MODDE enforces a hierarchy of model terms, where simpler terms are higher in the hierarchy. You cannot delete the constant term, it is at the top of the hierarchy. A linear term can only be deleted if the second and third order terms involving the factor have been removed first.

## **Scaling**

### **Scaling the factor matrix, $\mathbf{X}$**

When the model is fit with multiple linear regression, the design matrix,  $\mathbf{X}$ , is scaled and centered by the method specified in the **Factor Definition** dialog box, in the **Advanced** tab, by changing the value in the **MLR scaling** box. The default scaling for MLR is *Orthogonal* scaling.

When the model is fit with PLS regression, the design matrix,  $\mathbf{X}$ , is always scaled by *Unit variance scaling* (see below).

If needed, the scaled design matrix,  $\mathbf{X}$ , is extended with higher order terms, such as squares, cubes and interaction terms, according to the selected model.

The available scaling choices are:

**Orthogonal scaling**, where

$$z_{ij} = (x_{ij} - M_j)/R_j$$

in which  $z_j$  denotes a scaled factor and  $x_j$  denotes the original factor,  $M_j$  is the mid-range of the factor and  $R_j$  the range.  $M_j$  and  $R_j$  are computed as

$$M_j = (\min_i(x_{ij}) + \max_i(x_{ij}))/2$$

$$R_j = (\max_i(x_{ij}) - \min_i(x_{ij}))/2$$

**Mid-range scaling**, where

$$z_{ij} = (x_{ij} - M_j)$$

**Unit variance scaling**, where

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j},$$

where

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^n x_{ij}$$

and

$$\sigma_j = \sqrt{\frac{1}{N-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

are the mean and standard deviation of the  $j$ th variable of  $\mathbf{X}$ , where  $N$  is the number of rows without missing data in the current column.

---

*Note: Orthogonal and mid-range scaling are only available for use with MLR.*

---

## Scaling the response matrix, $\mathbf{Y}$

The responses in the matrix  $\mathbf{Y}$  are not scaled with MLR. With PLS regression, on the other hand, they are by default scaled to *unit variance*. This may be changed by changing the value in the **PLS scaling** box in the **Response Definition** dialog box to something else.

---

## Missing data

### Missing data in the design matrix, $\mathbf{X}$

Missing data in the design matrix,  $\mathbf{X}$ , are not allowed, and will disable the fit. This does, however, not apply to uncontrolled  $\mathbf{X}$  variables. For an MLR model with missing data in the design matrix, in an uncontrolled factor, the row with missing data is excluded from the calculations.

PLS regression can handle missing data in uncontrolled factors.

### Missing data in $\mathbf{Y}$ with multiple linear regression

For MLR models, each row with missing data in  $\mathbf{Y}$  is excluded for that specific response. Therefore, the number,  $N$ , displayed in plots and lists is the number of runs (experiments) with no missing data at all.

### Missing data in $\mathbf{Y}$ with PLS regression

When the model for all responses is the same and the model is fit with PLS regression, missing data is handled differently compared to how it is handled in MLR models.

When all values in a row of  $\mathbf{Y}$  are missing, that row is excluded from the analysis. When there are only some values missing in a row of  $\mathbf{Y}$ , those rows are kept in the model.

This leads to minor differences in the displayed number of runs,  $N$ , and degrees of freedom, DF, at the bottom of plots and lists when MLR or PLS regression models are used.

When different models are fit to different responses, missing data is handled the same as it is with an MLR model, see above.

---

## Number of observations

The number of observations (experiments),  $N$ , as used in ANOVA and for the computation of the adjusted  $R^2$ , is the actual number of non-missing observations for each response column. This value of  $N$  and the degrees of freedom,  $DF = N - p$ , are displayed at the bottom of ANOVA plots and lists, and on all residual plots (including plots of observed vs. predicted  $\mathbf{Y}$ ).

---

## Summary of fit

### R2

The first bar in the Summary of fit plot is the  $R^2$ , which is the fraction of the variation of the response explained by the model, *i.e.*

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where  $SS_{\text{res}}$  is the sum of squares of the residual, corrected for the mean and  $SS_{\text{tot}}$  is the total sum of squares of  $\mathbf{Y}$  corrected for the mean.

Note that the  $R^2$  overestimates the goodness of fit.

The  $R^2$  value is always between 0 and 1. A value close to 1 indicates that the model fit the data very closely.

$R^2$  *Adjusted* is the fraction of variation of the response explained by the model, but adjusted for the degrees of freedom of the model, *i.e.*

$$R_{\text{adj}}^2 = 1 - \frac{SS_{\text{res}}/DF_{\text{res}}}{SS_{\text{tot}}/DF_{\text{tot}}} = 1 - (1 - R^2) \frac{n - 1}{DF_{\text{res}}},$$

where  $DF_{\text{tot}} = n - 1$  is the degrees of freedom of the population variance of the dependent variable and  $DF_{\text{res}}$  is the degrees of freedom of the estimate of the population error variance.

### Q2

The second bar in the Summary of fit plot is  $Q^2$ . The  $Q^2$  estimates the predictive ability of the model (also known as model predictive power), *i.e.*, its ability to generalize to new, unseen data. The  $Q^2$  is either computed in closed form as in generalized cross-validation or computed in a cross validation-like procedure and is expressed in the same units as  $R^2$ . The  $Q^2$  is defined as

$$Q^2 = 1 - \frac{\text{PRESS}}{SS_{\text{tot}}} = 1 - \frac{\text{PRESS}}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where PRESS is the prediction residual sum of squares, which differs for MLR and PLS regression models,  $SS_{\text{tot}}$  is the total sum of squares of  $\mathbf{Y}$  corrected for the mean and

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

is the mean response. See the Measures of goodness-of-fit subsection later in this chapter for more details about  $Q^2$ .

The  $Q^2$  is less than or equal to 1, and a value greater than zero indicates that the model is significant (it performs better than just predicting the mean value,  $\bar{Y}$ , for each response).

If the regression model is a PLS model, negative  $Q^2$  values are set to zero, for computational purposes.

### Model validity

The third bar in the Summary of fit plot is the *Model Validity*. When the model validity column is larger than 0.25, the model has no *lack-of-fit*. This means that the model error is not significantly larger than the pure error.

When the model validity is less than 0.25 there is a significant lack-of-fit and the model error is significantly larger than the pure error (poor reproducibility).

The model validity is computed as

$$\text{Validity} = 1 + 0.57647 \log_{10}(p_{\text{tof}}),$$

where  $p_{\text{tof}}$  is the  $p$ -value for the lack-of-fit test and the value 0.57647 is selected such that  $p_{\text{tof}} \geq 0.05$  gives  $\text{Validity} \geq 0.25$ .

### Reproducibility

The fourth bar in the Summary of fit plot is the *Reproducibility*, which is the variation of the response (often at the center points) under the same conditions as the total variation of the response. It is computed as

$$\text{Reproducibility} = 1 - MS_{pe}/MS_{tot},$$

where  $MS_{pe}$  is the mean square of the pure error and  $MS_{tot}$  the total mean square of  $Y$ .

A reproducibility of 1 signifies perfect reproducibility.

## Residual standard deviation (RSD)

If the model is a PLS regression model, and it is the same for all responses, the residual standard deviation is computed by using the total number of observations (experiments, runs), i.e., without excluding the rows with missing values. This is displayed in the summary table and at the bottom of all plots and lists, including the ANOVA.

For MLR and PLS regression with different models for the responses, the residual standard deviation is instead computed by using the number of experiments without missing values for each response.

This residual standard deviation is used in interval estimation, for coefficients and in predictions.

---

## Analysis of variance, ANOVA

Analysis of variance (ANOVA) decomposes the total sum of squares of a selected response (sum of squares corrected for the mean) into a part due to the regression model and a part due to the residuals. *I.e.*, such that

$$SS_{tot} = SS_{reg} + SS_{res}.$$

With corresponding decomposition of the degrees of freedom as

$$DF_{tot} = DF_{reg} + DF_{res}.$$

If there are replicated experiments (observations, runs), the residual sum of squares can be further decomposed in a part due for the *pure error* and one part due to the *lack-of-fit* of the model, *i.e.* such that

$$SS_{res} = SS_{pe} + SS_{lof}.$$

where the degrees of freedom is decomposed as

$$DF_{res} = DF_{pe} + DF_{lof} = \sum_k (n_k - 1) + \left( n - p - \sum_k (n_k - 1) \right),$$

where  $n_k$  is the number of replicates in the  $k$ th set of replicates. Note that rows with missing values have been excluded.

---

**Note:**  $DF_{lof}$  (degrees of freedom for lack of fit) is the same as  $RDF$  (real degrees of freedom).

---

A *goodness-of-fit* test is performed by an  $F$ -test between the lack-of-fit and pure error sum of squares. The test statistic is

$$F = \frac{SS_{lof}/DF_{lof}}{SS_{pe}/DF_{pe}},$$

which has an  $F$ -distribution with the corresponding number of degrees of freedom.

Two ANOVA plots are displayed:

- The regression goodness-of-fit test.
- The goodness-of-fit test of the lack-of-fit.

### **Identifying replicates**

MODDE checks the rows of the worksheet for replicates. Rows in the worksheet are considered replicates if they match all factor values plus or minus a small tolerance. The default tolerance is 10 % of half the range of a factor, but it can be altered in the MODDE options page, in the **File | Options** dialog box.

---

## Measures of goodness-of-fit

MODDE computes and displays the following statistics:

### Q2

The regression models are validated by the  $Q^2$  (sometimes also called  $q^2$ ,  $R_{cv}^2$ , or  $R_{pred}^2$  in the literature). The  $Q^2$  estimates the predictive ability of the model, i.e., its ability to generalize to new, unseen data. The  $Q^2$  is defined as

$$Q^2 = 1 - \frac{\text{PRESS}}{\text{SS}_{\text{tot}}},$$

where

$$\text{SS}_{\text{tot}} = \sum_{i=1}^n (y_i - \bar{y})^2$$

is the sum of squares of the training samples around the mean (proportional to the variance of the response) and

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

is the mean response.

When the regression model is an MLR model, the "PRESS" is computed as in *generalized cross-validation* (see e.g., Golub et al. (1979)), i.e., by

$$\text{PRESS}_{\text{MLR}} = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - x_i(X'X)^{-1}x_i'} \right)^2,$$

where the predictions  $\hat{y}_i$  are computed from a model of all samples.

When the model is a PLS regression model, the PRESS is computed in a cross-validation procedure. The cross-validation procedure keeps a part of the data (e.g.,  $1/K$ th of the samples) out of the model and predicts the left-out samples from a model of the kept samples (i.e., the rest of the  $(K - 1)/K$  samples). In this case the PRESS is computed as

$$\text{PRESS}_{\text{PLS}} = \sum_{i=1}^n (y_i - \hat{y}_i^{cv})^2,$$

where  $\hat{y}_i^{cv}$  is the predicted left-out response values from the cross-validation procedure.

For PLS regression models, the  $Q^2$  is computed both for all responses and such that a  $Q_j^2$  is computed for each individual response,  $j = 1, 2, \dots, m$ .

A computed  $Q^2$  value greater than zero indicates that the model is significant (it performs better than just predicting the mean value for each response). A high  $Q^2$ , e.g.,  $Q^2 > 0.5$  indicates that the model has good predictive ability and will have only a small prediction error on new samples. The  $Q^2$  is considered large if  $Q^2 > 0.7$ .

### References

Golub, Gene H.; Heath, Michael and Wahba, Grace. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21 (2), 1979.

## R<sup>2</sup>

The regression models are also validated by the  $R^2$ , the coefficient of determination (also sometimes called the multiple correlation coefficient). The  $R^2$  denotes the fraction of the response that is explained by the model and is defined as

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where the predicted values,  $\hat{y}_i$ , are computed from a model of all samples.

The  $R^2$  could be considered an upper bound on the estimate for how well the model predicts outcomes of new experiments.

## Adjusted R<sup>2</sup>

The *adjusted*  $R^2$  is the fraction of variance of the response explained by the model, but adjusted for the degrees of freedom of the model, *i.e.*

$$R_{\text{adj}}^2 = 1 - \frac{SS_{\text{res}}/DF_{\text{res}}}{SS_{\text{tot}}/DF_{\text{tot}}} = 1 - (1 - R^2) \frac{n - 1}{DF_{\text{res}}},$$

where  $DF_{\text{tot}} = n - 1$  is the degrees of freedom of the population variance of the dependent variable and  $DF_{\text{res}}$  is the degrees of freedom of the estimate of the population error variance. Note that  $R_{\text{adj}}^2$  may be negative.

The value of  $R_{\text{adj}}^2 \leq R^2 \leq 1$  only increases when new explanatory variables explain the data more than what can be attributed to chance, unlike the  $R^2$  which increases when new explanatory variables are added to the model.

## F-test

An *F*-test can be used to test the significance of a regression model. This tests the null hypothesis

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

against the alternative hypothesis

$$H_A: \beta_j \neq 0, \text{ for at least one } j=1, \dots, p.$$

The *F* quotient is

$$F = \frac{\frac{SS_{\text{reg}}}{DF_{\text{reg}}}}{\frac{SS_{\text{res}}}{DF_{\text{res}}}} = \frac{SS_{\text{reg}} \cdot DF_{\text{res}}}{SS_{\text{res}} \cdot DF_{\text{reg}}},$$

Where  $SS_{\text{reg}}$  is the sum of squares of the regression, corrected for the mean, and  $SS_{\text{res}}$  is the sum of squares of the residual, corrected for the mean. The  $DF_{\text{reg}}$  and  $DF_{\text{res}}$  are the corresponding degrees of freedom.

The null hypothesis is rejected if the computed value, *F*, is greater than the critical value of the *F* distribution at some significance level, typically at 0.05.

## Degrees of freedom and saturated models

MODDE computes the *real* degrees of freedom of the residual,  $RDF_{res}$ , as

$$RDF_{res} = N - p - \sum_k (n_k - 1),$$

where  $n_k$  is the number of replicates in the  $k$ th set of replicates,  $N$  is the number of runs without missing data and  $p$  is the number of terms in the model (including the constant).

If

$$p > N - \sum_k (n_k - 1)$$

the *real* degrees of freedom of the residual is defined as

$$RDF_{res} = \begin{cases} \sum_k (n_k - 1) & \text{if } n_i > 0 \text{ for any } i, \\ 0 & \text{otherwise.} \end{cases}$$

When  $RDF_{res} = 0$ , the model is saturated. In this case, MODDE does not compute or display  $R^2$ ,  $R^2$  *adjusted* or  $Q^2$  if the model is fitted by MLR. If the model is fitted by PLS regression, only the  $Q^2$  is computed and displayed.

The *real* degrees of freedom are used in the residual plots to determine which type of residuals to display when the degrees of freedom is small.

---

*Note: RDF is listed in the ANOVA and Descriptive statistics tables as DF for lack of fit.*

---

## Coefficients

### Scaled and centered coefficients

The regression coefficients that are displayed in MODDE are computed for centered and scaled data. It is also possible to select to display "unscaled and uncentered" coefficients.

The scaled and centered coefficients are the coefficients of the fitted model, for which the factors were centered and scaled. The default scaling in MLR is *orthogonal* scaling. With PLS, the factors are centered and scaled to *unit variance*.

### Orthogonal coefficients in PLS

The "centered and scaled" coefficients in PLS regression models are computed from factor values scaled to unit variance.

The *orthogonal coefficients* in PLS regression re-express the coefficients such that they correspond to factors that are *centered* and *orthogonally scaled*, i.e., by using the mid-range and low and high values in the factor definition (coded as -1 and 1), see the Scaling subsection for more details. Orthogonal coefficients in PLS regression models are not available when there are only formulation factors in the investigation.

With process and mixture factors, the PLS orthogonal coefficients refer to process factors scaled orthogonally, and mixture factors unscaled (original units).

The orthogonal coefficients in PLS regression models are only available when the model is fit with PLS regression.

---

*Note: The orthogonal coefficients in PLS regression models are only meant for comparison with the corresponding MLR coefficients. They are incorrect unless the design is balanced and the mean is equal to the mid-range.*

---

## Unscaled

The unscaled coefficients are the coefficients corresponding to unscaled, uncentered data. When exporting unscaled coefficients to use in other applications, be sure to use the *E-format* in order to obtain maximum precision in the coefficients.

## Normalized coefficients

To make the coefficients comparable between responses when the responses have different ranges, the "centered and scaled" coefficients are normalized with respect to the variation in  $Y$ . That is, they are divided by the standard deviation of their respective response, *i.e.*, by the standard deviation in the corresponding  $Y_i$ , for  $i = 1, 2, \dots, m$ .

## Confidence interval

Intervals for coefficients (such as *e.g.*, a confidence interval) and predictions are computed using the total number of observations, regardless of missing values when the regression model is a PLS regression model and all polynomial models (*i.e.*, the polynomial model specified by the factors) for all responses are the same. For MLR and PLS regression models with different polynomial models for different responses, the total number of observations is the number of elements in the response without missing values. This total number of observations is displayed as  $N$  at the bottom of plots and lists.

## Extended or compact format

For qualitative factors at  $q$  levels, with  $q > 2$ , MODDE generates  $q - 1$  dummy variables, indexed from 2 to  $q$ .

By default the Coefficient plot displays all  $q$  settings using the **Extended** format. See the Investigation options subsection in Chapter 4, File, for details.

---

## Analysis wizard features

### Tukey's and variability tests

On the Replicates page in the Analysis Wizard two tests are performed, Tukey's test that checks for outliers, and the Variability test that checks replicated point's variability.

#### *Tukey's test*

When Tukey's test fails this indicates an outlier. Information about this is displayed in the Advisor in the Analysis Wizard. Tukey's test is only available in the Analysis Wizard.

Tukey's test classifies an experiment as an outlier when outside the range  $Q1-k(Q3-Q1)$  to  $Q3+k(Q3-Q1)$ .

In MODDE;

- $k = 3$  is used
- $Q1 = Q(25\%)$ , listed in the Descriptive statistics list.
- $Q3 = Q(75\%)$ , listed in the Descriptive statistics list.

#### *Variability test*

The variability test compares the replicate variability, for each replicate group, with the overall variability and displays information about the outcome in the Advisor pane in the Analysis Wizard. The displayed information:

1. Replicate range  $> IQR$  results in a warning that since the replicate variability is large there is also a large probability for a non-significant model.
2. Replicate range  $0.5 IQR < \text{Replicate range} < IQR$  results in a warning that the large replicate variability may result in a weak model.
3. Replicate range  $< 0.5 IQR$  results in information that the replicate variability is fairly small and it is likely to end up with a useful model.
4. Replicate range  $< 0.25 IQR$  results in information that the replicate variability is small and it is likely to end up with a useful model.
5. Replicate range  $< 0.02 IQR$  results in information that the replicate variability is very small and although positive, this may give artificial model validity problems.

---



---

**Hint:**  $IQR = Q3 - Q1 = Q(75\%) - Q(25\%)$

---



---

### Auto transform and Auto tune

When using One-Click, Auto transform and Auto tune are performed automatically. Auto transform is located on the Histogram page and Auto tune on the Coefficient page.

#### *Auto transform*

*Auto transform* is available when the Skewness test value is higher than 2. The default log transform is applied and if there are values that are  $\leq 0$  the constant C2 is increased such that the smallest value is 0.1.

When the Skewness test value is smaller than -2 there is also a warning suggesting a transformation, but no automatic transformation is available.

Clicking Auto transform log-transforms the response if transforming it results in higher  $Q^2$  and lower Skewness.

### **Auto tune**

*Auto tune* is available when both p-values for the model terms can be calculated and  $Q^2$  can be calculated. It is available for all designs but mixture designs, Stability testing designs and Generalized Subset Designs.

Auto tune works as follows:

1. Finds the largest p-value that is above 0.05.
2. Removes that term and recalculates the model.
3. If  $Q^2$  was increased, Auto tune restarts at step 1.
4. If  $Q^2$  was not increased, but there is yet another term with a p-value larger than 0.05, that term is removed in conjunction with the other term, and Auto tune continues at step 3. This step is repeated until all terms with a p-value larger than 0.05 have been tested.
5. Auto tune is complete after having tried excluding all terms with a p-value larger than 0.05, and no improvement in  $Q^2$  can be found.

---

*Note: Auto tune is available in on the Coefficient page in the Analysis Advisor and on the shortcut menu of Coefficient plots created outside the Analysis wizard.*

---

### **Square and interaction tests**

On the Coefficients page of the Analysis wizard there are two tests, the **Square test** and the **Interaction test**.

#### **Square test**

The *Square test* can be performed if adding a square term does not result in a condition number that exceeds 100.

The following test will make the Square test available, if successful. For each factor:

1. A square term of the factor is added to the current model.
2. The test measures whether the square term is significant at the 95 % confidence level and the  $Q^2$  increases or does not decrease by more than 0.05. If this is the case the **Square test** will be available.
3. If the square term is non-significant at the 95 % confidence level, the test measures whether the  $Q^2$  value increases by 0.05 or more. If this is the case, the **Square test** will be available.
4. Steps 1-3 are repeated until a square term of all factors have been added once.

#### **Interaction test**

The *Interaction test* is performed when the design is of resolution IV and the confounding pattern is known. That is, the design was created by MODDE.

The following test will make the Interaction test available, if successful. For each interaction term of each factor:

1. An interaction term is added to the current model.

2. The test measures whether the interaction term is significant at the 95 % confidence level and  $Q^2$  increases or does not decrease by more than 0.05. If this is the case the **Interaction test** will be available.
3. If the interaction term is non-significant at the 95 % confidence level, the test measures whether  $Q^2$  increases by 0.05 or more. If this is the case, the **Interaction test** will be available.
4. Steps 1-3 are repeated until all interaction terms have been added once.

## Qualitative factors with more than 2 levels

If a term in the model comprises a qualitative factor,  $C$ , with  $k$  levels, there will be  $k - 1$  expanded terms associated with that term. For example, if the levels of a qualitative factor  $C$  are  $(a, b, c, d)$  the three expanded terms  $C(j)$  are as follows:

<b>C</b>	<b>C(2)</b>	<b>C(3)</b>	<b>C(4)</b>
<b>a</b>	-1	-1	-1
<b>b</b>	1	0	0
<b>c</b>	0	1	0
<b>d</b>	0	0	1

The coefficients of these expanded terms are given as the coefficients for level 2 ( $b$ ), 3 ( $c$ ), and 4 ( $d$ ) of  $C$ , while the coefficient for level 1 ( $a$ ) is computed as the negative sum of the three others. MODDE displays all four coefficients in the **Coefficient List** but notes that they are associated with only three degrees of freedom.

*Note: When excluding all experiments for a certain setting, the factor is treated as if that setting never existed.*

## Residuals

### Raw residuals

The *raw residual* is the difference between the observed and the predicted values, *i.e.*

$$e_i = y_i - \hat{y}_i.$$

### Standardized residuals

The *standardized residual* is the raw residual divided by the residual standard deviation

$$e_{std,i} = e_i / s,$$

where  $s$  is the residual standard deviation.

Residual plots for PLS regression present standardized residuals by default.

### Deleted studentized residuals

MODDE defaults to plotting *Deleted studentized residuals* when fitting with MLR and the model has at least three (3) *real* degrees of freedom. Deleted studentized residuals are not available when fitting with PLS regression.

The Deleted studentized residual is the raw residual ( $e_i$ ) divided by an estimate of its standard deviation. The estimation of the standard deviation is computed from the “deleted” standard deviation ( $s_i$ ), which is the residual standard deviation ( $s_i$ ) computed with observation  $i$  left out. When the  $i$ th residual is excluded like this, the residual is sometimes said to be *externally studentized*. The variance of the  $i$ th residual is defined as

$$\text{Var}(e_i) = \sigma^2(1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i')$$

where  $\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'$  is the  $i$ th element of the hat matrix, sometimes denoted *leverage*. Thus the standard deviation of the residual is

$$\text{sd}(e_i) = \sqrt{\text{Var}(e_i)} = \sigma \sqrt{1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'}$$

Hence, the Deleted studentized residual is computed as

$$e_{del,i} = \frac{e_i}{\text{sd}(e_i)} = \frac{e_i}{s_{-i}\sqrt{1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'}}$$

where  $s_{-i}$  is the estimate of the residual standard deviation with the  $i$ th sample left out.

For more information see *e.g.*, Belsley, Kuh and Welsch (1980).

---

**Note:** Deleted studentized residual requires at least three degrees of freedom.

---

## References

Belsley, David A.; Kuh, Edwin and Welsch, Roy E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley and Sons, New York, 1980.

---

## Condition number

The condition number of the orthogonally scaled and centered extended design matrix is computed using the singular value decomposition (SVD). The condition number can be seen in the status bar at the bottom of MODDE's main window. The  $\mathbf{X}$  matrix is taken from the worksheet. The computation of the condition number depends on the selected fit method (MLR, PLS regression) and which factors are involved.

### Condition number definition

The condition number is defined as

$$\kappa(\mathbf{X}) = \|\mathbf{X}\|_2 \|\mathbf{X}^+\|_2,$$

where

$$\|\cdot\|_2$$

is the operator norm induced by the  $\ell_2$  (the Euclidean) norm and  $\mathbf{X}^+$  denotes the Moore-Penrose pseudo inverse of the matrix  $\mathbf{X}$ . This definition is equivalent to the ratio between the largest and the smallest singular values of  $\mathbf{X}$  (square root of the eigenvalues of  $\mathbf{X}'\mathbf{X}$ ), such that

$$\kappa(\mathbf{X}) = \frac{\sigma_{\max}(\mathbf{X})}{\sigma_{\min}(\mathbf{X})},$$

where  $\sigma_{\max}(\mathbf{X})$  is the largest singular value of  $\mathbf{X}$  and  $\sigma_{\min}(\mathbf{X})$  is the smallest non-zero singular value of  $\mathbf{X}$ .

The condition number is a measure of the sphericity of the design. All factorial designs, without center points, have a condition number of 1 and the design points are situated on the surface of a hypersphere.

The condition number is computed for the extended design matrix. The extended design matrix is created as follows:

1. The factor values, taken from the worksheet, are centered and scaled according to the selected settings: *MLR scaling*, for MLR models and (always) unit variance with PLS regression models.
2. The design matrix is then extended according to the selected polynomial model.

---

*Note: If you select Mid-range scaling and your factors have different ranges, the condition number of the worksheet may become very large. This is due to the fact that MODDE uses the SVD to compute the MLR model.*

---

### Condition number with mixture factors

The condition number with mixture data depends on the method of fit and the type of model.

### PLS and Cox reference mixture model

When the model is fit by PLS regression (or a Cox reference mixture model) the data are scaled and centered. The condition number is computed from the worksheet, with the slack variable model (mixture factor with the largest variance removed) and all mixture factors scaled with orthogonal scaling.

### MLR regression and the Cox reference mixture model

The condition number displayed is the condition number of the Cox reference mixture model (with the Cox constraints), derived from the worksheet, with no mixture factors scaled or centered.

### MLR and Scheffé models

The condition number of the Scheffé model, derived from the worksheet, with no mixture factors scaled or centered.

---

*Note: With formulation factors and fitting with PLS, the condition number is computed by excluding the factor with the largest range, and scaling the remaining ones orthogonal. When fitting with MLR the condition number is computed without centering and scaling the factors.*

---

## Interval estimates

MODDE supports three types of intervals: confidence intervals, prediction intervals and tolerance intervals. The figure below illustrates how the intervals relate to each other.

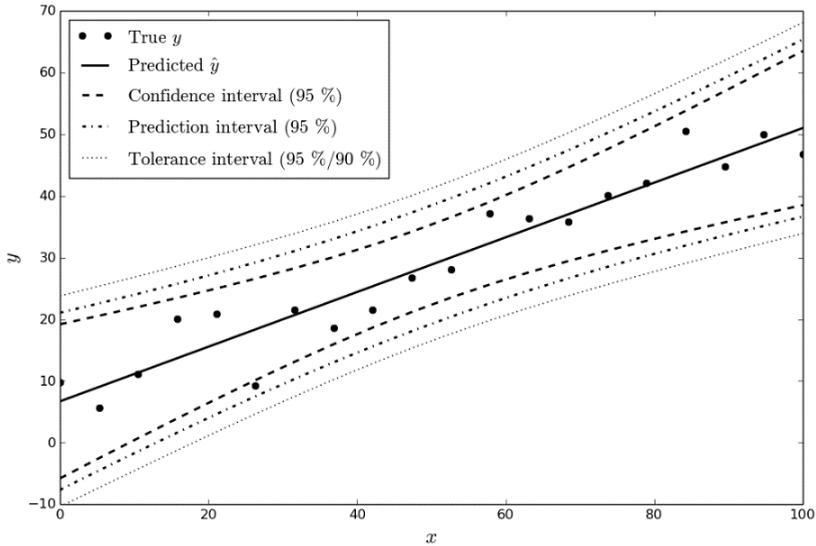


Figure. An illustration of the relation between the confidence, prediction and tolerance intervals. The *confidence interval* tells us that we are *e.g.* 95 % confident that the "true" mean prediction (*i.e.* the "true" underlying output) is contained by our interval, which means equivalently that if we repeat the study many times, 95 % of the corresponding confidence intervals will contain the true mean.

A  $100(1 - \alpha)$  % confidence interval around the mean predicted value is computed as

$$\mathbf{x}_i' \hat{\boldsymbol{\beta}} \pm t_{\alpha/2, n-p} \sqrt{s^2 \mathbf{x}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i},$$

for each sample (*i.e.*, row of  $\mathbf{X}$ ),  $\mathbf{x}_i$ , where

$$s^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}})^2$$

is the sample noise variance,  $t_{\alpha/2, n-p}$  is the critical value of a Student's *t*-distribution on the two-sided  $\alpha$  level with  $n - p$  degrees of freedom.

The *prediction interval* tells us that with a particular confidence, *e.g.*, 95 %, the interval will contain the *next* observation. *I.e.*, the prediction interval will enclose a *single* future observation with some confidence. A  $100(1 - \alpha)$  % prediction interval, for the next observation, is computed as

$$\hat{y}_i \pm t_{\alpha/2, n-p} \sqrt{s^2 (1 + \mathbf{x}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_i)}.$$

The *tolerance interval* states with some confidence, say 95 %, that some percent, say 99 %, of future observations fall within the interval (this is the proportion of future samples that fall within the interval, *i.e.* the *tolerance proportion*). The tolerance interval is formally stated as

$$P(P(\hat{y}_i - \sigma k_i \leq y_i \leq \hat{y}_i + \sigma k_i) \geq P) \geq 1 - \alpha,$$

where  $k_i$  is unknown and must be found, and  $\sigma$  is the standard deviation of the noise (approximated by  $s$ ). The value of  $k_i$  is approximated by

$$k_i \approx \sqrt{\frac{d \cdot \chi_{1,p}^2(\mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i')}{\chi_{d,\alpha}^2}},$$

where  $d = n - p$  is the number of degrees of freedom and  $\chi_{a,b}^2(c)$  is the  $b$  quantile in a non-central  $\chi^2$  distribution with  $a$  degrees of freedom and center parameter  $c$ .

Both the MLR and PLS regression models will compute an interval around the average predicted  $\mathbf{Y}$  if the independent matrix,  $\mathbf{X}$ , has a condition number less than or equal to 3000.

Matrices with condition number greater than 3000 are considered singular and it is only possible to build PLS regression models. When the condition number of  $\mathbf{X}$  is greater than 3000, or the model is a Cox reference mixture model, PLS regression computes the standard error of the average predicted  $\mathbf{Y}$  as

$$se_i = \sqrt{s^2(1 + \mathbf{t}_i(\mathbf{T}'\mathbf{T})^{-1}\mathbf{t}_i')},$$

where  $\mathbf{T}$  is the score matrix of the PLS regression model and  $\mathbf{t}_i$  is the  $i$ th row of  $\mathbf{T}$ . The interval is then computed as

$$y_i \pm se_i.$$

The type of interval to use depends on the purpose of computing the interval. The confidence interval is an interval around the mean prediction, and therefore does not give any information about where future samples will fall, but only where future predictions will fall. The prediction interval really only gives information about the next sampled point. A tolerance interval should be used if more than one new point is sampled and you want to estimate where they fall. Note that the prediction interval is wider than the confidence interval, and that the tolerance interval (usually, because this depends on both  $\alpha$  and  $P$ ) is wider than the prediction interval.

## PLS plots

MODDE can present score and loading plots when the model was fitted with PLS regression.

### Loadings - WC plots

Plots of the  $\mathbf{X}$  and  $\mathbf{Y}$  weights ( $\mathbf{w}$  and  $\mathbf{c}$ , respectively), with one PLS component against another, illustrate how the  $\mathbf{X}$  variables influence the  $\mathbf{Y}$  variables, and the correlation structure between  $\mathbf{X}$  and  $\mathbf{Y}$  variables. These plots simplify the understanding of how the responses vary in relation to each other and which ones provide similar information.

## Scores - TT, UU, and TU plots

The plots of the  $X$  and  $Y$  scores (e.g.,  $t_1$  vs.  $t_2$ , and  $u_1$  vs.  $u_2$ ), can be interpreted as windows into the  $X$  and  $Y$  spaces, respectively, and illustrate how the design points (experimental conditions,  $X$ ) and responses profiles ( $Y$ ) are situated with respect to each other. These plots illustrate the presence of outliers, groups, and other patterns in the data.

The  $X$ - $Y$  score plots, e.g.,  $t_1$  vs.  $u_1$ ,  $t_2$  vs.  $u_2$ , etc. show:

- the relationship between  $X$  and  $Y$ ,
- the degree of fit (good fit corresponds to small scatter around the straight line),
- indications of curvature, and
- outliers.

---

## PLS regression coefficients

The PLS regression model computes regression coefficients,  $B_m$ , for each response,  $Y_m$ , expressed as a function of the  $X$  variables according to the assumed polynomial model (i.e., linear, interaction, quadratic, etc.). These coefficients are (columns of  $B$ ) computed as:

$$B = W(P'W)^{-1} C',$$

where  $W$  and  $C$  are  $p$ -by- $A$  and  $m$ -by- $A$  matrices, respectively, whose columns are the vectors  $w_a$  and  $c_a$  and  $p$  is the number of variables in  $X$ ,  $m$  is the number of variables in  $Y$  and  $A$  is the number of components.

---

## Box-Cox plot (only MLR)

If the response values vary more than an order of magnitude in the experimental domain, a transformation is often recommended.

When the  $Y$  variables are positive, a useful transformation is the Box-Cox transformation

$$y_{\lambda,i} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(y_i) & \text{if } \lambda = 0, \end{cases} \equiv \begin{cases} y_i^\lambda & \text{if } \lambda \neq 0, \\ \ln(y_i) & \text{if } \lambda = 0, \end{cases}$$

where the second equality (with three horizontal lines) indicates equivalence under linear transformation.

The Box-Cox transformation computes a transformation of the response in order to achieve:

- A simple and parsimonious model,
- An approximately constant model error variance,
- An approximately normal model error distribution.

The parameter  $\lambda$  is found by maximum likelihood estimation of the regression of the transformed  $y_{\lambda,i}$  on the independent variables in  $X$ . The log-likelihood function is defined as

$$L_{max}(\lambda) = -\frac{1}{2} n \ln \left( \frac{SS_{res,\lambda}}{n} \right) + (\lambda - 1) \sum_{i=1}^n \ln(y_i).$$

MODDE computes the value, denoted  $\lambda^*$ , that maximizes  $L_{max}(\lambda)$  in the interval  $-2 \leq \lambda \leq 2$ , and presents a  $100(1 - \alpha) \%$  confidence interval of the maximum by

$$L_{max}(\lambda^*) - L_{max}(\lambda) < \frac{1}{2} \chi_{DF_\lambda}^2(\alpha),$$

where  $DF_\lambda$  is the degrees of freedom in  $\lambda$ .

The value of  $\lambda$  that maximizes  $L_{max}(\lambda)$  in the interval, *i.e.*,  $\lambda^*$ , is the value that gives a transformation of the response that gives the best fit of the model. This is the maximum likelihood estimation of  $\lambda$ .

See Box and Cox (1964) for more details.

The **Box-Cox Plot** displays  $\lambda$  vs.  $L_{max}(\lambda)$ .

---

*Note: The Box-Cox plot is not available for PLS regression models*

---

### References

Draper, Norman R. and Smith, Harry. *Applied Regression Analysis*, second edition. John Wiley & Sons, New York, 1981.

Box, George E. P. and Draper, Norman R. *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York, 1987.

Box, George E. P. and Cox, David R. An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B (Methodological)*, 26 (2), 211-252, 1964.

---

## Mixture data in MODDE

Mixture data in MODDE is available as either mixture factors only or as a combination of process and mixture factors.

---

## Mixture factors only, Model types

When the investigation only includes mixture factors, there are three model types (or forms) available:

- Slack variable model
- Cox reference mixture model, and
- Scheffé model.

## Slack variable model

When you define a mixture factor as filler, MODDE generates the *slack variable* model by omitting the filler factor from the model. The model is generated according to the selected objective and is treated as a non-mixture model. You may select MLR or PLS regression to fit the model, just as with ordinary process factors. With MLR the factors will be orthogonally scaled.

## Cox reference mixture model

When all mixture factors are formulation factors, MODDE generates, by default, the Cox reference mixture model, and therefore the complete polynomial model linear or quadratic. MODDE also supports a special cubic and a full cubic model.

## Scheffé model

You may select to fit a Scheffé model. MODDE expresses the mixture model in the Scheffé form. The full cubic model is not supported in a Scheffé model.

# Mixture factors only, fitting a model

## Analysis and fit method

MODDE's default fit method with mixture factors is PLS regression, and the model form is the Cox reference mixture model. All factors, including mixture factors, are scaled by default to unit variance prior to the model fitting. This is also done with mixture factors that have been transformed to *pseudo components*.

## Cox reference mixture model

The Cox reference mixture model can be fitted by MLR (when obeying the mixture hierarchy) or by PLS regression.

The coefficients in the Cox reference mixture model are meaningful and easy to interpret. They represent the change in the response when going from a standard reference mixture ( $s$ , with coordinates  $s_k$ , for  $k = 1, \dots, K$ ) to the vertex  $k$  of the simplex. In other words, when a component (a factor)  $x_k$  changes by an amount  $\Delta x_{ik}$ , the change in the response is proportional to  $\beta_k$ , the regression coefficient for the corresponding factor  $k$ , such that

$$\Delta y_i = \frac{\beta_k \Delta x_{ik}}{1 - s_k},$$

where  $\Delta y_i$  is the change in the response.

Terms of second or higher degrees are interpreted as in regular process variable models. The presence of square terms, though they are not independent, facilitates the interpretation of quadratic behavior, or departure from non-linear blending. The constant term is the value of the response at the standard reference mixture.

## Changing the standard reference mixture

On the **Design** tab, click **Reference mixture** to change the coordinates,  $s_k$ , of the standard reference mixture. By default, MODDE selects the centroid of the constrained region as the reference mixture.

### **Mixture Hierarchy with the Cox reference mixture model**

By default, all Cox reference mixture models (linear and quadratic) obey a "mixture hierarchy". First we give the notation

$$y_i = \sum_j \beta_j x_{ij} + \sum_{j < k} \beta_{jk} x_{ij} x_{ik},$$

for the regression model. *I.e.*, that the regression coefficients corresponding to quadratic terms have double indices. Then, the group of terms are constrained as follows for linear models;

$$\sum_{k=1}^K \beta_k s_k = 0,$$

with the additional constraint for quadratic models;

$$\sum_{k=1}^K \beta_{jk} s_k = 0,$$

for  $j = 1, \dots, K$ , where  $s_k$  is coordinate  $k$  in the standard reference mixture.

The groups are treated as a unit, and terms can therefore not be removed individually.

When these constraints are imposed on the regression coefficients, the change in the response is proportional to the change in a factor, such that

$$\Delta y_i = \frac{\beta_k \Delta x_{ik}}{1 - s_k},$$

when factor  $k$  is changed from  $x_{ik}$  to  $x_{ik} + \Delta x_{ik}$ .

### **Removing individual terms**

If you want to remove terms individually, as with regular process models, you must clear the **Enforce the mixture model hierarchy** check box in the **Edit Model** dialog box. This dialog box can be opened by clicking **Edit model** on the **Home** tab.

When the mixture hierarchy is not enforced (this includes *e.g.* cubic models), the Cox reference mixture model can only be fitted by PLS regression. The coefficients are the regular PLS regression coefficients and they are not re-expressed relative to a stated standard reference mixture.

---

*Note: Linear mixture terms cannot be excluded. Also, the model hierarchy is always enforced; a term cannot be removed if a higher order term containing that factor is still in the model.*

---

### **ANOVA with the Cox reference mixture model**

In the ANOVA table, the degrees of freedom for regression are the real degrees of freedom, taking into account the mixture constraints. These are the same as for the corresponding slack variable model.

### **Screening plots**

When the objective is to find the components' effects on the response, the coefficients of the Cox reference linear model are directly proportional to the Cox effects. The Cox effect is the change in the response when component  $k$  varies from 0 to 1 along the Cox axis. That is, the axis joining the reference point and the  $k$ th vertex.

### Effect plot

The **Effect Plot** displays the adjusted Cox effects. The adjusted effect of component  $k$  is

$$k = r_k * t_k$$

$$r_k = U_k - L_k$$

$$t_k = b_k / (T - s_k)$$

where  $r_k$  is the range of the  $k$ th factor,  $t_k$  is the total Cox effect,  $T$  is the mixture total (in most cases  $T = 1$ ),  $b_k$  is the unscaled, uncentered coefficient and  $s_k$  is the value of the factor at the reference mixture.

The Effect Plot is only available for screening designs using the Cox reference mixture model.

### Main effect plot

For a selected mixture factor,  $X_k$ , the Main effect plot displays the predicted change in the response when  $X_k$  varies from its *low* to its *high* level, adjusted for all other mixture factors, that is, by default, the relative amounts of all other mixture factors are kept in the same proportion as in the standard reference mixture (MODDE does not check if the other mixture factors are kept within their ranges). For example, if the main effect of the mixture factor  $X_i$  is being displayed, when  $X_i$  takes on the value  $\tilde{X}_i$ , the other mixture factors are assigned the values

$$X_j = (T - \tilde{X}_i) \cdot \frac{s_j}{T - s_i},$$

for  $j \neq i$ , where  $s_k$  are the coordinates of the standard reference mixture. The standard reference mixture is the one used in the model.

You can change the default and choose to have the other mixture factors kept in the same proportion as their ranges (this ensures that no extrapolation is performed).

### Interaction Plot

The interaction plot is not available when only mixture factors are present.

## MLR and the Cox reference mixture model

In MODDE it is only possible to fit Cox reference mixture models (linear or quadratic) with MLR when they obey the mixture hierarchy. When fitting the model with MLR the mixture factors are not scaled and they are only transformed to *pseudo components* when the region is regular, i.e., when all axes have equal length. The model is fitted by imposing the mixture hierarchy.

## PLS and the Cox reference mixture

When fitting a model with PLS regression, the standard reference mixture is not stated a priori as with MLR, and no constraints on the coefficients are explicitly imposed. PLS regression fits the mixture models, and handles collinearities. The PLS regression coefficients can be interpreted as in the Cox reference mixture model, relative to a reference mixture resulting from the projection, but not explicitly stated.

### **Expressing PLS coefficients relative to an explicit standard reference mixture**

With linear and quadratic models that obey the mixture hierarchy, it is easy to re-express the PLS regression coefficients relative to a stated reference mixture with coordinates  $s_k$ .

---

*Note: The  $s_k$  are expressed in pseudo components, if a pseudo component transformation was used.*

---

The mixture hierarchy is imposed on the uncentered, unscaled coefficients in the fitted PLS regression model.

The scaled and centered coefficients are recomputed afterwards.

---

*Note: In MODDE, linear and quadratic models obeying the mixture hierarchy can only be removed as a group, and not individually. By default, PLS regression coefficients are always expressed relative to a stated standard reference mixture.*

---

With models that contain third order terms (cubic terms), or that disobey the mixture hierarchy, no constraints are imposed on the PLS regression coefficients. The coefficients are in this case the regular PLS coefficients and the reference mixture is implicit.

### **Scheffé models derived from the Cox reference mixture model**

With the linear or quadratic Cox reference mixture model, the unscaled coefficients can be re-expressed as those in a Scheffé model. The following relationships hold.

#### **Linear**

$$b_{\text{Scheffé},k} = b_{\text{Cox},0} + b_{\text{Cox},k}$$

#### **Quadratic**

$$b_{\text{Scheffé},k} = b_{\text{Cox},0} + b_{\text{Cox},k} + b_{\text{Cox},kk}$$

$$b_{\text{Scheffé},kj} = b_{\text{Cox},kj} - b_{\text{Cox},kk} - b_{\text{Cox},jj}$$

### **Scheffé models**

The Scheffé models are only fitted with MLR and among the effect plots only the **Main Effect Plot** is available.

Scheffé models are only available for investigations where there are only mixture factors.

#### **ANOVA**

The degrees of freedom in the ANOVA table for a Scheffé model are computed in the same way as with the slack variable model (see Marquardt and Snee (1974)).

## Using the model

### **Prediction plot**

The **Prediction Plot** is available for all objectives and all models. As with process factors, this plot displays a spline representing the variation of the fitted function when the selected mixture factor varies over its range, adjusted for the other factors. As with the main effect plot, this means that the relative amounts of all other mixture factors are kept in the same proportion as in the standard reference mixture. If no standard reference mixture is specified, the centroid of the constrained region is used by default.

### **Mixture contour plot**

Trilinear contour plots are available with mixture factors but no response surface plots.

### **References**

Marquardt, Donald W. and Snee, Ronald D. Test Statistics for Mixture Models. *Technometrics*, 16 (4), 533-537, 1974.

## Process and mixture factors

When both process and mixture factors are present, you can select to treat them as one model, or you can specify separate models for the mixture factors and the process factors.

With both mixture and process factors, the only model form available is the Cox reference mixture model.

When the model obeys the mixture hierarchy, the PLS regression coefficients are expressed relative to a stated standard reference mixture.

If  $\gamma$  is a parameter vector that contains the coefficients of the interactions between the process and mixture factors, then

$$\sum_{k=1}^q \gamma_k s_k = 0.$$

*Note: When the model contains terms of the third order, or it contains qualitative and formulation factors, the PLS regression coefficients are not adjusted relative to a stated standard mixture.*

### **MODDE plots**

All plots in MODDE are available when you have both mixture and process factors. For both Main Effect Plots and for Prediction Plots, when you select to vary a process factor, all of the mixture factors are set to the values of the standard reference mixture. When you select to vary a mixture factor, process factors are set to their average and the other mixture factors are kept in the same proportion as in the standard reference mixture or their ranges.

---

## Optimizer

MODDE's optimizer uses the Nelder-Mead method (also known as the downhill simplex method) on the fitted response surface (Nelder and Mead, 1965), in order to minimize an *overall desirability* function that combines the *individual desirability* of each response.

### References

Nelder, John and Mead, Roger. A simplex method for function minimization. *The Computer Journal*, 7 (4), 308-313, 1965.

### Desirability

The individual desirability functions are highly flexible functions, whose shape can be manipulated by adjusting a few parameters. Depending on the form of the individual desirability functions, the Optimizer can be configured to suit various optimization objectives.

The optimizer can be set up for different objectives:

1. *Limit optimization* – where the objective is to reach a solution in which the response is within the specification limits (Min and Max limits). This is the default approach in MODDE.
2. *Target optimization* – where the objective is to reach a solution in which the response is as close to target as possible. For the target optimization to work properly, it is necessary that the response can be optimized close to or on target; otherwise the search may end up with an unacceptable solution.
3. *Custom optimization* – user defined customization of the Target optimization.
4. *Focus optimization* – where the objective is to favor one or several responses over the others; accomplished by manipulating the individual weights.
5. *Robust setpoint* – where the most robust setpoint is found. Depends on the existence of a solution based on objectives 1-4.

MODDE supports two types of desirability functions: a quadratic spline function and an exponential spline function. The shape of these splines is influenced by a few parameters, input by the user.

### Limit optimization

In *Limit optimization*, the desirability function is defined as

$$d_k = 100(e^{\lambda_k g(y_k)} - 1),$$

where

$$g(y_k) = 100 \frac{y_k - P_k}{T_k - P_k},$$

for every response  $y_k$ , for  $k = 1, \dots, m$ , and where  $T_k$  is the user-defined target for response  $y_k$  and  $P_k$  is the maximum of the worst response value computed from the starting simplex and  $1.1L_k$ , in which  $L_k$  is a user-defined worst acceptable response value for response  $y_k$ . The value of  $P_k$  is never closer to the target than  $L_k$ .

When a response is to be maximized,  $L_k$  is the smallest acceptable value: when the response is to be minimized  $L_k$  is the largest acceptable value. When the response is to be on target, the user gives the smallest and largest acceptable values.

When the response is to be minimized,  $T_k$  must be less than  $L_k$ . When the response is to be maximized  $T_k$  must be greater than  $L_k$ .

For responses to be on target the user must supply lower and upper limits such that

$$L_{k,lower} < T_k < L_{k,upper}.$$

$L_k$  is generated internally if not supplied by the user.

The  $\lambda_k$  is a scaling parameter computed as

$$\lambda_k = -\frac{\ln\left(\frac{100}{100 - Limit_k}\right)}{100 \frac{L_k - P_k}{T_k - P_k}},$$

where  $Limit_k = 90 + 80\log_{10}(w_k)$  and  $w_k$ , for  $k = 1, \dots, m$ , are weights assigned to each response by the user. The weights are  $0.1 \leq w_k \leq 1$  and  $w_k = 1$  is the default for all  $k = 1, \dots, m$ .

When the user wants the response to be on target,  $L_{k,upper}$  is used in the calculation of  $\lambda_k$  when  $y_k > T_k$  and  $L_{k,lower}$  is used in the calculation when  $y_k < T_k$ .

This definition of  $\lambda_k$  makes  $d_k = -Limit_k$  when  $y_k = L_k$  and makes the desirability function range from 0 to -100 (a value of -100 can only be reached when  $y_k$  gets close to  $T_k$ ). Note, though, that the plots rescale the desirability functions to range from -1 to 0.

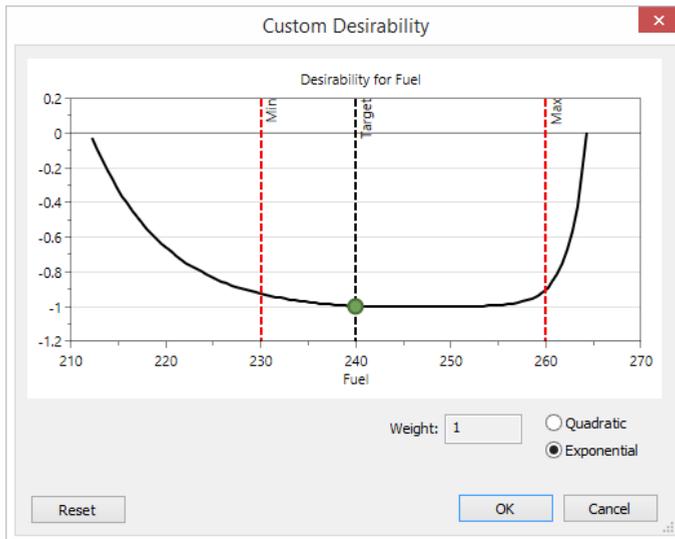


Figure. A plot that shows the default desirability for the exponential desirability function (*Limit desirability*).

**Target optimization**

In *Target optimization*, the desirability function is defined as

$$d_k = 100(a + b \cdot \text{Dist}_{sc} + c \cdot \text{Dist}_{sc}^2 - 1),$$

when the prediction is outside of the limit interval, and

$$d_k = 100 \left( a + \left( \frac{\text{Dist}_{sc}}{2} + \text{Dist}_{sc}^2 \right) \cdot CC \right) - 100$$

when the prediction is inside the limit interval and where  $a$ ,  $b$  and  $c$  are adjustable parameters, (the default settings are  $a = 0$ ,  $b = 0.2$  and  $c = 0.8$ .)

$$\text{Dist}_{sc} = \begin{cases} \frac{T_k - \text{Dist}}{T_k - PMin} & \text{if } y_k < T_k \\ \frac{\text{Dist} - T_k}{PMax - T_k} & \text{otherwise} \end{cases}$$

represents a scaled (between 0 and 1) distance between the predicted value and  $T_k$ ,  $CC$  is a correction factor that makes the transition from outside to inside the limit interval continuous; and  $PMin$  and  $PMax$  are the highest and lowest predicted values for the specific response, respectively.

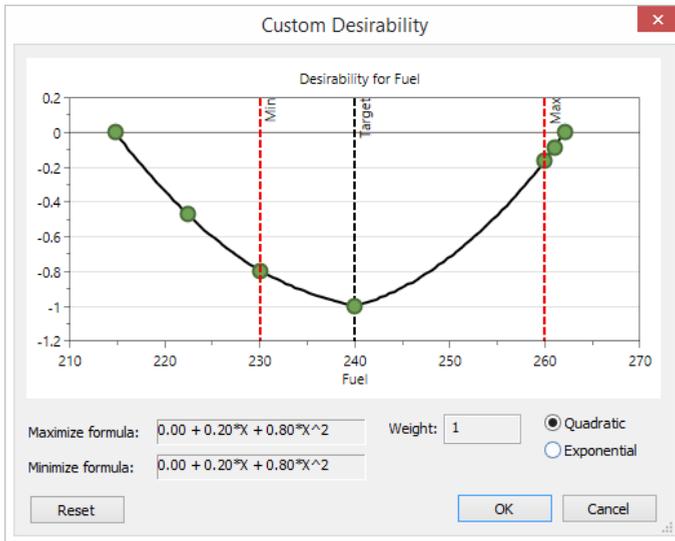


Figure. Plot showing the default desirability function for Target optimization.

**Custom optimization**

*Custom optimization* means a user-defined customization of the *Target optimization* function. This objective is used when the shape of the quadratic spline desirability function is changed by the user. The shape of the spline can be manipulated interactively by dragging and releasing the blue dots in the desirability plot. A customization of the desirability function may be warranted if there is a certain functional relationship that is known to exist between factors and responses, or if there is deemed an advantage trying to approach the target response value from either the high (max) side or the low (min) side.

### Focus optimization

*Focus optimization* can also, like *Custom optimization*, be seen as an extension of the *Target optimization* function. The rationale for the Focus optimization objective is that initial searches might have indicated that it is rather difficult to obtain a stable solution for a case that involves multiple response variables and perhaps also with partly conflicting goals. In such a case, it might be warranted to make prioritizations among the responses by modifying their weights in the optimization. Technically, in MODDE such changes in priorities are accomplished by down-weighting the least important response variables. Thus, if the weights are set differently for different responses, responses with higher weights take priority in the search for a solution inside the specifications. The overall optimization criterion is used to reach the lowest value of the individual desirability functions.

### Overall desirability

The *overall desirability* function is the mean of the individual desirability functions for all responses. It quantifies how far the predicted results are from their corresponding target values. The overall desirability is defined as

$$f(\mathbf{ds}) = \frac{1}{m} \sum_{k=1}^m d_k,$$

where  $\mathbf{ds} = (d_1, \dots, d_m)$  are the individual desirability functions.

### Overall distance to target

The *Overall distance to target*,  $D$ , is computed as

$$D = \log_{10} \left( \frac{\sum_{k=1}^m w_k \left( \frac{\hat{y}_k - T_k}{T_k - L_k} \right)^2}{m} \right),$$

where the  $w_k$  are the user-defined weights on the individual desirability functions, the  $\hat{y}_k$  are predicted response values, the  $T_k$  are user-defined targets,  $L_k$  are user-defined worst acceptable response values and  $m$  is the number of responses.

The overall distance to target is defined as  $D = -10$  when all responses have reached their targets,  $T_k$ .

The value of  $D$  is not used in the optimization procedure, but is displayed as  $\log(D)$  in the run list.

## Starting simplexes

The optimizer will start simplexes from a number of starting positions that are selected as follows:

- a. The (up to) 6 most important factors are used to generate the full or fractional factorial design for start points as follows:

Factors	Start point design
1 continuous	3 levels
2 continuous	$2^2$
3 continuous	$2^3$
4 continuous	$2^4$
5 continuous	$2^{5-1}$
6 continuous	$2^{6-2}$

- b. The last 4 runs are the 3 runs from the worksheet with the "best" predictions *i.e.* the lowest  $\log(D)$  plus a center point, where  $D$  is the overall distance to target.

The start points of continuous factors are 20 % smaller than the original design factor range. The start settings of qualitative factors are random.

The user cannot modify these start runs but can add their own runs.

Each simplex is generated from the start point by adding an additional point for each factor with an offset of 20 % of the distance from the center to the maximum value, the other factors being kept at the same values. A check is made that all runs are within the defined experimental region.

## Sensitivity analysis

Sensitivity analysis measures the probability for a setpoint (the optimum) to fail for each individual response or all combined. The sensitivity analysis is performed by a Monte Carlo simulation and is reported as Probability of failure in percent or *defects per million opportunities outside of specifications*, DPMO.

The regression model is defined as

$$y = x_1 \beta_1 + x_2 \beta_2 + \dots + x_p \beta_p + \varepsilon,$$

where we predict  $y$  as

$$\hat{y} = x_1 \beta_1 + x_2 \beta_2 + \dots + x_p \beta_p.$$

The Monte Carlo procedure includes by default the model prediction error and can also add an estimated operational factor precision by disturbing the factors,  $x_i$  for  $i = 1, 2, \dots, p$ . This is done by adding *e.g.* normally distributed random noise to the factors; either the default  $\pm 5$  % of the range of each factor, or a user specified factor precision. The prediction error is assumed to follow a Student's  $t$ -distribution with standard deviation defined by a confidence or a prediction interval surrounding the predicted mean, or it is assumed to follow a normal distribution with standard deviation defined by a tolerance interval surrounding the predicted mean. For details about intervals, see the Interval estimates subsection in the Statistical appendix

The Monte Carlo procedure samples points from this distribution and counts the number of samples that falls outside of the user specified bounds for the response. This result in an estimate of the total Probability of failure, in DPMO or percent, for all active responses. If the Probability of failure is zero or *very low*, this means that the proposed factor settings represent a robust point of operations.

The user can modify the factor precision in the **Factor Definition** dialog box.

## Factor contribution

The **Factor contribution**, displayed in the **Selected setpoint** page in the **Optimizer**, is defined as the relative contribution of a factor on the optimizer result. The value can range from 0 to 100. A high factor contribution affects the result setting more than a low factor contribution. The calculation of the contribution uses by default 5 % of the factor definition range as the range around the selected setpoint. The factor contribution range can be changed in the **Optimizer settings** dialog box.

### Factor contribution calculation

The details of the factor contribution calculations are listed in this section.

#### Terminology

$F_{max}$  is the maximal factor value according to design specification.

$F_{min}$  is the minimal factor value according to design specification.

$F_{opt}$  is the factor setting at the found optimum.

$F_{low}$  is the factor value below the optimum, see equation in step 2.

$F_{high}$  is the factor value above the optimum, see equation in step 2.

$C_{Range}$  is the contribution range (default is 0.05).

$RS_{max}$  is the maximum limit for response value specified in the optimizer.

$RS_{min}$  is the minimum limit for response value specified in the optimizer.

$Rd_{max}$  is the largest response value from worksheet.

$Rd_{min}$  is the smallest response value from worksheet.

$R_{Range}$  is the response range.

$R_{opt}$  is the response values at the found optimum.

$R^+$  is the response value when factor value is above the optimum, see step 2.

$R^-$  is the response value when factor value is below the optimum, see step 2.

$F_{Contr}$  is the factor contribution.

### Step 1 - Computing the response ranges

The following equations are used in the computation of the range for each response.

If the **Weight factor contribution against distance to response limit** check box in the **Optimizer Properties** dialog box is selected,

$$R_{Range} = \min(\text{abs}(R_{opt} - RS_{min}), \text{abs}(RS_{max} - R_{opt})),$$

else

$$R_{Range} = Rd_{max} - Rd_{min},$$

**Step 2 - Computing of factor contribution for each factor/response combination**

The following equations are used to compute the factor contribution for the available factor/response combinations.

For every factor  $j$ :

$$F_{low} = F_{opt} - C_{Range} * (F_{max} - F_{min})$$

$$F_{high} = F_{opt} + C_{Range} * (F_{max} - F_{min})$$

$R_j^-$  is the predicted response using  $F_{low}$  for factor  $j$  (all other factors are kept at optimal settings)

$R_j^+$  is the predicted response using  $F_{high}$  for factor  $j$  (all other factors are kept at optimal settings)

For every response  $k$ :

$$F_{contr,j,k} = \frac{\text{abs}(R_j^+ - R_{opt}) + \text{abs}(R_{opt} - R_j^-)}{2 \cdot C_{range,j} \cdot R_{range,k}}$$

**Step 3 - Computing the factor contributions over all responses**

To compute the factor contributions over all responses for each factor,  $j$ , we use

$$F_{contr,j} = \sum_k F_{contr,j,k}$$

**Step 4 - Normalization of factor contributions**

To normalize the factor contributions, we use

$$F_{contr,j} = 100 \cdot \frac{F_{contr,j}}{\sum_i F_{contr,i}}$$

**Orthogonal blocking**

When you cannot perform all experiments in a homogeneous way, randomizing the run order of the experiments may not be sufficient in order to deal with the extraneous sources of high variability. You may want to run the experiments in homogeneous groups, i.e., blocks, in such a way that the external source of variability does not influence the effects of the factors.

For example, if you are using a full factorial design with 5 factors and 32 runs, and the batch size of raw material allows you to perform only 8 runs per batch. You may want to run your experiments in 4 blocks, each composed of 8 runs using homogeneous material.

*Orthogonal blocking* divides 32 runs in 4 blocks of 8 runs, each divided so that the difference between the blocks (the raw material) does not affect the estimate of the factors.

MODDE supports *Orthogonal blocking* for the 2 level factorial, fractional factorials, Plackett Burman, CCC, and Box-Behnken designs.

MODDE also supports blocking of D-Optimal designs. These designs are more flexible with respect to the number of blocks and the block size, but the blocks in D-Optimal designs are usually not orthogonal to the main factors. The only restriction with D-Optimal designs is that the number of runs must be a multiple of the block size.

---

*Note: Blocking introduces extra factors in the design, and hence reduces the degrees of freedom of the residuals, and the resolution of the design. You should only block when the extraneous source of variability is high and cannot be dealt with by randomizing the run order.*

---

### **Block interaction**

An interaction between a main effect and a block effect is called a block interaction.

When the design supports interactions between the block effects and the main effects, the **Block interactions** check box is available, in the Select model and design page in the Design Wizard. You can select the check box if you want to add the block interactions to your model.

### **Recoding the blocking factors**

When blocking factors are generated, the blocks are assigned according to the combination of signs of the blocking factors.

For example to generate 4 blocks, the following scheme of signs of the blocking factors is used.

<b>\$B1</b>	<b>\$B2</b>	
–	–	<b>Block 1</b>
+	–	<b>Block 2</b>
–	+	<b>Block 3</b>
+	+	<b>Block 4</b>

The **Design Matrix** is displayed in coded units that include the block factors. The Design Matrix can be accessed via the **Design** tab, in the **Show** group, click **Design matrix**.

When the worksheet is generated, the block factors are recoded and the model is re-parameterized. Rather than keeping the  $d$  blocking factors, MODDE generates one qualitative variable called  $\$BlockV$ , with  $k = 2^d$  levels, denoted  $B1, B2, \dots, Bk$ , for the  $k$  blocks.

### **Inclusions and blocks**

Adding inclusions to a blocked design is not supported, unless the inclusions belong to one of the blocks present.

## **Blocking screening designs**

### **Full and fractional factorial designs**

The block size and the number of blocks of a 2 level factorial design are always powers of 2.

The maximum number of blocks supported by MODDE is 8, with a minimum block size of 4.

The designs are blocked by introducing blocking factors, denoted \$B. There is one blocking factor for 2 blocks, two for 4 blocks and three for 8 blocks. The block effects consist of the effects of the blocking factors and all their interactions.

Hence, with 8 blocks, there are 7 block effects using 7 degrees of freedom. (It is equivalent to adding 7 extra factors to your design).

MODDE selects the generators of these blocking factors to achieve the highest possible pseudo-resolution of the design.

The pseudo-resolution of the design is the resolution of the design when all the block effects (blocking factors and all their interactions) are treated as main effects under the assumption that there are no interactions between blocks and main effects, or blocks and main effects interactions.

### **Blocking Plackett Burman designs**

Plackett Burman designs can only be split into two blocks by introducing one block variable, and using its signs to split the design.

### **Blocking RSM designs**

RSM designs are orthogonally blocked when they fulfill the following two conditions:

1. Each block must be a first order orthogonal block.
2. The fraction of the total sum of squares of each variable contributed by every block must equal the fraction of the total number of runs (observations, experiments) allotted to the block.

### **Central Composite Circumscribed designs**

The Central Composite Circumscribed designs can be split into two blocks, the cube portion and the star portion, and satisfying the two above conditions when  $\alpha$  (the distance of the star points to the center) is equal to

$$\alpha = \sqrt{\frac{k(1 + p_s)}{1 + p_c}},$$

where  $k$  is the number of factors,  $p_s = n_{so}/n_s$  is the proportion of center points in the star block,  $n_{so}$  is the number of center points in the star block,  $p_c = n_{co}/n_c$  is the proportion of center points in the cube block,  $n_{co}$  is the number of center points in the cube block,  $n_c$  is the number of star points runs and  $n_e$  is the number of runs from the cube block.

This is the value of  $\alpha$  implemented in MODDE when you select blocking in a CCC design.

### **Smaller blocks**

The cube block of the CCC designs can be split into further blocks if:

- a. The factorial or the fractional factorial part of the design can be split into orthogonal blocks of pseudo resolution 5.

and

- b. Each one of these blocks have the same number of center points.

### **Box Behnken designs**

These designs can be orthogonally blocked. See details in Box and Behnken (1960) or Box and Draper (1987).

### **Central Composite Face designs**

These designs cannot be blocked.

### **Central Composite Orthogonal designs**

These designs cannot be blocked.

## References

Box, George E. P. and Behnken, Donald. Some new three level designs for the study of quantitative variables. *Technometrics*, 2, 455-475, 1960.

Box, George E. P. and Draper, Norman R. *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York, 1987.

## Blocking D-Optimal designs

MODDE can block D-Optimal designs, but usually the blocks are not orthogonal to the main factors.

The following restrictions apply to blocking D-optimal designs:

1. The blocks must be of equal size.
2. You cannot have interactions between the block factor and the other factors in the model.
3. The selected number of runs of the design must be a multiple of the number of blocks.

MODDE blocks the D-Optimal design by generating a qualitative factor, denoted  $\$BlockV$ , with as many levels as the selected number of blocks. By default, it then selects only balanced designs with respect to the blocking factor.

It is not always possible to generate a balanced D-Optimal design with respect to the blocking factor. In this case you may want to change the model, the number of blocks, or generate an unbalanced design.

## Random versus fixed block factor

You can select to have the block factor treated as a fixed or random effect and the predictions computed accordingly.

Select the block factor as *Fixed* when the external variability can be *set at will* (it is controlled) and the primary objective for blocking is to eliminate that source of variability.

A fixed block can be modeled as a controlled qualitative factor with a limited number of levels. All predictions of the responses and contour plots will be made for a selected block level.

The block is a fixed effect if you have 32 experiments and 8 runs are done on, for instance, each of 4 different machines. There is no other machine available other than the four used, not now, nor in the future.

You may want to have 4 blocks to eliminate the variability introduced by the machines, but all predictions of the responses are made for one of the specific machines.

Set the block factor as *Random* effect when the external variability *cannot* be controlled and set at will, and the primary objective is to make predictions without specifying the block level, and taking into account the external variability.

Since the block level of future experiments is unknown, all predictions of the responses for random block effects are made without specifying the block level. The interval estimations for the responses are increased to account for the uncontrolled external block variability.

The block is a random effect if you have 32 experiments and each block of 8 runs is made with a different batch of, for instance, raw material. Your primary objective is to make predictions for the next unknown batch of raw material.

### ***Analysis with random effects***

When you treat the block factor as a *Random* effect, it is often desirable to investigate the consistency of the factor effects by including in the model all the interactions of the block factor with the main factors, if possible.

A model in MODDE is always fit as a fixed effect model. That is, with the block factor treated as a controlled qualitative variable, even when the blocks are specified as random.

If the random block interaction effects with the main factors are large and significant, the effect of the main factors varies from block to block, and the interval estimates of the prediction will be large due to this uncontrolled variability.

If the random block interaction effects are small and insignificant, the effects of the main factors are consistent from block to block and the uncertainty of the predictions is greatly reduced.

To have a realistic size of the interval estimations, trim the model and remove all insignificant block interaction effects.

If the block factor is specified as a fixed effect, the interactions of the main factors with the block factor are of less interest.

### ***Predictions with random effect***

When the block effect is specified as random, the predictions of the responses are computed without specifying the block level.

MODDE uses the average block level to predict the response, but the interval estimate is increased to take into account the variability of the response due to the different blocks, plus the variability of the response due to uncertainty on the coefficients of the model including all the terms with the block factor.

---

## **Setpoint statistics**

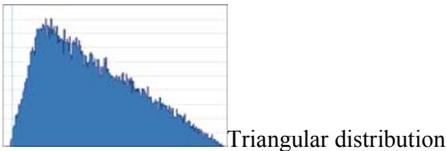
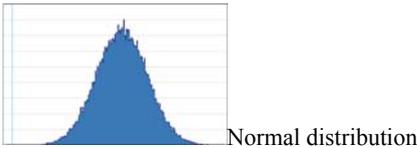
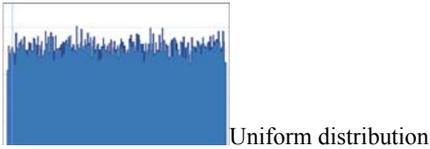
*Setpoint analysis* extends the possible factor ranges from a *setpoint* (the optimum) to the largest possible range where all response predictions are still within the specifications. The range is an estimation of the largest possible deviation that is accepted for the factors at a given setpoint combination.

Predictions in the setpoint analysis are performed by Monte Carlo simulations. The resulting distribution of predictions simulates a real situation with a random combination of factor setting disturbances within a given range.

*Setpoint validation* tests the robustness of the model by disturbing the model a large number of times (a Monte Carlo simulation) in the specified region. The result is shown as the distribution of the random samples, including model prediction errors. The result can be expressed in general statistics as well as capability index,  $C_{pk}$ , or Probability of failure.

## Monte Carlo simulations

The factors may be perturbed by three different distributions, namely *Uniform*, *Normal* and *Triangular* distributions:



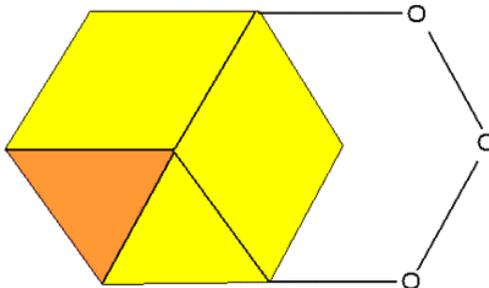
The default is that the randomization follows a normal distribution. In the **Distribution** column in the Factor list you can also select *Target*, in which the interval surrounding the predicted point is zero.

The low and high factor settings are the distribution boundaries. For a normal distribution, 95 % of the distribution is found within the boundaries, by default. The automatic search procedure extends the range for each factor until one or more response limits are exceeded according to the specified Probability of failure limit. The automated search procedure extends symmetrically around the *setpoint*.

### Doehlert designs

The Doehlert designs are quadratic RSM designs with some special properties (buildable and extendable to other factor intervals). They allow the estimation of all main effects, all first order interactions, and all quadratic effects without confounding. They are saturated designs with similar properties to the CCF, CCO and CCC designs. Geometrically they are polyhedrons based on hyper-triangles (simplexes), with a hexagon in the simplest two-factor case.

Doehlert design in 2 factors with 6 runs + center points can be extended to a new design by adding 3 experiments. Usually also one or two new center points are added in the new design (*i.e.*, in the figure the right-most point in the old design).



The Doehlert designs are well suited for the RSM objective with up to 5 or 6 factors (respectively 33 and 45 runs with 3 center points). The intent with these RSM designs is to get a precise model that can be used for optimization and for detailed understanding.

The required number of runs  $N$ , except for replicated center points, for the quadratic Doehlert designs with  $k$  factors is:

$$N = 1 + k + k^2$$

It is recommended to add 3 to 4 center points to these designs.

## Probability of failure and process capability indices

Probability of failure can be expressed in % (default) and DPMO and is used as a stopping criterion in the setpoint analysis/validation. DPMO stands for *defects per million opportunities outside specifications*.

In the **Properties** dialog box of **Setpoint Analysis/Validation** (opened by right-clicking the window and then clicking **Properties**) it is possible to select two stopping criteria, namely total Probability of failure or Probability of failure for the individual response.

The difference comes from the fact that the same point may be found outside of the limits for more than one response. The individual DPMO/percentage may count the same point several times (at most once per response), while the total Probability of failure only counts each point once. Default is individual and total is activated by selecting the **Use total Probability of failure** check box.

See the DPMO and Cpk definition section for details and equations for DPMO, Cp and Cpk.

The relation between  $C_{pk}$ , DPMO and %Outside is summarized in the following table:

Cpk	Probability of failure DPMO	Probability of failure %
0.4	115070	11.51
0.6	35930	3.59
0.8	8198	0.82
1	1350	0.13
1.1	483	0.05
1.2	159	0.02
1.3	48	0.0048
1.4	13	0.0013
1.5	3.40	0.0003
1.6	0.79	7.93328E-05
1.7	0.17	1.69827E-05
1.8	0.03	3.33204E-06
1.9	0.0060	5.99037E-07
2	0.0010	9.86588E-08

## References

Kane, Victor E. Process Capability Indices. *Journal of Quality Technology*, 18 (1), 41-52, 1986.

## DPMO and Cpk definition

### DPMO

The *DPMO* is computed as

$$DPMO = 1\,000\,000H/N_s,$$

where  $H$  is the number of hits outside of the specifications and  $N_s$  is the number of iterations in the simulation.

The precision in the DPMO estimate is related to the number of simulations. The default settings are 20 000 simulations for the acceptable factor range estimation and 50 000 simulations for the final response profile prediction.

We recommend that you increase the number of simulations by a factor 2-5 when you perform the final analysis (e.g., for your documentation).

Since Monte Carlo simulations are random, there will be variation in the predictions between repeated runs. The results are a function of the number of simulation used, where more simulations give better reproducibility.

### Cpk

The *Process Capability Indices* defined below originate from the *Six Sigma* statistics techniques.

The *process potential index*,  $C_p$ , estimates what a process is “capable” of producing if the target is centered between the specification limits under the “natural” tolerance  $6\sigma$ . If  $C_p = 1$ , the process is said to be “capable”. The computation of  $C_p$  assumes that the results are approximately normally distributed. The process potential index is computed as

$$C_p = \frac{USL - LSL}{6\sigma},$$

where  $USL$  is the upper specification limit (user input),  $LSL$  is the lower specification limit (user input) and  $\sigma$  is the estimated standard deviation for predictions.

The *process performance index*,  $C_{pk}$ , is another way of expressing the probability of obtaining results outside the specification limits, but does not assume that the target is centered between the specification limits. The  $C_{pk}$  is defined as

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right),$$

where  $\mu$  is the predicted average of the simulation. The value is  $C_{pk} < 0$  if the simulation mean falls outside of the user supplied limits. The computation of  $C_{pk}$  assumes that the simulation results are approximately normally distributed.

The *process deviation index*,  $k$ , estimates the amount of deviation of the process mean from the target. It is computed as

$$k = \frac{|T - \mu|}{\min(T - LSL, USL - T)},$$

where  $T$  is the target value.

### References

Kane, Victor E. Process Capability Indices. *Journal of Quality Technology*, 18 (1), 41-52, 1986.

## Predictions including model error

Predictions from the Monte Carlo simulations include the model error by default. You can select to not include the model error in the property page.

The regression model is defined as

$$y = f(x) + \varepsilon,$$

and predictions are computed as

$$\hat{y} = f(x).$$

The Monte Carlo procedure is described in detail in the Sensitivity analysis subsection earlier in this appendix, and restated here slightly more formally for completeness:

1. Perturb each variable,  $x_j$ , for  $j = 1, \dots, p$ , such that  $\tilde{x}_j \sim \mathcal{D}(x_j, \sigma_j^2)$  where  $D$  is the distribution selected to perturb the variables (e.g., normal, triangular or uniform) in which  $x_j$  is the mean and  $\sigma_j^2$  is a spread parameter (e.g., the variance) computed from user input.
2. Sample a random point from a Student's  $t$  distribution with mean  $\tilde{\mathbf{x}}' \hat{\boldsymbol{\beta}}$  standard deviation  $I(\tilde{\mathbf{x}})/t_{\alpha/2, n-p}$  and  $n-p$  degrees of freedom, where  $I(\mathbf{x})$  is the interval width (e.g., a confidence, prediction or tolerance interval) at the point  $\mathbf{x}$ .

Repeat these steps many times to obtain the distribution of the prediction under perturbation of the variables.

---

*Note: The above is true for PLS regression and MLR models, but PLS regression models with condition numbers greater than 3000 and non-hierarchical mixture models will display an interval computed directly from the standard error of prediction.*

---

## General subset designs algorithm and principles

The algorithm that we developed generates designs by utilizing a number of well-known combinatorial structures like orthogonal arrays and Latin squares. In brief, it transforms the original design space into symmetrical/hypercube space where optimal design planes are generated, to be later transformed back into the original space. The algorithm was implemented in MODDE 12.

The steps comprising the algorithm are presented schematically in Figure 1 and summarized below:

1. Choose  $n$ -dimensional design space ( $n$  – number of studied factors) with  $K$  runs in its candidate set ( $K$  – levels of each factor) and a reduced design  $D_p$  so that the final design matrix  $D$  has dimensions  $m \times n$ , where  $m = K/p$ .
2. Decompose factors into  $p$  sets.
3. Generate  $p$  mapping matrices  $M$ , where  $M = 2^{-(p, n, p^{n-3})}$  orthogonal arrays (OA) from Latin Squares (separate algorithm, see below); the purpose of the matrix  $M_i$ ,  $i=1, 2, \dots, p$  is to map from the hypercube space to the original space.
4. Reduce  $M$  to all rows mapping to non-empty sets
5. Map each row of  $M$  to the decomposed factor sets and make a full factorization of the elements in the active sets:  $M_i$ .
6. The final design is produced by concatenating the mappings  $M_i$  from step 5.

The algorithm for generation of orthogonal arrays from Latin Squares can be summarized as follows:

1. Select a desired number of factors  $n$  and a level of reduction  $p$ .
2. Start with a set  $X$  of cardinality  $p$  and assign each value to a matrix  $A_i, i \in [1, p]$ .
3. Create a Latin square of size  $p, L(p)$ ; denote each row  $i \in [1, p]$  of  $L(p)$  by  $L_i$ .
4. For all matrices  $A_i, i \in [1, p]$ , and for all  $j \in [1, p]$  elements in  $X$  let  $A_{ij} = (X_j A_y)$  where  $y = L_i(j)$ .
5. Return to step 3 unless  $A_i$  has  $n$  columns, then end algorithm.

An example of the Generalized subset designs for the three factors at 2, 3 and 4 levels respectively (24 possible experimental combinations), where the full design was reduced to 3 unique sets of 8 combinations, is presented in Figure 2. Each design set has the best possible balance of combinations, any combination of sets gives the best possible design and no settings are duplicated.

The algorithm allows all types of factor combinations to be reduced to design subsets that are balanced and uniformly distributed in space. The resulting design sets are less optimal when the ranges of the factors settings are large and the factor settings are prime numbers. These two problems need to be addressed with the future improvement of the algorithm.

Decomposition into several subsets typical  $> 10$  will for design matrixes with many initial combination  $> 250$  become a challenge. The algorithm will do random starts in the search for a solution and might require a significantly longer search time specifically for large subset decomposition. The resampling function allow setting focus and time for the resampling of a specific subset decomposition.

More:

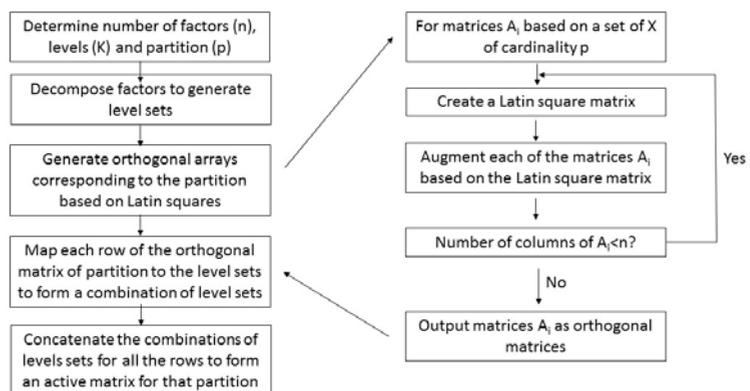


Figure 1. Schematic representation of the steps comprising our algorithm.

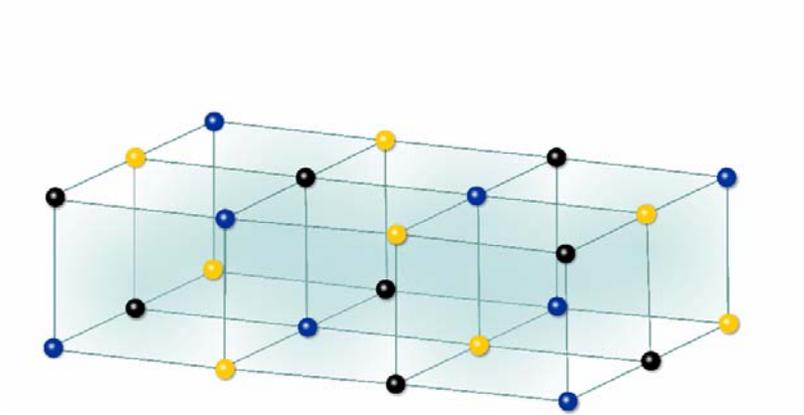


Figure 2. The perfect complementary designs for the three factors at 2, 3 and 4 levels used in the stability study, where the full design was reduced to 33% at time points 3, 6 and 9 months; each design set is indicated by a different color: blue – 3 months, yellow – 6 months, black – 9 months. Each design set has the best possible balance of combinations, any combination of sets gives the best possible design and no settings are duplicated.

# Design appendix

---

## Designs for process factors

### Screening designs

Screening designs are used in the early stages of an investigation to find which factors are important and if it is necessary to modify their ranges.

All screening designs support linear models and some support interaction models.

The designs you can select with *interaction models*, *i.e.* all interactions, are the full factorial, the fractional factorial, Rechtschaffner, and D-Optimal designs of resolution V. Fractional factorials of resolution V are supported for up to 12 factors.

MODDE supports the following screening designs:

### **Full factorial designs**

Full factorial designs can be created at 2 or more levels.

These designs comprise all the possible combinations of the factor levels. For  $p$  factors at 2 levels you need  $N=2^p$  runs. Full factorial designs are orthogonal (balanced) designs. Hence, the estimated effect of a factor is independent of the effects of all other factors.

Full factorials with one or more factors at more than two levels are called Full Factorial Mixed.

### **Fractional factorial designs**

Fractional factorial designs are 2 level designs with resolution III, IV, V or more.

These designs are balanced subsets (fractions) of the full factorials. The resolution of the design depends on the size of the subset, *i.e.* the number of runs selected. The possible resolutions are:

- Resolution III designs where main effects are confounded with 2 factor interactions.
- Resolution IV designs where two factor interactions are confounded with each other.
- Resolution V designs where main effects and all two-factor interactions are clear of each other (unconfounded). MODDE supports resolution V designs.

With both resolution III and IV designs, you can only select the linear model. You may edit the model and enter selected interactions. In that case, you may have to edit the generators of the design.

With resolution V designs, MODDE generates the interaction model.

The default generators used by MODDE for fractional factorial designs are those recommended by Box, Hunter and Hunter (page 410). You may edit and change the generators in the **Generators** dialog available on the **Design** tab and by clicking **Settings** on the **Select model and design** page in the **Design Wizard**. When you update the confounding, MODDE will warn you if some of the effects in your model are confounded with each other, i.e. if your model is singular.

### ***Three level fractional factorial designs***

Fractional factorial designs at three levels are fractional factorial designs from the Graeco-Latin square family. The available designs are:

- L9: design with up to 4 factors at three levels.
- L27: design with up to 13 factors at three levels.
- L36: design with 5 to 13 factors at three levels.
- L18 is called mixed as it has one factor at 2 levels and up to 7 factors at three levels.

With these 3 level designs MODDE (objective = screening) lets you select only the *linear* model, because these designs do not support interactions. In **Edit model** you may edit the model and include square terms.

## **Generalized Subset Designs**

The generalized subset designs, GSD, is a new entry in MODDE providing a possibility to accomplish reduced designs even when handling multiple multilevel factors. This design setup generates a series of reduced designs, subsets, that are logically linked, such that, when combined, all subsets will add up to a full multilevel multifactorial design where all factor combinations are encoded by the global design. Conceptually, the output of GSD is similar to how two-level fractional factorial designs represent complementary reductions of two-level full factorial designs.

For more, see the Generalized Subset Designs section in the Generalized subset designs appendix.

### ***Reduced combinatorial designs***

Reduced combinatorial designs are generated from a strict combinatorial perspective in order to create a reduced design with a balanced distribution of all factor settings. The algorithm selecting the points treats multilevel and qualitative factors equally.

The default start for an RC design is a linear model and the proposed number of experiments are a minimalistic proposal according to the number of degrees of freedom required by the model. More complex models with higher order terms can be specified and RC designs with sufficient number of runs created.

The evaluation of the design quality is done with G-efficiency and Condition number in MODDE.

The algorithm for reduced combinatorial optimizes the  $J_2$  (Hongquan Xu (2002)) criteria, described as a function that aims to reach as close to an orthogonal array as possible. The resulting design may include more replicated (duplicated) runs than specified, due to the properties of this algorithm. Receiving some duplicates is a trade-off between a fast and functional algorithm and the optimally reduced combinatorial design.

---

---

**Hint:** To find duplicates, add a random response (run order) to the Worksheet and on the **Worksheet** tab click **Replicates** and look for the blue plot symbols. If there are more replicates than desired, either exclude them, or create a design with more runs and exclude replicates. Carefully evaluate the condition number for the final design.

---

---

## Stability testing design

A pharmaceutical product in storage may change its quality characteristics with time. Hence, it is important to know how well a product retains its quality characteristics over the life span of the product. A product is considered stable as long as its quality characteristics remain within specifications. The shelf life of a product corresponds to the number of days it remains stable at the recommended storage conditions. The process of collecting experimental data for estimating and verifying a product's shelf life is called *stability testing*.

For more, see the Stability testing design section in the Generalized subset designs appendix.

### **Plackett Burman designs**

Plackett Burman designs are fractional factorial designs of resolution III, generated with 4, 8, 12, 16, 20, 24, and 28 and more runs. Plackett Burman designs support only linear models; *i.e.* you cannot estimate any two-factor interactions.

### **Plackett Burman Super-Saturated designs**

Plackett Burman Super-Saturated designs, PBSS, are fractional factorial designs of resolution II, generated with 4, 6, 8, 10, 12, 14, 16, and 18 runs. A super-saturated design is a Resolution II design with fewer runs than factors. Main effects are confounded with main effects. It is assumed that only very few of the factors investigated are active.

### **D-Optimal designs**

D-Optimal designs are computer generated designs that maximize the determinant of the  $X'X$  matrix,  $X$  being the extended design matrix.

D-Optimal designs are available for all objectives.

For more see the D-Optimal designs section later in this chapter.

### **Onion designs**

Like regular D-Optimal designs, D-Optimal Onion designs can be used both in screening and in RSM with quadratic models. The Onion designs comprise layers of designs, usually D-Optimal, where the outermost layer determines which type of model (linear, interaction or quadratic) that the Onion design supports. For more see the D-Optimal onion designs section later in this chapter.

### **Rechtschaffner designs**

Rechtschaffner designs are orthogonal, saturated fractions of resolution V of the 2<sup>n</sup> and 3<sup>n</sup> factorial designs. They allow the estimation of all main effects and all first order interactions without confounding. They are saturated designs, with no degrees of freedom remaining for the estimation of residuals and diagnostics.

The 2<sup>n</sup> Rechtschaffner designs are well suited when the objective is screening, with 6 or more factors, and little knowledge about the importance of each individual first order interaction. In this case it is of interest to estimate all first order interactions, unconfounded, and then eliminate the insignificant (small) ones, hence recovering some degrees of freedom for diagnostics and residual analysis.

The required number of runs  $N$  for the 2<sup>n</sup> Rechtschaffner designs with  $k$  factors is:

$$N = 1 + k + k(k - 1)/2$$

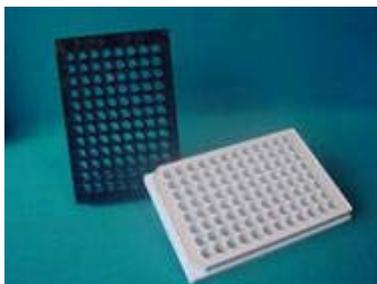
It is recommended to add 3 to 4 center points to these designs.

### **RED-MUP designs**

The RED-MUP designs are custom designs developed for the use with 96 well plates (see figure) and larger (384, 1536, etc.). These are widely used platforms for experimentation in biochemistry, microbiology, pharmaceutical development, etc., with some special properties (buildable and extendable to other factor intervals). The RED-MUP designs consist of two sub-designs corresponding to the vertical and horizontal directions of the plates, i.e., 8 and 12, respectively, for 96-well plates. The total design is made by multiplying the two sub-designs together. Hence, this total design supports a model with all interactions between the factors in the sub-designs, *plus* from each sub-design, the main effects, and when these sub-designs so support, interactions, and quadratic effects.

Below, we use  $n_1$  and  $n_2$  for the number of rows and columns in the plate, i.e., 8 and 12 in a 96 hole plate. A 96-well plate can handle from 5 full “RSM factors” up to 18 factors for a stretched screening situation.

The layout of a 96 well plate has 8 rows and 12 columns. Hence, the vertical direction has  $n_1 = 8$ , and the horizontal direction has  $n_2 = 12$ .



If both sub-designs support only main effects, using for example Plackett Burman sub-designs, up to  $n_1 + n_2 - 2$  factors can be investigated, *i.e.*, up to 18 factors in a 96 well plate. Such a sparse design without center points is not recommended. More reliable designs with center points in the larger sub-design would allow  $n_1 - 3 + n_2 - 5 = n_1 + n_2 - 8$  factors, *i.e.*, 12 factors for a 96 well plate.

When specifying the RED-MUP design, it is important to distribute the factors over the two sub-designs (vertical and horizontal) so that (a) the actual experimental protocol remains simple and doable, (b) the sub-designs and final design make chemical/biological/engineering sense, and (c) a-priori interesting interactions and higher order terms can be estimated. Note that all the interactions between each of the factors in the vertical design and each of the factors in the horizontal designs can always be estimated. Hence, factor pairs for which interactions are expected should be split into the two sub-designs. Then their interaction can always be estimated, regardless of choice of sub-design.

## Special designs

When selecting to create a RED-MUP design, there are special designs for the 96 well plates (8 x 12) which aim to make better use of the plate.

### **Definitive screening designs**

Definitive screening designs are available when there are at least 4 regular quantitative factors and no other factor types. These designs are parsimonious and the minimum number of runs is  $2^k+1$  (excluding replicates) for  $k$  factors. With definitive screening designs all factors are performed at 3 levels. The default regression model is a model with linear and quadratic terms, but no two-factor interactions. It should be observed that the quadratic terms are partially correlated and that these correlations increase with increasing number of factors in the investigation.

## RSM designs

RSM designs are used in later stages of an investigation to develop more elaborate models (quadratic) in the few important factors, usually not more than 5 or 6.

MODDE supports the following RSM designs:

### **Full factorial design at three levels**

Full factorial design at three levels is the full factorial design, with every factor varied at three levels.

### **Central composite designs CCC, CCO and CCF**

The three central composite designs available in MODDE are the Central Composite design Circumscribed (CCC), Orthogonal (CCO) and Face Centered (CCF). CCO is recommended thanks to its advantage as a compromise between CCF, which only has 3 levels, and CCC, which star points may end up impractically far away from the Low and High settings.

MODDE supports CCC, CCO and CCF designs for up to 12 factors.

These designs are composed of:

- A full or fractional factorial design.
- Star points.
- Replicated center points.

MODDE also supports a **reduced CCC and CCF** for four factors, with the fractional part of the design reduced from 16 to 12 runs.

Note that with the CCC and CCO designs you may edit the model and include cubic terms, if you wish.

### **Box Behnken designs**

Box Behnken designs are three level designs. All the design points are located at the center of the edges of the cube or hypercube, and are all situated on the surface of a sphere.

### **D-Optimal designs**

D-Optimal designs are computer generated designs that maximize the determinant of the  $X'X$  matrix,  $X$  being the extended design matrix.

D-Optimal designs are available for all objectives.

For more see the D-Optimal designs section later in this chapter.

### **Onion designs**

Like regular D-Optimal designs, D-Optimal Onion designs can be used both in screening and in RSM with quadratic models. The Onion designs comprise layers of designs, usually D-Optimal, where the outermost layer determines which type of model (linear, interaction or quadratic) that the Onion design supports. For more see the D-Optimal onion designs section later in this chapter.

### **Rechtschaffner designs**

Rechtschaffner designs are orthogonal, saturated fractions of resolution V of the 2<sup>n</sup> and 3<sup>n</sup> factorial designs. They allow the estimation of all main effects and all first order interactions without confounding. They are saturated designs, with no degrees of freedom remaining for the estimation of residuals and diagnostics.

The 3<sup>n</sup> Rechtschaffner designs are well suited for the RSM objective with 6 or more factors as they require fewer runs than the classical CCC or CCF non saturated designs. The intent with these designs is to estimate quadratic terms but performing fewer runs than with CCC or CCF. Eliminating insignificant terms, after performing the experiments, results in recovering some degrees of freedom.

The required number of runs N for the 3<sup>n</sup> Rechtschaffner designs with k factors is:

$$N = 1 + 2k + k(k - 1)/2$$

It is recommended to add 3 to 4 center points to these designs.

### **RED-MUP designs**

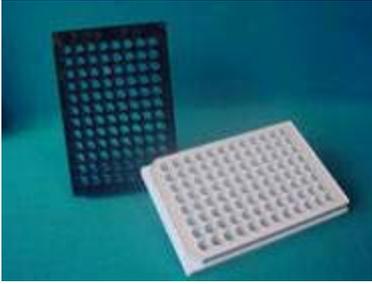
The RED-MUP designs are custom designs developed for the use with 96 well plates (see figure below) and larger (384, 1536, etc.). These are widely used platforms for experimentation in biochemistry, microbiology, pharmaceutical development, etc., with some special properties (buildable and extendable to other factor intervals). The RED-MUP designs consist of two sub-designs corresponding to the vertical and horizontal directions of the plates, i.e., 8 and 12, respectively, for 96-well plates. The total design is made by multiplying the two sub-designs together. Hence, this total design supports a model with all interactions between the factors in the sub-designs, *plus* from each sub-design, the main effects, and when these sub-design so support, interactions, and quadratic effects.

The RED-MUP designs are well suited for the RSM objective with up to 5 or 6 factors. The intent with these designs is to get a precise model that can be used for optimization and for detailed understanding.

The maximum number of “RSM factors” depends on the sizes of the sub-designs. An 8 run sub-design, e.g., a Doehlert design with 2 center points, supports 2 “RSM factors” (1 constant, 2 linear, two quadratic, and one interaction terms), and a 12 run sub-design, e.g., a three level Rechtschaffner design with 2 center points, supports 3 “RSM factors” (1 constant, 3 linear, 3 quadratic, and 3 interaction terms) for a total of 5 “RSM factors” for a 96-well plate.

### **Mixed objective**

Since the RED-MUP designs are constructed from two sub-designs, one of these can be an RSM design and the other a screening design. In such a case the objective is said to be *mixed*.



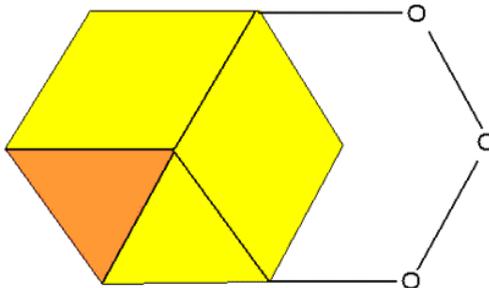
### Special designs

When selecting to create a RED-MUP design, there are special designs for the 96 well plates (8 x 12) which aim to fill up the plate.

### Doehlert designs

The Doehlert designs are quadratic RSM designs with some special properties (buildable and extendable to other factor intervals). They allow the estimation of all main effects, all first order interactions, and all quadratic effects without confounding. They are saturated designs with similar properties to the CCF, CCO and CCC designs. Geometrically they are polyhedrons based on hyper-triangles (simplexes), with a hexagon in the simplest two-factor case.

Doehlert design in 2 factors with 6 runs + center points can be extended to a new design by adding 3 experiments. Usually also one or two new center points are added in the new design (*i.e.*, in the figure the right-most point in the old design).



The Doehlert designs are well suited for the RSM objective with up to 5 or 6 factors (respectively 33 and 45 runs with 3 center points). The intent with these RSM designs is to get a precise model that can be used for optimization and for detailed understanding.

The required number of runs  $N$ , except for replicated center points, for the quadratic Doehlert designs with  $k$  factors is:

$$N = 1 + k + k^2$$

It is recommended to add 3 to 4 center points to these designs.

---

## Designs for mixture factors

In a mixture experiment the responses of interest depend only on the relative proportions of the components (called mixture factors) that make up the mixture or formulation. Hence, the sum of all the mixture factors is a constant  $T$ , usually equal to 1 when no mixture factors are kept constant.

### Mixture and process factors

Mixture factors are expressed as the fraction of the total amount of the formulation. Their experimental ranges lie between 0 and 1.

Regular factors (i.e., temp, pH, etc.) that are not part of the mixture or formulation are referred to as process factors. These are expressed as amounts or levels, and can be either quantitative (measured on a continuous scale) or qualitative (have only discrete values).

MODDE supports both mixture and process factors in the same experiment.

### Mixture factors definition

A mixture factor can be a formulation factor or a filler factor. Only one mixture factor can be defined as filler.

#### *Formulation factor*

Formulation factors are the usual mixture factors used in formulations with specifically defined experimental ranges. Most mixture experiments have only formulation factors.

#### *Filler factor*

The presence of filler is typical of certain types of simple mixture experiments. For example in a synthesis the solvent is typical filler, as is water in a juice punch. A filler is a mixture component, usually of little interest, making up a large percentage of the mixture, and added at the end of a formulation to bring the mixture total to the desired amount.

It is recommended to define a mixture factor as filler when all three conditions below are fulfilled;

- The factor is always present in the mixture,
- The factor accounts for a large percentage of the mixture and there is no restriction on its range. It is added at the end to bring up the mixture total to the desired amount (usually 1 when no mixture factors are kept constant), and
- You are not interested in the effect of the filler per se.

When you specify a filler factor, MODDE checks that these conditions are met and defaults to a slack variable model, with the filler factor omitted from the model.

### *Use*

All mixture factors are controlled or constant. The **Uncontrolled** option is unavailable for both formulation and filler factors.

Formulation factors can be defined as **Constant** when you want to keep them constant in the experiment.

When mixture factors are constant, the mixture total  $T = 1 - \text{Sum}$  (constant mixture factors). When no formulation factors are defined as constant, the mixture total has to be equal to 1. MODDE issues an error message and stops whenever the mixture total is not equal to  $T$  or 1.

---

*Note: A filler factor cannot be **Constant**.*

---

### **Scaling**

Mixture factors are always unscaled when you fit the model with MLR. When you fit the model with PLS, all mixture factors are scaled to unit variance.

---

*Note: When the mixture region is regular, mixture factors are first transformed to pseudo components, and then scaled with PLS models.*

---

### **Mixture constraint**

In a mixture experiment the mixture total (*i.e.* the sum of all the mixture factors in the experiment) is equal to a constant  $T$ . The mixture Total  $T$  is generally equal to 1 when no mixture factor is kept constant. This mixture constraint implies that the mixture factors are not independent, and this collinearity has implications on the mixture experimental region, the mixture designs, and the mixture model formulation.

### **Mixture experimental region**

When all mixture factors vary from 0 to  $T$  (the mixture total), the shape of the experimental region is a *Simplex*. With constraints on their ranges, the experimental region is usually an irregular polyhedron inside the simplex. In some constrained cases, as for example, with lower bounds constraints only, the experimental region is a small simplex inside the original simplex. See Crosier (1984).

MODDE checks for consistent bounds, and computes

$$R_U = \sum U_i - T,$$

$$R_L = T - \sum L_i.$$

$L_i$  and  $U_i$  are the lower and upper bound of the  $i^{\text{th}}$  mixture factors.

From  $R_L$ ,  $R_U$  and  $R_i$  (the range of every formulation factor) MODDE determines if the experimental region is a Simplex (the  $L$  simplex oriented as the original one, or the  $U$  simplex with opposite orientation) or an irregular polyhedron.

### **Regular region pseudo components transformations**

When the mixture region is the  $L$  or  $U$  simplex, MODDE defaults to transforming the mixture factors to pseudo component to make all their ranges vary between 0 and 1. This is very similar to orthogonal scaling of process factor, to make their ranges vary between -1 and +1.

With a regular mixture region, MODDE uses classical mixture designs.

The design is expressed in pseudo components and the worksheet is of course always displayed in original units.

The analysis is performed on the mixture factors transformed to pseudo component, as the coefficients of the Cox model can then be directly interpreted as the mixture factors effects.

### **Region is the L simplex**

When the mixture region is the  $L$  simplex, the  $L$  pseudo component transformation is defined as

$$P_i = (X_i - L_i) / (R_L).$$

The transformed mixture factors  $P_i$  vary from 0 to 1.

### **Region is the U simplex**

When the mixture region is the  $U$  simplex, the  $U$  pseudo component transformation is defined as:

$$P_i = (U_i - X_i) / (R_U)$$

The transformed mixture factors  $P_i$  vary from 0 to 1, but in this case the new simplex in the  $P$ 's has an opposite orientation to the original simplex in  $X$ , that implies that effects in  $P$  are reversed from those in  $X$ .

## **Classical mixture designs**

When all factors are mixture factors and the shape of the region is a simplex, the designs available in MODDE are the following classical mixture designs (all classical mixture designs are displayed in pseudo components in the design matrix, and by default the analysis is done with the formulation factors transformed to pseudo components).

### **Screening designs**

MODDE provides three variants of the axial design. Axial designs locate all the experimental points on the axis of the simplex and are recommended for screening, see Snee (references).

#### **Standard Axial (AXN)**

The standard axial design includes the following  $2 * q + m$  runs ( $q$  = number of mixture factors,  $m$  centroid points as specified by user).

1. All the  $q$  vertex points. The coordinates of the  $i^{\text{th}}$  Vertex point is  $x_i = (0, 0, 0..1, 0, 0..)$ .
2. All  $q$  interior points of the simplex. The coordinates of the  $i^{\text{th}}$  Interior point is  $x_i = (1/2q, 1/2q, 1/2q,..(q+1)/2q, 1/2q, 1/2q..)$ .
3. The overall centroid of the simplex with coordinates  $x = (1/q, 1/q, ....., 1/q..)$  replicated  $(m-1)$  times.

#### **Extended Axial (AXE)**

The extended axial design includes the following  $3*q + m$  runs ( $q$  = number of mixture factors,  $m$  specified by user).

1. All the  $q$  vertex points. The coordinates of the  $i^{\text{th}}$  Vertex point is  $x_i = (0, 0, 0..1, 0, 0..)$ .
2. All  $q$  interior points of the simplex. The coordinates of the  $i^{\text{th}}$  Interior point is  $x_i = (1/2q, 1/2q, 1/2q,..(q+1)/2q, 1/2q, 1/2q..)$ .
3. All the  $q$  End points. The coordinates of the  $i^{\text{th}}$  End point is  $x_i = (1/(q-1), 1/(q-1), 1/(q-1), 0, 1/(q-1), 1/(q-1)..)$ .
4. The overall centroid of the simplex with coordinates  $x = (1/q, 1/q, ....., 1/q..)$  replicated  $(m-1)$  times.

**Reduced Axial (AXR)**

The reduced axial design includes the following (q+m) (specified by user) points:

1. All the q vertex points.
2. A subset or none (specified by user) selected from the q interior points.
3. The overall centroid replicated as desired.

**RSM**

MODDE provides 2 variants of the quadratic model designs, one special cubic and one cubic. The simplex centroid design has all the experimental points on the vertices, and on the center of the faces of consecutive dimensions.

**Modified simplex centroid (SimM)**

The modified simplex centroid design supports a quadratic model and includes the following:

1. The q vertex points. The coordinates of the  $i^{\text{th}}$  Vertex point is  $x_i = (0, 0, 0..1, 0, 0..)$ .
2. The  $(q(q-1))/2$  Edge centers. The coordinates of the  $ij^{\text{th}}$  edge point is  $x_{ij} = (0, 0, 1/2, 1/2, 0, 0..)$ .
3. The q Interior check points. The coordinates of the  $i^{\text{th}}$  interior point is  $x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..)$ .
4. The overall centroid with coordinates  $x = (1/q, 1/q, \dots, 1/q)$ , replicated as desired.

**Modified simplex centroid Face center (SimF)**

The modified simplex centroid face center design supports a quadratic model and includes the following:

1. The q vertex points. The coordinates of the  $i^{\text{th}}$  Vertex point is  $x_i = (0, 0, 0..1, 0, 0..)$ .
2. The  $(q(q-1))/2$  Edge centers. The coordinates of the  $ij^{\text{th}}$  edge point is  $x_{ij} = (0, 0, 1/2, 1/2, 0, 0..)$ .
3. The q Face centers of dimension (q-1). The coordinates of the  $i^{\text{th}}$  face center is:  $(1/q-1, 1/q-1, \dots, 0, 1/q-1..1/q-1)$ .
4. The q Interior check points. The coordinates of the  $i^{\text{th}}$  interior point is  $x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..)$ .
5. The overall centroid with coordinates  $x = (1/q, 1/q, \dots, 1/q)$ , replicated as desired.

**Simplex centroid Special Cubic (SimSC)**

The simplex centroid special cubic design supports a special cubic model and includes the following:

1. The q vertex points. The coordinates of the  $i^{\text{th}}$  Vertex point is  $x_i = (0, 0, 0..1, 0, 0..)$ .
2. The  $(q(q-1))/3, 2/3$  Edge points. The coordinates of the  $ij^{\text{th}}$  edge point is  $x_{ij} = (0, 0, 1/3, 2/3, 0, 0..)$ ,  $x_{ji} = (0, 0, 2/3, 1/3, 0, 0..)$ .
3. The  $q(q-1)(q-2)/6$  Face centers of dimension 2. The coordinates of the  $i^{\text{th}}$  face center is  $(0, 0, 0, 1/3, 1/3, 1/3..0, 0, )$ .

4. The  $q$  Interior check points. The coordinates of the  $i^{\text{th}}$  interior point is  $x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..)$ .
5. The overall centroid with coordinates  $x = (1/q, 1/q, \dots, 1/q)$ , replicated as desired.

### Simplex Centroid Cubic (SimC)

The simplex centroid cubic design supports a cubic model and includes the following:

1. The  $q$  vertex points. The coordinates of the  $i^{\text{th}}$  Vertex point is  $x_i = (0, 0, 0..1, 0, 0..)$ .
2. The  $(q(q-1))/3, 2/3$  Edge points. The coordinates of the  $ij^{\text{th}}$  edge point is  $x_{ij} = (0, 0, 1/3, 2/3, 0, 0..)$ ,  $x_{ji} = (0, 0, 2/3, 1/3, 0, 0..)$ .
3. The  $q(q-1)/2$  Edge centers. The coordinates of the  $i^{\text{th}}$  edge center is  $x_i = (0, 0, 0, 1/2, 1/2, 0..0)$ .
4. The  $q(q-1)(q-2)/6$  Face centers of dimension 2. The coordinates of the  $i^{\text{th}}$  face center is  $(0, 0, 0, 1/3, 1/3, 1/3..0, 0, )$ .
5. The  $q$  Interior check points. The coordinates of the  $i^{\text{th}}$  interior point is  $x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..)$ .
6. The overall centroid with coordinates  $x = (1/q, 1/q, \dots, 1/q)$ , replicated as desired.

## D-Optimal designs

### What are D-Optimal designs?

**D-Optimal** designs are computer generated designs, tailor made for a specific problem. They allow great flexibility in the specifications of your problem. They are particularly useful when you want to constrain the region and no classical design exists.

“D-Optimal” means that these designs maximize the information in the selected set of experimental runs with respect to a stated model.

For a specified regression model  $Y = X*\beta + \epsilon$  where:

$Y$  is a  $(N \times 1)$  vector of observed responses,

$X$  is a  $(N \times p)$  extended design matrix, i.e. the  $n$  experimental runs extended with additional columns to correspond to the  $p$  terms of the model (i.e., the added columns are for the constant term, interaction terms, square terms, etc.)

$\beta$  (beta) is a  $(p \times 1)$  vector of unknown coefficients to be determined by fitting the model to the observed responses.

$\epsilon$  (epsilon) is a  $(N \times 1)$  vector of residuals (the differences between the observed and predicted values of the response  $y$ ). They are assumed to be independent of each other, normally distributed and with constant variance  $\sigma^2$

The D-Optimal design maximizes the determinant of the  $X'X$  matrix, which is an overall measure of the information in  $X$ . Geometrically; this corresponds to maximizing the volume of  $X$  in a  $p$  dimensional space.

## Candidate set

D-Optimal designs are constructed by selecting N runs from a candidate set. This candidate set is the discrete set of “all potential good runs”.

MODDE generates the candidate set as follows:

**I)** For a regular process region, the candidate set consists of one or more of the following sets of points (depending on your model and the number of factors):

- The full factorial for up to 10 factors, reduced factorial for up to 32 factors.
- Centers of edges between hyper-cube corners.
- Centers of the faces of the hyper-cube.
- Overall centroid.

**II)** For constrained regions of mixture and/or process factors, the candidate set consists of one or more of the following set of points:

- The extreme vertices of the constrained region.
- The centers of the edges. If these exceed 200, the centers of the 200 longest edges.
- The centers of the various high dimensional faces.
- The overall centroid.

MODDE has implemented an algorithm to compute the extreme vertices, center of edges, center of faces etc. as described by Piepel (1988).

## When do I use D-Optimal designs?

Whenever possible you should use classical designs and these are the default designs of MODDE. However when classical designs are impossible to apply, D-Optimal designs are the preferred choice.

MODDE suggests a D-Optimal design when:

1. There is a **linear constraint** on the factor settings, reducing the experimental region to an irregular polyhedron. There are no classical designs that can well investigate an **irregular region**. A D-Optimal design is then the preferred choice as it makes efficient use of the entire experimental space.
2. There are **formulation** factors, with lower and upper bounds, and possibly additional constraints, making the region an **irregular polyhedron**.
3. There are **qualitative** factors, with **more than two levels** and there is no mixed level design available. Or the mixed level design suggests too many runs to be acceptable.
4. The objective is **RSM** and there are **qualitative** factors.
5. The **number of experimental runs** affordable is smaller than the number of runs of any available classical design.
6. **Both process and mixture factors** are present.

## D-Optimal algorithm

D-Optimal designs have been criticized for being too dependent on an assumed model. To reduce the dependence on an assumed model, MODDE has implemented a Bayesian Modification of the K-Exchange algorithm of Johnson and Nachtshiem (1983), as described by W. DuMouchel and B. Jones in “A Simple Bayesian Modification of D-Optimal designs to reduce dependence on an Assumed Model”, *Technometrics* (1994).

With this algorithm one can add to the “primary terms” i.e. the terms in the model, “potential terms”, i.e. additional terms that might be important. The objective is to select a D-Optimal design, rich enough to guard for potential terms, and enable the analysis to detect possibly active ones.

In order not to increase the number of runs  $N$ , and to avoid a singular estimation, one assumes that the coefficients of the potential terms are likely to have a mean of 0 and a finite variance  $(\tau, \tau)^2$ .

## Implementation of the D-Optimal algorithm in MODDE

### *K-exchange algorithm*

The k-exchange algorithm is a compromise between the exchange algorithm of Wynn (1972) with  $k = 1$  and the Federov algorithm with  $k = N$  (the selected number of runs). In MODDE  $k$  is set to 3, that is at every iteration of the procedure, the algorithm considers an exchange between  $k = 3$  points in the design with the smallest prediction variance and points in the candidate set. If any exchange increases the determinant, the point(s) (up to 3) are exchanged.

### *Variance of the coefficients of the potential terms*

As recommended by W. DuMouchel,  $\tau, \tau$ , is set to 1 in MODDE.

## Potential terms

Potential terms are higher order terms not included in the model but taken into account during the creation of the candidate set. Potential terms are default added but can be removed by clearing the **Use potential terms** box.

Depending on the number of factors, the objective and the model, MODDE adds the following potential terms:

### *Process Factors with constraints*

#### Screening

Factors	Model	Potential terms
2 - 12	Linear	All interactions
2 - 12	Linear + interactions	All squares

#### RSM

Factors	Model	Potential terms
2 - 8	Quadratic	All cubes

## Process Factors without Constraints

### Screening

Factors	Model	Potential terms
2 - 20	Linear	All interactions
21 - 32	Linear	Interactions between the first 20 factors
2 - 17	All interactions	All squares

### RSM

Factors	Model	Potential terms
2 - 6	Quadratic	All cubes
7 - 12	Quadratic	None

## Mixture Factors and irregular regions

### Screening

Factors	Model	Potential terms
2 - 20	Linear	All squares + interactions

### RSM

Factors	Model	Potential terms
2 - 12	Quadratic	All cubes

---

*Note: No potential terms are added for investigations with all factors defined as qualitative.*

---

## Design evaluation

To evaluate and compare D-Optimal designs, MODDE computes the following criteria;

### **LogDetNorm**

The log of the determinant of  $X'X$  normalized for number of terms in the model  $p$ , and number of runs  $N$ .

This is the criterion used, by default, to select the best design. MODDE selects the design with the largest value (closest to 0) of *LogDetNorm*.

$$\text{LogDetNorm} = \text{Log}_{10} [\text{Det}(X'X)^{1/p} / N]$$

The maximum value of *LogDetNorm*, for an orthogonal design, is 0.

### **LogDet**

The Log of the determinant of the  $X'X$  matrix

### Condition No

The condition number of the  $X$  design matrix coded orthogonally and extended according to the model.

### G efficiency

G efficiency is a lower bound on D efficiency, which compares the efficiency of a D-Optimal design to a fractional factorial.

G efficiency is defined as:

$$G_{eff} = (100 * p) / (n * d)$$

Where

$p$  = number of terms in the model

$n$  = number of runs in the design

$d$  = maximum relative prediction variance  $v$  over the candidate set, where the prediction variance  $v = x(X'X)^{-1}x'$

$x$  = a row in the candidate set

$X$  = the selected design

### Inclusions and design augmentation

MODDE allows you to specify a set of experimental runs as **Inclusions** specified under **Design | Inclusions**. If you enter experiments in Inclusion before creating your design these runs are by default a part of the resulting D-Optimal design.

Inclusions are useful for design augmentation. If you already have performed a few experiments, and want to add  $M$  additional experiments, add the old experiments in **Inclusions**, ask for  $N + M$  runs and state the desired model. The  $M$  runs are selected D-Optimally from the candidate set with respect to your model.

For more on design augmentation see the Complement design section in Chapter 4, File.

---

*Note: All of these statistics are computed from the runs selected D-optimally and do not include the possible center points added to the worksheet.*

---

### Irregular region

#### Screening

When the mixture region is an irregular polyhedron, MODDE computes the extreme vertices (corners) delimiting the region. These extreme vertices constitute the candidate set and the centers of the high dimensional faces are added to support potential terms. The design is a D-optimal selection of  $N$  (specified by user) runs from the candidate set.

#### RSM

MODDE computes the extreme vertices, 1/3, 2/3 centers of edges, centers of faces of dimension  $(q-1)$  and the overall centroid of the experimental region. When there are too many extreme vertices, only the centers of the 25% longest edges are computed. These experimental points constitute the candidate set.

The design is a D-Optimal selection of  $N$  runs (specified by the user) from the Candidate set.

### **Pseudo component transformation**

You can always select to have the mixture factors expressed in pseudo components for the analysis. MODDE uses the  $L$  pseudo component transformation when  $R_L \leq R_U$  and the  $U$  pseudo component when  $R_U < R_L$ .

Pseudo component transformation is the MODDE default when the method of fit is MLR as it stretches the experimental region and alleviates the problem of ill conditioning.

---

*Note: All mixture designs are displayed in pseudo components.*

---

### **Mixture models**

Because of the mixture constraint, (the mixture factors are not independent) the analysis of mixture data with multiple regression requires a special model form.

The traditional approaches have been:

- Defining the model omitting one mixture factor, hence making the others independent. This is the *slack variable* approach.
- Omitting some terms from the model, so that the terms remaining in the model are independent. This is the *Scheffé* model, with the constant term removed from the linear model and the quadratic terms removed from the quadratic model.
- Using the complete model including all the mixture terms, but putting constraints on the coefficients to make them estimable. This is the Cox reference model, and the constraints on the coefficients are defined with respect to a standard reference mixture. This standard reference mixture serves the same function as the centering constant with process variables models.

### **Process and mixture factors together**

When you have both process and mixture factors, you can select to treat them as one model, or to specify separate models for the mixture factors, and the process factors.

With both mixture and process factors, the only model form available is the Cox reference mixture model.

When the model obeys mixture hierarchy, the PLS coefficients are expressed relative to a stated standard reference mixture. The following constraints are imposed on the coefficients:

#### **For linear models**

$$\sum b_{ks} s_k = 0$$

#### **For quadratic models**

$$\sum b_{ks} s_k = 0 \quad (1)$$

$$\sum c_{kj} b_{kjs} s_k = 0 \text{ for } k = 1, \dots, q \quad (1) \text{ and for } j = 1, \dots, q \quad (2)$$

Here  $c_{kj} = 1$  when  $j \neq k$  and  $c_{kj} = 2$  when  $k = j$ .

$s_k$  are the coordinates of the standard reference mixture.

If  $\gamma$  (gamma) are the coefficients of the interactions between the process and mixture factors:

$$\sum \gamma_{ks} s_k = 0$$

---

*Note: When the model contains terms of order 3, or contains qualitative and formulation factors, the PLS coefficients are not adjusted relative to a stated standard mixture.*

---

## D-Optimal Onion designs

Onion designs can be created for regular process factors, using an imported candidate set, using scores from SIMCA as design variables, or letting MODDE create the candidate set from the factor setup.

When importing *scores*, the scores are automatically loaded from the SIMCA *usp*-file, and the candidate set for the Onion design is comprised of all objects (rows) in the workset of the SIMCA model selected as the basis of the Onion design.

The design is made in a number of layers (shells), with a separate D-optimal design for each layer. Typically the number of layers is two or three.

D-Optimal onion designs are similar to space filling designs in that design points are situated also in the interior of the design space.

D-Optimal onion designs are available in MODDE only when the factors are quantitative.

Observations in the candidate set are sorted by their distance to the center of the multivariate space, expressed as percentiles from the center.

The candidate set is then divided into layers, by default three, layer one being the innermost layer and layer three the outermost layer.

A D-Optimal design is then performed on each layer separately and the final design and worksheet includes all the runs selected D-Optimally in each layer. This makes the selected runs fill the multivariate design space.

The model and the number of runs in each layer as well as the percentile of observations included in each layer can be specified by the user.

The D-Optimal onion design selects runs from each layer separately, ensuring that the design will have points that fill the space.

SIMCA-P+ 12 or later needs to be installed.

## Screening onion designs

When the objective is screening, two D-Optimal Onion design are available. The recommended design has a full interaction model in the outer layer. The second choice is with a linear model in the outer layer.

The default number of layers (three when the candidate set allows it) can be changed from the **Layer** box. The outer layer is the last layer, and the innermost layer is number 1. The maximum allowed number of layers is default 10. You can change the **Max number of layers in onion designs** in **File | Options** on the **MODDE options** page in the **General** section.

For more about D-Optimal designs, see the D-Optimal section earlier in this appendix.

## RSM onion designs

When the objective is RSM, the quadratic D-Optimal Onion design is available. The recommended design has a full quadratic model in the outer layer.

The default number of layers (three when the candidate set allows it) can be changed from the **Layer** box. The outer layer is the last layer, and the innermost layer is number 1. The maximum allowed number of layers is default 10. You can change the **Max number of layers in onion designs** in **File | Options** on the **MODDE options** page in the **General** section.

For more about D-Optimal designs, see the D-Optimal section earlier in this appendix.

## Candidate set

The D-Optimal onion design in MODDE is created from a candidate set. The candidate set can be created by MODDE, imported from one of the supported file formats, or imported from SIMCA.

### ***Candidate set created by MODDE***

When you have defined only process factors and enter the **Select model and design** page, you can select to generate an **Onion** design.

MODDE will then create the same number of candidate sets as layers specified. The high and low limits for each factor in the candidate set will be based on the percentile defined for each layer.

E.g. for a factor with Low = -1 and High = 1 with the four layers, 0% – 15%, 15% – 30%, 30% – 75% and 75% – 100%, the candidate sets will be generated with the low/high settings -0.15/0.15, -0.3/0.3, -0.75/0.75 and -1/1. The number of points generated in each candidate set depends on the number of factors.

### ***Candidate set imported from a file***

For how to import a file as candidate set, see the Design from candidate set section in Chapter 4, File.

After importing the candidate set, selecting an Onion design and selecting the intervals (percentile) for the layers, the candidate set is divided into sub candidate sets for each layer, based on the experiments distance. The most distant experiment will define 100% and the center will define 0%.

The distance of a specific experiment from the center is calculated as the “geometrical distance”, i.e. the square root of the sum of squared factor values. The factors are orthogonally scaled.

### ***Candidate set from SIMCA***

For how to import scores from SIMCA as design variables (factors), see the Design from scores section in Chapter 4, File.

When you select to create an onion design, the candidate set is parted in layers.

When the main effects can be evaluated independently from one another (not confounded) the perfect reduced design is an orthogonal array (OA). The classical  $2^{k-p}$  fractional factorial designs are orthogonal arrays. When the factors have more levels than 2 (i.e. a combination of 2, 5, 3, 2, 3 levels over 5 factors) the complexity when creating an OA escalates considerably and the OA might not even exist. The concept of OA dates back to Rao (1947).

To create a design with a reasonable number of runs, the concept of nearly orthogonal arrays NOAs was invented and has been described by Wang and Wu (1992). The concept used in MODDE to solve the problem in creating OAs and NOAs is called the J2 algorithm, a fast and efficient algorithm for Design generation.

## Definition OA

Definition of orthogonal array (OA) can be found at [http://en.wikipedia.org/wiki/Orthogonal\\_array](http://en.wikipedia.org/wiki/Orthogonal_array) (Wikipedia).

## Definition of NOA

"Let  $S_i$  be a set of  $s_i$  levels denoted by  $0, 1, \dots, s_i - 1$  for  $1 \leq i \leq v$  for some positive integer  $v$ . We define a nearly-orthogonal array NOA  $(N, s_1^{k_1}, s_2^{k_2}, \dots, s_v^{k_v})$  to be an array of size  $N \times k$  such that  $k = k_1 + k_2 + \dots + k_v$  and the first  $k_1$  columns have symbols from  $S_1$ , the next  $k_2$  columns have symbols from  $S_2$ , and so on, such that the array is optimal according to some criterion.", from Hongquan Xu (2002)

## References

Rao, C. R. (1947), Factorial Experiments Derivable from Combinatorial Arrangements of Arrays, *Journal of the Royal Statistical Society*, Supplement,9, 128–139.

Wang, J. C., and Wu, C. F. J. (1992), Nearly Orthogonal Arrays With Mixed Levels and Small Runs, *Technometrics*, 34, 409–422.

Hongquan Xu (2002), An Algorithm for Constructing Orthogonal and Nearly-Orthogonal Arrays With Mixed Levels and Small Runs, *Technometrics*, 44, 356–368.

---

## Power of the design

This section describes how the power analysis is performed.

We will consider a regression situation with a dependent output variable,  $y$  ( $n \times 1$ ), that is predicted from a set of independent predictors,  $X$  ( $n \times p$ ), such that  $(\mathbf{x}_i, y_i)$ , for  $i = 1, \dots, n$  denote our set of measurements.

Power denotes our ability to detect a significant effect, given that there is one. It is often used to determine the number of runs that are required in a set of experiments in order to be confident that we would detect a significant effect of a particular size.

In our case, with a multiple regression model, we are interested in determining whether we have a sufficient number of runs to confidently detect whether at least one regression coefficient is significantly different from zero (see details below), given that it truly should be non-zero.

## Power of the model

We have a linear regression model

$$y = x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p + \varepsilon,$$

where we predict  $y$  as

$$\hat{y} = x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p,$$

and we want to test the null hypothesis that (Liu, 2014)

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

against the alternative hypothesis that

$$H_A: \beta_j \neq 0, \text{ for at least one } j = 1, 2, \dots, p.$$

I.e., that at least one regression coefficient is significant. We will test this hypothesis using an  $F$ -test.

The  $F$ -test is

$$F = \frac{\frac{SSR}{d_r}}{\frac{SSE}{d_e}} = \frac{SSR \cdot d_e}{SSE \cdot d_r},$$

where  $d_r = p$  is the degrees of freedom of the regression, and  $d_e = n - p$  is the number of degrees of freedom of the error. The sum of squares of the regression is

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , and the sum of squares of the error is

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

The coefficient of determination, also called the multiple correlation coefficient, is defined as

$$R^2 = \frac{SSR}{SST},$$

where  $SST = SSR + SSE$ .

Then  $SSE = (1 - R^2)SST$ , and the  $F$ -test may be written equivalently as

$$F = \frac{R^2}{1 - R^2} \frac{d_e}{d_r} = f^2 \frac{d_e}{d_r}$$

Where  $f^2 = R^2/(1 - R^2)$  is called the effect size.

Under the alternative hypothesis, the  $F$  statistic follows a non-central  $F$  distribution. The power of the  $F$ -test is then (Liu, 2014)

$$1 - \beta = 1 - P(F_{d_r, d_e}(\lambda) < F_{1-\alpha, d_r, d_e})$$

where  $F_{d_r, d_e}(\lambda)$  is the non-central  $F$  distribution with center parameter, where  $d_r$  and  $d_e$  are the degrees of freedom and  $F_{1-\alpha, d_r, d_e}$  is the critical  $F$ -value on the  $\alpha$  level. The center parameter,  $\lambda$ , is computed as

$$\lambda = f^2(d_r + d_e).$$

The interpretation is that, if the data is sampled many times and the regression model refitted for each sampled data set, we would correctly identify at least one non-zero regression coefficient  $100(1 - \beta)\%$  of the time.

Thus, the model's  $R^2$  value must be known, in order to compute the effect size, which in turn is required to compute the power. The  $R^2$  could be determined from previous studies, or from experience or assumptions about the current study.

The power is computed using an estimated  $R^2$  value, the  $\alpha$  value and the total number of runs in the currently selected design. The total number of runs is determined by the selected design and the presence of center points.

Also, in order to plan the study, the power analysis also presents the *suggested number of runs*, in order to achieve a power  $1 - \beta \geq 0.8$ .

## Post-hoc power analysis

The notion of post-hoc power (also called retrospective power, observed power or achieved power) has received a great deal of criticism, much of it fair. In fact, the notion of power is really only meaningful in the design phase of a study. If you set up a study, it is relevant to ask yourself how many runs are required in order to confidently detect an effect of a particular size. However, after the study has been performed, you no longer have any questions like that. Now everything is fixed: your data is already sampled, you have an actual number of runs that fell outside of the confidence range, your null hypothesis is now actually declared either true or false, etc. This last point is particularly important. We note that power is formally defined as (Zumbo and Hubley, 1998)

$$P(\text{reject } H_0 | H_0 \text{ is false}),$$

*i.e.*, the probability of rejecting the null hypothesis given that it is false. The post-hoc power, on the other hand is defined as

$$P(H_0 \text{ is false} | \text{reject } H_0),$$

*i.e.* the probability that the null hypothesis actually is false, given that we rejected the null hypothesis. The relation between a priori power and post-hoc power thus follows from Bayes' theorem, and is

$$P(H_0 \text{ is false} | \text{reject } H_0) = P(\text{reject } H_0 | H_0 \text{ is false}) \frac{P(H_0 \text{ is false})}{P(\text{reject } H_0)}.$$

A priori and post-hoc power are thus equivalent in the case when  $P(H_0 \text{ is false}) = P(\text{reject } H_0)$ , which may only happen in the limit as  $n \rightarrow \infty$  of the sample size. Note in particular that this is very likely to be false for a single study.

The main critique against the post-hoc power analysis is thus the following: The power states what is likely to happen in  $100(1 - \beta)\%$  of all cases (e.g. if the study were repeated many, many times). Your current study was performed once, and it is not possible to compute a probability like this from a single sample. It is therefore not possible to use the post-hoc power to interpret the power of the current study.

It is also common to falsely conclude that, if a result from a study was deemed non-significant,  $\beta$  states the probability that this is a false negative. By extension, this would imply that a significant result could have been obtained if more runs were included. To be perfectly clear: This is a *false* claim. There is a relation between the computed  $p$ -value and the power, and a non-significant  $p$ -value always correspond to a low power (Hoenig and Heisey, 2001).

Nevertheless, the post-hoc power is sometimes used, and may sometimes be useful. For instance, post-hoc power may be used in meta studies; and the post-hoc power is sometimes used in order to design future studies, e.g. if you performed a pilot study (Hoenig and Heisey, 2001; O'Keefe, 2007). MODDE therefore includes the post-hoc power in the **Descriptive Statistics** list.

## How high should the power be?

High power gives us confidence that we would detect a significant model if there is one. Ultimately, the required power is determined for the study at hand, and depends on the consequences of a false negative.

Conventionally, a power that is  $1 - \beta \geq 0.8$  is typically said to be good (just as  $\alpha$  is typically set to 0.05), and is motivated by  $\beta = 4\alpha$  (Cohen, 1988). A higher power is of course better.

## References

- Cohen, Jacob (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New-Jersey, 2nd edition.
- Hoenig, J. M. and Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55, 19-24.
- Liu , Xiaofeng Steven (2014). *Statistical Power Analysis for the Social and Behavioral Sciences: Basic and Advanced Techniques*. Taylor & Francis, 1st edition.
- O’Keefe, Daniel J. (2007). Post Hoc Power, Observed Power, A Priori Power, Retrospective Power, Prospective Power, Achieved Power: Sorting Out Appropriate Uses of Statistical Power Analyses. *Communication Methods and Measures*, 1(4), 291-299.
- Zumbo, Bruno D. and Hubley, Anita M. (1998). A note on the misconceptions concerning prospective and retrospective power. *The Statistician*, 47, 385-388.



# Optimizer appendix

---

## Introduction

The optimizer works according to a given set of specifications and the specification of the factors and responses are selected according to the desired result. Therefore, if the response specifications are unrealistic, it might be impossible for the optimizer to reach the best possible solution. With a good strategy and by using complementary tools, such as contour plots, probability of failure estimates, sweet spot plots, Design Space estimates, robust setpoint, setpoint analysis and the predicted min/predicted max values listed in the Optimizer, a good understanding of the possible specifications can be obtained. Note here that the most important consideration in order to obtain optimal optimization results is to certify that the *Pred. min* and *Pred. max* range and the response specifications at least partly overlap.

The optimizer is used to find an experimental *setpoint* that fulfills various criteria. The optimizer uses a search function to find the best possible solution to an equation that depends on a number of operating criteria.

This appendix describes the possibilities and limitations of the optimizer function. The first part is a description of how the optimizer works and the second part discusses how different objectives can be reached by selecting different start criteria for the optimization.

---

## Search function

The optimizer uses desirability functions,  $d_k$ , for each response,  $k=1, \dots, m$ , and searches for the combination of factor settings that predicts a result inside the response specifications and as close as possible to the targets for all responses. When searching for a solution with many criteria, the result will be a compromise between those criteria. This compromise is expressed as the *overall desirability function*,  $f(\mathbf{ds})$ , a sum of all  $d_k$ . This compromise is also expressed as a normalized distance to target ( $D$ ) for all responses.

The success of the desirability function depends on the optimizer specification (Min, Target, Max) and the selected **Desirability** objective (Limit/Target/Custom/etc.). It must be possible to reach the optimizer objective for the current data in order for the desirability function to succeed.

The search for a robust setpoint, **Find robust setpoint**, is based on Monte Carlo simulations and is available if a setpoint can be found that predicts all responses within their limits.

---

## Optimizer objectives

The optimizer can be set up for different objectives:

1. *Limit optimization* – where the objective is to reach a solution in which the response is within the specification limits (Min and Max limits). This is the default approach in MODDE.
2. *Target optimization* – where the objective is to reach a solution in which the response is as close to target as possible. For the target optimization to work properly, it is necessary that the response can be optimized close to or on target; otherwise the search may end up with an unacceptable solution.
3. *Custom optimization* – user defined customization of the Target optimization.
4. *Focus optimization* – where the objective is to favor one or several responses over the others; accomplished by manipulating the individual weights.
5. *Robust setpoint* – where the most robust setpoint is found. Depends on the existence of a solution based on objectives 1-4.

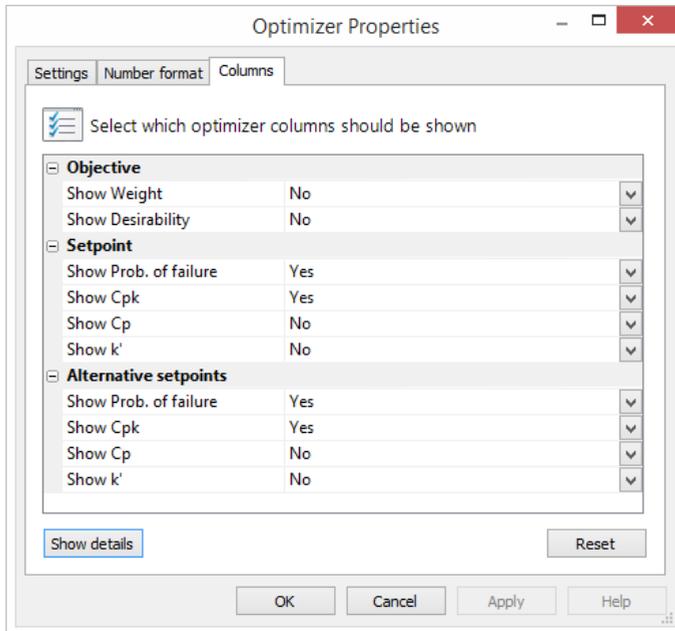
To control the optimization, the overall desirability function,  $f(ds)$ , plays a key role, as well as reasonable limits and targets for the responses. The optimizer will strive to reach the lowest possible value of the overall desirability function,  $f(ds)$ , and will strive to reach the lowest possible value. The shape of the function is controlled by the settings of the criteria (Min, Target, Max) for each response and the choice of the  $d_k$  functions.

With weight 1, the lowest possible value of the individual desirability functions,  $d_k$ , is -100. Note that in the plots, the individual desirability function,  $d_k$ , is translated to  $[-1, 0]$ , so that a weight of 1 has a lowest possible value in the plot of -1, and so on. The goal is to reach a minimum in the overall desirability function,  $f(ds)$ .

### Accessing the individual desirability functions

Open the **Optimizer Properties** window by clicking **Optimizer settings** in the **Optimizer** window or right-clicking the **Optimizer** window and then click **Properties**. Then click the **Columns** tab of the **Optimizer Properties** dialog box.

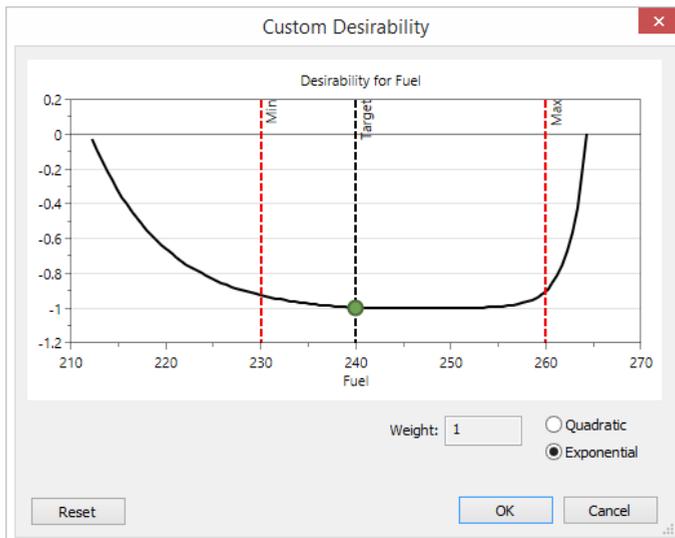
This presents additional columns for the optimizer. Once selected, the additional columns will appear in the spreadsheet in the optimizer.



*Note: For a weight change to have the expected result, the selected desirability needs to be **Target**.*

### Limit optimization

The default function is an exponential desirability function. This means that the desirability function decreases rapidly close to the limit and then flattens out. As a consequence it will be easier to reach a compromise where all responses are inside the specification limits but may not be as close to the desired target as is possible.



When **Weight** = 1 the desirability function reaches the lowest possible value just inside the specification limit. This works well to find a compromise when many responses strive to be inside the specification limits. A possible drawback can be that the optimizer will not drive the solution to the **Target** even when that is possible. There is also no relative symmetry in the function as it depends on the limit – target distance as shown in the plot above.

### Target optimization

The quadratic function corresponds to *Target optimization*, meaning that a response is expected to reach the target or be close to the target.

If that is not possible, the optimizer might find a solution that will predict some responses very close to target and some outside the specification limits. In such a case, a good recommendation is to start with the exponential desirability functions (*i.e. Limit optimization*) and proceed with *Target optimization* to reach inside the given specifications.

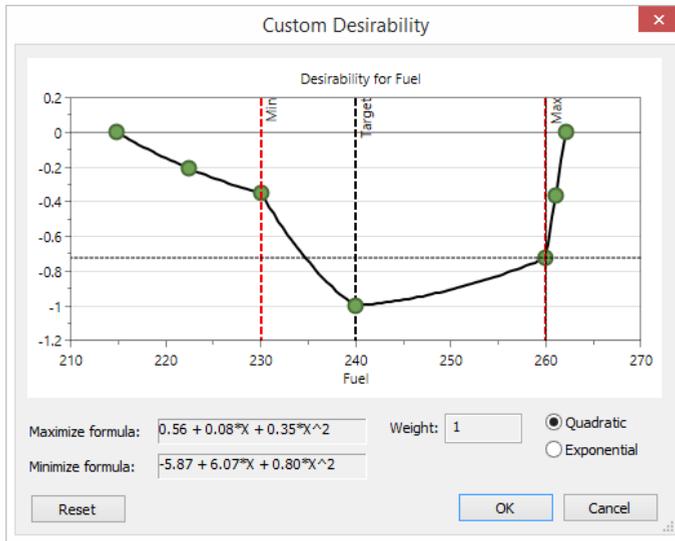


In Target optimization, the desirability function has a quadratic form and can be customized by adjusting the dots. The dots correspond to the optimizer specifications (Min, Target, Max), the predicted min and max (the extreme points) and a setting in the middle of the specification and predicted ranges for Min and Max.

### Customized desirability function

It is possible to change the shape of the quadratic desirability function at the nodes represented by the blue dots. A function like the one in the following figure will favor solutions inside the upper specification limit and will reach a solution from that direction.

To access the Custom Desirability plot the Desirability column needs to be shown on the Objective page in the Optimizer window. See the Optimizer columns subsection in Chapter 12, Optimizer for details.



## Focus optimization

If the weights are set differently for different responses, responses with higher weights will take priority in the search for a solution inside the specifications. The overall optimization criterion is to reach the lowest value of the overall desirability function,  $f(ds)$ .

The search will focus on minimizing desirability functions with high weights, since the objective is to minimize the overall desirability function.

*Note: The selected Desirability must be **Target** in order for the focus optimization to work well.*

## Define optimizer specifications

The start specifications for the optimizer are based on the initial factor and response definitions. If no response specification for Min, Target, and/or Max exists, the default criterion is 'Predict'.

Reaching an optimal result is often an iterative process. The criteria will need to be reevaluated if the response specifications are impossible to reach. If compromises need to be made on many responses, or a solution where the specification for one response is far out of reach, this can point the search function in the wrong direction.

To help the specification evaluation, MODDE uses the current model and estimates the predicted min and max values for each response. These values, complemented with a graphical presentation, can be used as a guide when evaluating the possible ranges. In a situation where one response specification is far away from the possible range, reconsider the specification or set the Criterion for the response to Predicted and optimize that specific response separately.

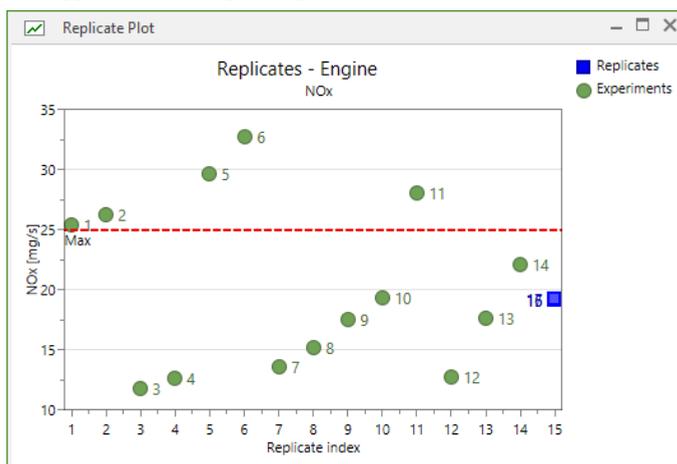
	Response	Criterion	Min	Target	Max	Pred. min	Pred. max	Graph	Desirability
1	Fuel	Minimize		214.097	250	214.785	262.105		Limit
2	NOx	Minimize		5	10	11.274	32.4767		Limit
3	Soot	Minimize		0.0182831	1	0.0650316	3.92071		Limit

In the optimizer specification above, the response NOx has a specification that is not inside the possible region. With these settings, the solution will never have NOx inside its optimizer specification and the optimizer specification should therefore be reconsidered.

With the help of some raw data analysis and some initial model analysis, you can get a reasonable understanding of the possibilities.

### Example of response specification in the optimizer

In this example we want to minimize the NOx emission with an upper specification limit (Max) of 10. Setting the target to 0 might seem logical but it is not possible for the optimizer to reach 0 in the investigated experimental region. An inspection of the raw data, see the **Replicate Plot** below, shows that all NOx data are between 10 and 35. Therefore it should be impossible to obtain a prediction close to 0 for the current experimental region. A more reasonable target in this case is 10 for the NOx emission with an upper max limit of 15. The target and max limit values combined with raw data are supported by the replicate plot.



**Replicate Plot:** All values for NOx emission can be found between 10 and 35.

## Optimizer search function

The optimizer will search for a solution to the specifications from different starting points using the Nelder-Mead simplex algorithm and will try to minimize the overall desirability function,  $f(ds)$ . The result is expressed as a normalized distance to target,  $Log(D)$ , and Probability of failure (percentage outside of specifications).

For more information about the computations, see the Optimizer section in the Statistical appendix.

Objective	Setpoint	Alternative setpoints					
		1	2	3	4	5	6
		Air	EGR%	Needlelift	Fuel	NOx	Soot
1	244.4		6.6	-5.202			
2	279.6		6.6	-5.202			
3	244.4		6.6	-0.578			
4	279.6		6.6	-0.578			
5	244.4		11.4	-5.202			
6	279.6		11.4	-5.202			
7	244.4		11.4	-0.578			
8	279.6		11.4	-0.578			

After running the optimizer the best proposal (*i.e.*, the one with the lowest  $\text{Log}(D)$ ) is selected. A  $\text{Log}(D) < -1$  means that all results should be within the specification limits or very close to them. The optimal value of  $\text{Log}(D)$  is -10; in this case all response predictions are on target. Probability of failure gives information about a point's robustness to small disturbances; the disturbances are governed by the precision specified for the factors.

Objective	Setpoint (#13)	Alternative setpoints							
		1	2	3	4	5	6	7	8
		Air	EGR%	Needlelift	Fuel	NOx	Soot	Iterations	log(D)
1	240.008	8.73562	-3.99504	232.621	18.1077	0.535261		66	-0.382708
2	283.994	9.60732	-1.23352	239.051	19.4476	0.658999		91	-0.258037
3	240.002	8.87056	-4.52436	234.777	17.5441	0.498541		91	-0.383884
4	283.998	9.62881	-1.11097	238.508	19.5066	0.682545		90	-0.257997
5	240.002	8.77252	-4.4801	234.519	17.8029	0.486195		64	-0.384194
6	283.856	9.65137	-1.21362	239.019	19.3538	0.673908		232	-0.257807
7	240.004	8.84796	-4.55291	234.869	17.5893	0.491333		132	-0.383881
8	283.985	9.67467	-1.203	238.972	19.3016	0.679721		115	-0.258424
9	240.008	8.73562	-3.99504	232.621	18.1077	0.535261		66	-0.382708
10	283.994	9.60732	-1.23352	239.051	19.4476	0.658999		91	-0.258037
11	240.002	8.87056	-4.52436	234.777	17.5441	0.498541		91	-0.383884
12	283.998	9.62881	-1.11097	238.508	19.5066	0.682545		90	-0.257997
13	240.002	8.77252	-4.4801	234.519	17.8029	0.486195		64	-0.384194
14	283.856	9.65137	-1.21362	239.019	19.3538	0.673908		232	-0.257807

In the Alternative setpoints spreadsheet the user can define start points for the optimization or create a focus search around a specific setpoint (by clicking **New from selected**).

## Robust setpoint

The **Find robust setpoint** tool (FRS) adds the possibility to search for a solution that accounts for model error and factor precision. The main purpose is to describe the effect of perturbations on a possible solution but also to give the user the possibility to elaborate on various kinds of constraints in the result interpretation.

The search for a robust setpoint starts with the creation of the space where all the response specifications are fulfilled (Design Space, DS) according to a specific quality statement (acceptance criterion). The default Acceptance limit is 1% (DPMO = 10 000), which means that 99% of the samples are within limits. For more about the Design Space generation, see the **Design space appendix**.

The precision of the robust optimization procedure depends on the selected resolution. The resolution is the number of discrete blocks which the investigated region is divided into. The FRS procedure fills the DS volume with acceptance blocks according to the selected resolution.

As an example, if the resolution is 16, the factor range will be divided into 16 blocks. Then if the DS volume represents only 3 blocks out of the 16, the robust point will be considered to be the block in the middle of the 3. If the resolution is increased the precision in the result increases to perhaps 4 blocks to the right and 5 blocks to the left in the DS volume and then a more precise estimate of the FRS location is obtained.

The result of the robust optimization is presented in the **Alternative setpoint** list in the optimizer as the **R** alternative. Detailed information is shown in the **Optimizer List** (see below) available using **Create list** (**Tools** tab or right-click).

Selected setpoint									
Factor	Role	Value	Unit	Factor contribution					
Air	Free	245.677	kg/h	35.2164					
EGR%	Free	7.93548	%	38.4595					
Needlelift	Free	-4.66129	BTDC	26.3241					
Response	Criterion	Value	Unit	log(D)	Prob. of failure	Cpk	Cp	k'	
Fuel	Minimize	239.454	mg/st	-0.302063	0.06%	1.16714			
NOx	Minimize	20.3686	mg/s	-0.343069	0%	1.87974			
Soot	Minimize	0.315024	mg/s	-0.295832	0.12%	1.94981			
Robust setpoint (R)									
Factor	Role	Value	Robust low edge	Robust high edge	Unit	Factor contri	Robust resol	Hypercube k	Hypercube h
Air	Free	245.677	240.177	251.177	kg/h	35.2164	4 <-> 4	240	248.516
EGR%	Free	7.93548	6.99798	8.87298	%	38.4595	5 <-> 5	7.16129	8.32258
Needlelift	Free	-4.66129	-5.74504	-3.21629	BTDC	26.3241	6 <-> 8	-5.03419	-4.10194
Response	Criterion	Value	Unit	log(D)	Prob. of failure	Cpk	Cp	k'	
Fuel	Minimize	239.454	mg/st	-0.313155		1.16714		0.704912	
NOx	Minimize	20.3686	mg/s	-0.313155		1.87974		0.673512	
Soot	Minimize	0.315024	mg/s	-0.313155		1.94981		0.318876	
Design space summary									
Resolution	Iterations	DS count	Relative volume %						
32	50000	1875	5.72205						

The *Robust resolution distance* column of the Robust setpoint section is the information about how many discrete blocks in the DS that surround the *Robust setpoint*. In this example, for the factor Air it is 4 to the low limit and 4 to the high limit. Using these discrete blocks the *Robust low edge* and *Robust high edge* estimates are calculated.

If the number of blocks is lower than 4 on each side, the recommendation is to increase the resolution in the FRS function. If the DS is relatively small it may be difficult to obtain a good estimate of the robust setpoint. Here you may be limited to generating the 2D or 4D Design Space plots from the optimizer and analyze them.

For further analysis of the Robust setpoint the Setpoint analysis tool is recommended. For more information, see the Design Space appendix.

# Design space appendix

---

## Introduction

The establishment of a Design Space (DS) is based on the current regression models and an estimation of the probability for failure. In many cases, the resulting DS is a highly irregular volume located inside the experimental design region. It designates where we can expect all response specifications to be fulfilled at a given probability level. The estimation process considers several sources of variability that can affect the size of the DS. Monte Carlo simulations are used to compile the necessary probability statistics.

This appendix outlines a recommended approach for finding a reliable DS. In so doing, an account is given of how MODDE handles sources like model error and precision in factor settings when estimating a DS. Some critical inputs for the calculation setup are also discussed. It is shown how DS results can be presented graphically not only in a 2D contour projection but also in a multidimensional representation.

The main topics discussed are:

- Design space
- Robust setpoint
- Proven acceptable range
- Setpoint analysis
- Setpoint validation.

---

## What is a design space?

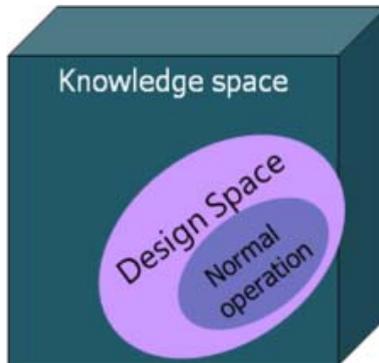
A common interpretation based on the ICH guidelines is that a DS is the region where all specifications are fulfilled at a given probability level.

.....

The “**International Conference on Harmonisation (ICH); Draft Guidance: Q8(R2) Pharmaceutical Development Revision 1**” (<http://www.fda.gov/cber/ich/ichguid.htm>) has outlined quality by design (QbD) principles for pharmaceutical development which introduced the concept of Design Space (DS). ICH Q8 defines DS as “the multidimensional combination and interaction of input variables (e.g. material attributes) that have been demonstrated to provide assurance of quality”.

.....

Conceptually, the DS is a region within the experimental design region. This investigated area is often denoted Knowledge Space (KS).



Some of the elements influencing the size and shape of a DS are the response specifications, the prediction models and the associated uncertainties, and the desired probability level. Since each response may have a unique model expressing its connection to the factors, the resulting DS may be constrained from different directions in factor space. The net result is often a highly irregular DS.

---

## Design space visualization

A key aspect in design space estimation is to be able to provide a graphical representation of the resulting region (area or volume). Unfortunately, however, a mathematical function to describe such an irregular volume becomes very complex, if at all existing. A solution to the problem is to divide the experimental region (knowledge space) into smaller sections and estimate the probability of fulfilling the response specifications in each section.

For example, if the above KS is divided into  $32 \times 32 \times 32$  sections, predictions can be performed to estimate how well the response specifications are fulfilled in each section. In order to accomplish such predictions we use simulations and the regression models. Simulations provide flexibility and enable many sources of uncertainty to be taken into account in the design space estimation. Subsequently, in each investigated section of the KS, the probability of getting predictions that are outside the response specifications is computed.

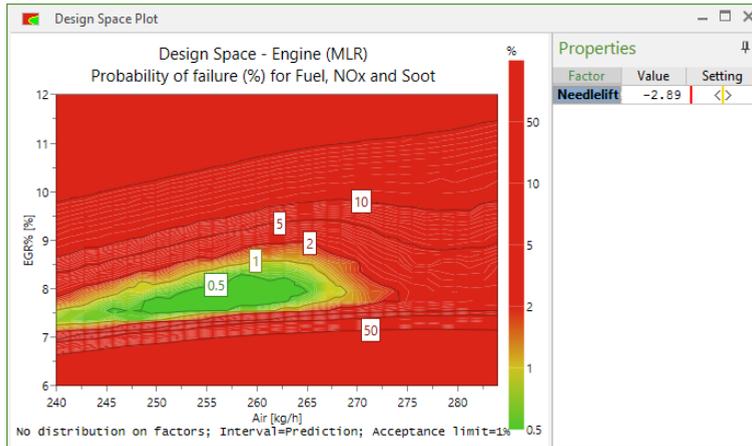
The probability of getting predictions outside the response specifications is quantified as probability of failure expressed in % or DPMO (defects per million opportunities). Using this strategy any type of DS region (area or volume) can be described and it will give a flexibility to use a variety of precision estimates for the factors that are involved in the regression models.

---

*Note: A DPMO of 10 000 equals 99% correct or 1% failure.*

---

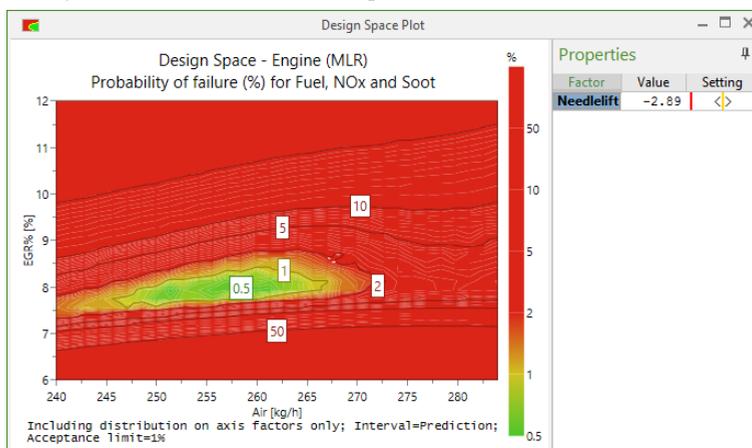
The picture below is a 2D example of a DS defined by 3 models predicting 3 responses and referencing the prediction distributions against the corresponding response specifications. The picture is built up from a grid of predictions in 32\*32 points that have been connected by lines at defined probability levels. The acceptance criterion is set to 1% and lower values are colored green and higher values are colored red.



As mentioned in the Statistical appendix, MODDE supports three interval estimates. The default setting in this context is the *Prediction interval*. The selected interval type will strongly influence the size of the resulting DS region (area or volume).

Furthermore, the size of the resulting DS region will also be critically influenced by which type of disturbance is applied to the factor settings. The DS estimate rendered in the above plot assumes that the factors can be set exactly numerically, and that the only disturbance applied to the factors in the simulations is in line with the size of the model error [ $y = f(x) + e$ ].

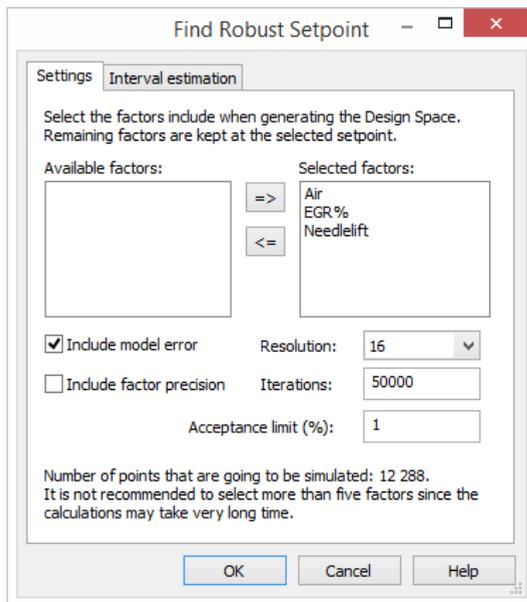
Alternatively, if it is assumed that the factors cannot be set exactly numerically, it is possible to invoke an additional source of uncertainty in the simulations, i.e., according to what is termed **Precision in factor setting** in the Factor Definition dialog box. For instance if Air can be set with a precision of +/- 8 kg/h and EGR% with +/- 0.4 % the DS region will shrink somewhat (see plot below).



Elaboration with various types of factor precision can be done in the Setpoint analysis tool. In the Setpoint analysis changing the precision is done by altering the Std. dev. or Low/High values. If the DS is created from another entrance than **Setpoint analysis** on the **Optimizer** contextual tab, the factor precision used is the one originally specified in the Factor definition dialog box used for the design specification. Adjustment of this precision can be done in the Factor definition dialog box or in the Setpoint analysis.

## In-depth assessment of Design Space

The Design Space (DS) graphs presented in the previous paragraph are simplistic in the sense that they are only two-dimensional representations of a complicated reality. An in-depth and more realistic assessment of a DS can be done by using the **Design Space Explorer** tool available after running the Optimizer. When first launching this tool a **Settings** page is opened, where it is possible to modify default settings for a number of options that have a profound impact on the course of the design space estimation.



As seen, the simulations by default take model error into account but not the factor precision. **Resolution** is the number of sections that each factor range will be divided into. **Iterations** is the number of simulations in each section for each response. **Acceptance limit** is the criterion for specifying the DS region.

If more than 5 factors are available the default selection will be the 5 with the highest factor contribution. The factor contribution is calculated as the prediction variation over all responses and is presented in the optimizer Setpoint tab.

The number of calculations will increase exponentially with the number of factors and the selected resolution. A recommendation is to start with resolution 16 and 50 000 simulations but for a final documentation the recommendation is to use 32 as resolution and leave the computer for an overnight calculation if necessary.

MODDE can address a maximum of  $64^4$  data points, but may take many hours before completion.

Factors/Resolution	Res 8	Res 16	Res 32	Res 64
3 factors	Possible	Possible	Possible	Possible
4 factors	Possible	Possible	Possible	Possible
5 factors	Possible	Possible	Not possible	Not possible
6 factors	Possible	Possible	Not possible	Not possible
7 factors	Possible	Not possible	Not possible	Not possible

When clicking **OK** in the Design Space Explorer dialog box, simulations commence and finally results are presented in terms of a Design Space Explorer window. This window consists of three parts.

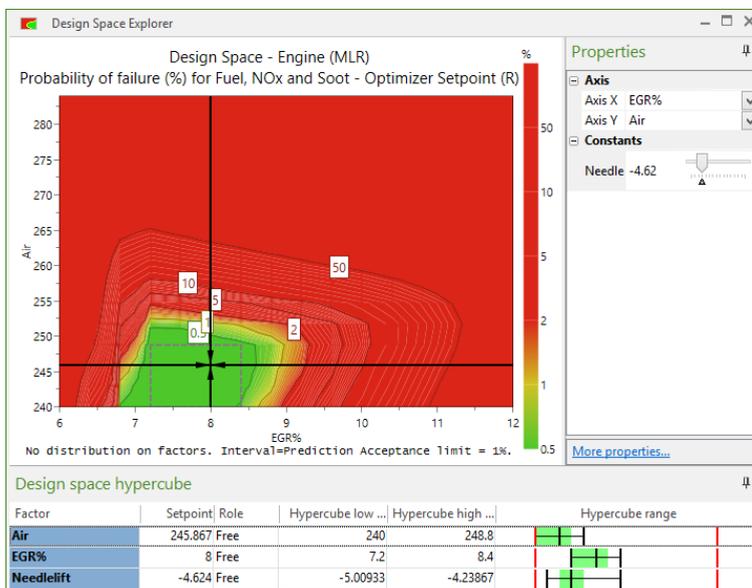
- The main part is a 2D plot showing the size and shape of the DS as viewed in the subspace defined by the two factors chosen for the X- and Y-axes.
- The second part is the Properties pane in which axis factors and constant factors can be selected. For each factor fixed at a constant value a slider is available to facilitate a step-less browsing from one setting to another.
- The third part is a Design space hypercube overview functionality, which is useful when trying to comprehend the size and shape of the investigated DS in the dimensions beyond the 2D representation of the corresponding plot.

Apart from a visualization of the design space in two dimensions, the design space explorer provides two additional graphical tools, i.e.,

- a cross-hair symbol that indicates the position of the robust setpoint and
- a dotted frame that shows the placement of the so-called design space hypercube.

The design space hypercube corresponds to the volume in which all factor combinations can be used without compromising the response specifications. Its extension in the 2D plot is given by the dotted frame, and its elongation across all dimensions is shown by the green color in the Design space hypercube overview. Note that the hypercube range can be changed interactively by clicking and dragging the low or high end for the range of a particular factor.

So, in summary, the green color in the Hypercube range field designates the *mutual ranges* within which all factors can be changed at the same time and without further restrictions. Slightly wider *individual ranges* are patterned by the black T-lines, which represent the allowable range of a process factor, while keeping all other factors constant at their setpoint value.



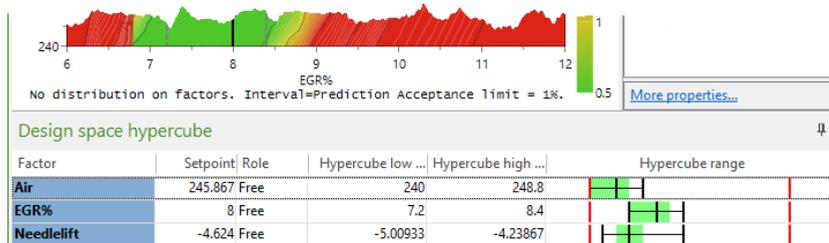
For more comprehensive evaluation and documentation, the design space estimation results can be exported to a text file by Creating a list (**Tools | List** or right-click and click **Create list**). The output is an N-dimensional DS description.

## Proven acceptable ranges

A closely related idea to the design space concept is the notion of *proven acceptable ranges* (PAR) for the factors. ICH Q8 defines a PAR as "a characterized range of a process parameter for which operation within this range, while keeping other parameters constant, will result in producing a material meeting relevant quality criteria". In fact, the PAR notion fits well into the capabilities of MODDE. There are three modes in which to state PAR in MODDE:

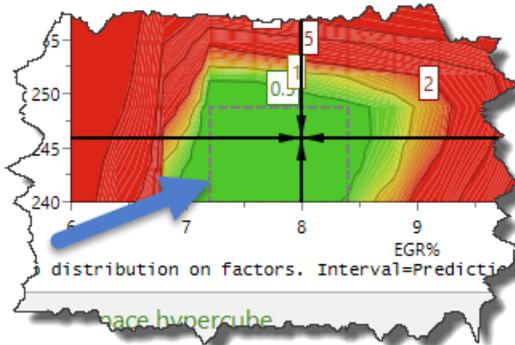
### 1 – Based on the robust setpoint

The hypercube range of the design space explorer is used. The black T-lines extending out from the robust setpoint coordinate indicate the individual ranges for the factors, i.e., the largest allowable range of a process parameter, while keeping all other parameters constant at their setpoint value. A full numerical documentation is available using **Create list** with the Optimizer window active (**Tools** tab or right-click).



## 2 – Based on the dotted hypercube frame

The dotted frame in the Design Space Explorer plot designates the largest possible *regular* hypercube that can be inserted into the *irregular* design space volume. How this regular hypercube extends into many dimensions is given by the green bars seen in the Hypercube range part in the picture above. The green bars mark the *mutual ranges* within which all factors can be changed at the same time and without further restrictions.



## 3 – Based on setpoint analysis

3 is reminiscent of 1, but is based on a *distribution* around a setpoint. Such distributions are attained using the Setpoint analysis functionality in MODDE. The details of the Setpoint analysis tool are given in the next paragraph. Briefly, Setpoint analysis is available after running the optimizer. It outputs the estimated acceptable range for the factors at every given setpoint.

Setpoint Analysis - Optimizer Setpoint (R)							
Factor	Low	Setpoint	High	Std. d...	Role	Distri...	Estimated acceptable range
Air	234.706	240	245.294	2.7009	Free	Normal	
EGR%	7.73143	8.57143	9.41143	0.428571	Free	Normal	
Needlelift	-5.79329	-4.95429	-4.115...	0.428062	Free	Normal	
Response	Min	Target	Max	Criterion	Prob. of failure	Predicted response profile	
Fuel		214.097	250	Minimize	0.04%		
NOx		10.8065	25	Minimize	0%		
Soot		0.01828...	1	Minimize	1%		

Total Prob. of failure: 1.1% (limit: 1%). Samples: 50000. Interval=Prediction

## Correlation in probability of failure

MODDE can calculate the probability of failure  $P(Y_i)$  for the responses based on the model error, and from that the total error for all responses. For example, for 2 responses it would be.

$$1) P_{tot} = P(Y_1) + P(Y_2) - P(Y_1 \cap Y_2)$$

The term  $P(Y_1 \cap Y_2)$  is the probability that both responses are outside limits at the same time. In the normal case when we treat the response residuals as unrelated this can easily be calculated as  $P(Y_1 \cap Y_2) = P(Y_1) * P(Y_2)$

$$2) P_{tot} = P(Y_1) + P(Y_2) - P(Y_1) * P(Y_2)$$

New in MODDE 12 is that that probability can also account for correlation between responses. When we consider correlation between the responses, the last term in equation 1 cannot be calculated so easily. Instead we consider this equation

$$3) P(Y_1 \cap Y_2) = P(Y_1) * P(Y_2|Y_1)$$

$P(Y_2|Y_1)$  is the probability of  $Y_2$  given  $Y_1$ , which is not equal to  $P(Y_2)$  if they are correlated.

In this case, MODDE will estimate the total failure of probability using Monte-Carlo simulations.

Including correlations in the calculation can be useful if there is causality between the responses or when there is an underlying unknown factor that both responses depend on but in most cases the correlation between the response residuals are just due to random errors and should be ignored.

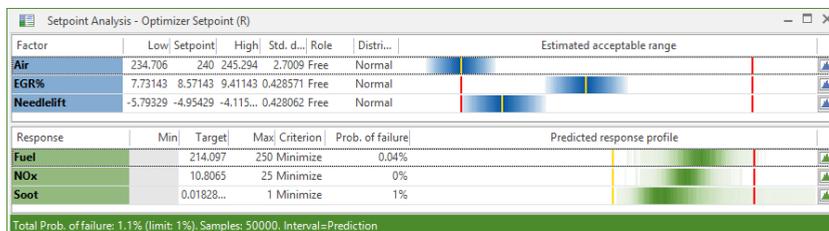
To calculate the correlated probability MODDE generates random values from a multivariate normal distribution based on model errors and the correlation between responses and from that calculates how many point fall outside specifications. Enable it by selecting **Yes** in **File | Options | Investigation Options, Correlation in probability of failure**.

In most cases the difference is too small to notice because we are generally interested in the areas with very low probability. And when  $P(Y_1)$  and  $P(Y_2)$  is small,  $P(P(Y_1) \cap P(Y_2))$  will be extremely small. Also the limits on the responses are more likely to be in different places in the X space, so where  $P(Y_1)$  is significant  $P(Y_2)$  is close to 0 or so high it is not interesting.

## Setpoint analysis

Setpoint analysis is available after running the optimizer. This tool can estimate the tolerance available for the factors at every given setpoint. It is an excellent platform for analysis of any given setting and also for imposing practical adjustments to the allowable factor ranges and evaluating the consequences of such changes.

In the presentation below the robust setpoint is the start, and from that co-ordinate the factor ranges are expanded until the Acceptance limit criterion is reached.



Factor variations possible to use are presented as Estimated acceptable ranges. Various types of distributions and interval estimates can be used in the estimation; the default is a normal distribution and a prediction interval.

In this example, the responses that limit the widening of the factor ranges are Fuel and Soot as these responses first reach the Acceptance limit of 1%. All estimations are based on the regression models and with the prediction interval.

## Properties - Setpoint analysis

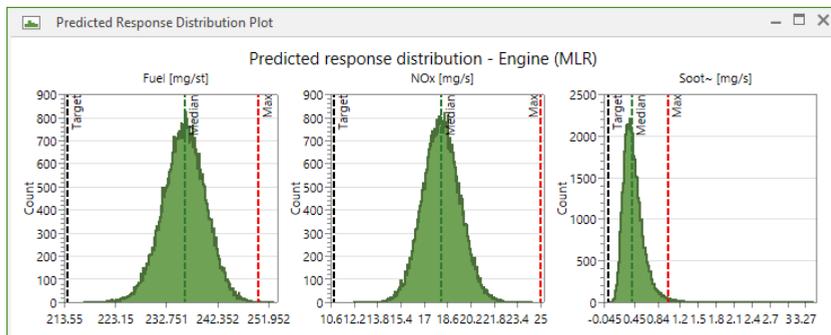
In the property page settings can be adjusted. The most important ones are the Interval type used for the predictions and the Acceptance limit. Because the predictions are based on simulations the stability in the results is dependent on the number of simulations. The default settings aim to give a reasonable estimate within a quite short time. The result can thus vary between simulations. To get a more stable result for documentation we recommend increasing the **Acceptable range simulations** and **Response profile simulations**.

### Monte Carlo simulations

The **Monte Carlo** simulations are:

- random factor settings according to the selected distribution,
- around their optimum value but within the Low and High limits,
- followed by predictions of the responses. In this case 50 000 predictions are performed. The distribution as well as the number of simulations and the range can be changed by the user.

The resulting distributions can be presented as a histogram, one for each response.

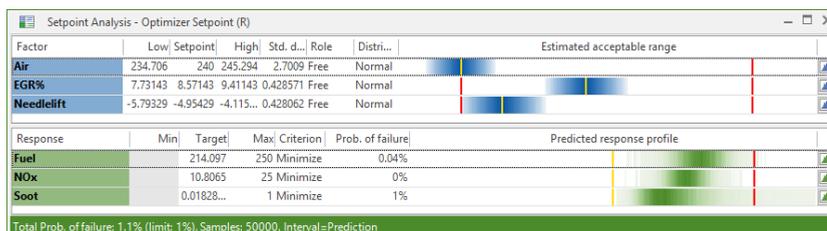


### Evaluate the results and make necessary adjustments

If the result in some way is not satisfactory, one option can be to change the starting point for the DS search (selected setpoint). Another option is to lock some factor ranges where setting tighter specifications won't be a problem. For more, see the Setpoint validation for robustness testing subsection later in this chapter.

For alternative start points it can be preferable to step back to the optimizer and select another start point based on Log(D), Probability of failure and Design Space evaluation. The modeling results, e.g. Coefficient Plot, can also be an information source to finding alternatives. All settings can be changed in the Setpoint analysis window for a user controlled search in order to find the most appropriate solution.

The setpoint shown below is a movement of the factor setting Needlelift to a higher setpoint.



The alternative factor settings give results within specifications but with narrower estimated acceptable ranges than for the previously proposed starting point. The alternative factor settings might be preferred for practical reasons.

## How to find the best Design Space

1. Develop the best model for each response.
2. Check the residual distribution for outliers or unresolved structures.
3. Find the optimal settings for the factors that comply with the response criteria (optimizer).
4. Check if the proposed optimal factor settings are critical (close to a limit) or in a safe region.
5. Perform a Robust Setpoint analysis.
6. Decide on the interval type and the acceptance level; default settings are Prediction interval and Probability of failure 1%.
7. Make an evaluation of the complete Design Space with Design Space Explorer.
8. Decide on how to express the PAR, from a setpoint or as a selected hypercube inserted into the Design Space.
9. Evaluate the results and make necessary adjustments using DS analysis and Setpoint analysis.
10. Set your preferred factor specifications with any necessary practical adjustments.
11. Document the final results.

---

## Setpoint validation for robustness testing

According to the FDA: "VALIDATION OF ANALYTICAL PROCEDURES: Definition and terminology. The robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage."

Setpoint validation is a way to test if the system investigated is robust against disturbances in the investigated region. Setpoint validation can be found on the Predict tab.

The aim of robustness testing is to evaluate if a process, or a system, performs satisfactory even when some influential factors are allowed to vary. In other words, we want to investigate the system's sensitivity (or preferably lack of sensitivity) to changes in certain critical factors. The advantages with having a robust process or system include simpler process control, a known range of applicability and an ensured quality of the product or process.

A robustness test is usually carried out before the release of an almost finished product, or analytical system, as a test to ensure quality. Umetrics recommends the use of DOE for robustness testing and such a design is usually centered on the factor combination, which is currently used for running the analytical system, or the process. We call this the setpoint. The setpoint may have been found through a screening design, an optimization design, or some other identification principle, such as written quality documentation. The aim of robustness testing is, therefore, to explore robustness close to the chosen setpoint.

In Setpoint validation we use simulations on the regression model and simulate random disturbances within the investigated range of operation for all factors. The regression model originates typically from a low resolution design supporting linear models since we assume that small disturbances have mainly linear effects. Fractional factorial resolution III and Plackett Burman designs are recommended.

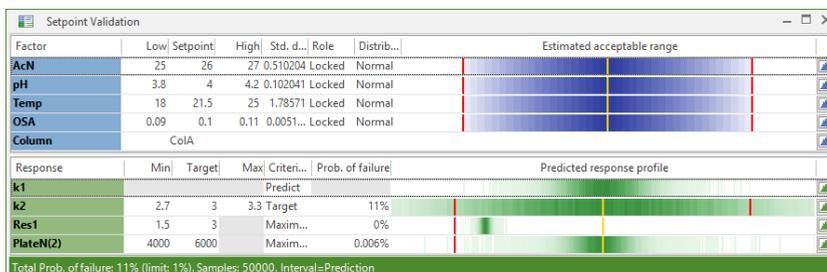
### Setpoint validation example

In this example we show that the DOE strategy in combination with Monte Carlo simulations here gives a proper estimate of the system's robustness.

The investigation chosen to illustrate the **Setpoint validation** feature originates from a pharmaceutical company. It represents a typical analytical chemistry problem within the pharmaceutical industry. In analytical chemistry, the HPLC method is often mounted for routine analysis of complex mixtures. It is therefore important that such a system will work reliably for a long time, and be reasonably insensitive to varying chromatographic conditions. For details about this example, see the tutorial "Robustness testing".

### Evaluation of the setpoint validation

The **Predict | Setpoint validation** tests the robustness by making a large number of random disturbances in the specified region with a selected confidence criterion. In this example the specified region is the experimental region or knowledge space (KS). In the Setpoint Validation window the factor part shows the original investigation settings with a specific setting for the qualitative factor Column. ColA was manually selected as this selection gave the worst results (see Tutorial example for more information about the specific example). The result is shown as a distribution of random samples including model prediction errors and it is well within the specification limits for some responses. The result can be expressed in general statistics as well as capability indices Cpk or Probability of failure.



A description of the details of this window is found in the Setpoint properties section in Chapter 13, Setpoint.

### Factor spreadsheet

All factors are varied within the design limits with Monte Carlo simulations according to a *Normal* distribution. These are the default settings and can be changed.

### Response spreadsheet

The result for response k1 is optional; there are no specific demands for this response.

The result for response k2 is outside the specification limits.

The result for response Res1 is above the low specification limit.

The result for response PlateN(2) is above the low specification limit

From the above we conclude that this system is robust against disturbances in the factors for Res1 and PlateN(2). k2 is not robust against disturbances in the factors.

### Final adjustments

Setpoint validation can be used to estimate the maximum accepted variability in factors that still predict all results within the specifications.

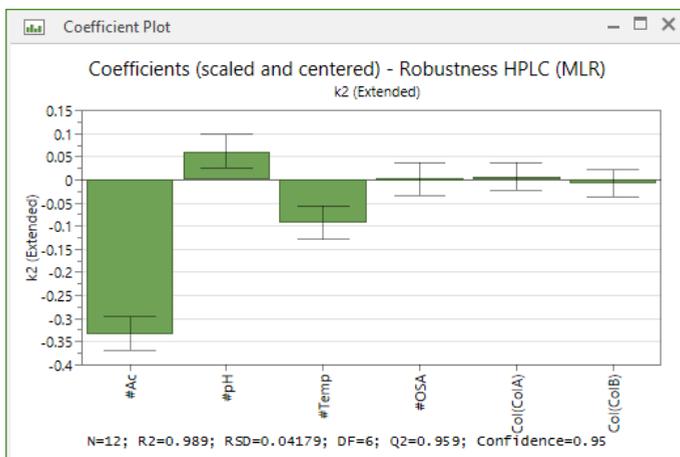
The problem in the described example is response k2. The requirement for k2 is that less than 1% of the predictions may be outside the specification limits.

There are constraints when handling this type of situation;

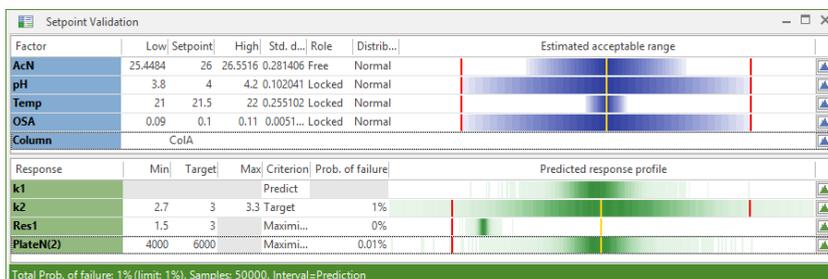
- Which factors affect the result?
- How can we adjust the factor limits without causing too much problems in the normal use of the procedure?

First we have to check the model to understand which factors are the most influential. The model has to be significant for an adjustment in factor ranges to have an effect on the result distribution.

In this example the model for k2 is very significant and the most important factor is Acetonitrile (ACN).

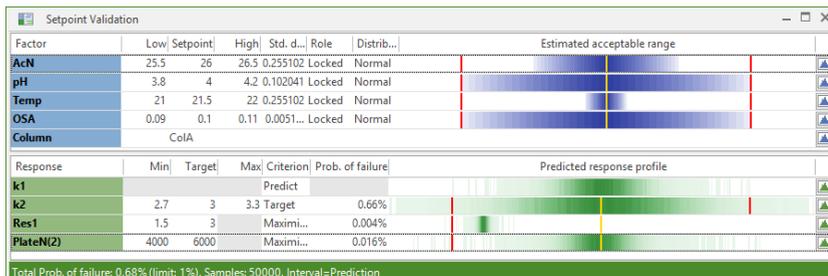


Assuming that the factor Temperature is easy to control with a narrower range we start with this factor by adjusting temperature to  $\pm 0.5$  °C. At the same time we can open the **Role** for ACN to *Free*. This instruction together with the specification limit for k<sub>2</sub> (Probability of failure = 1%) will give an estimate of a range for ACN where we can predict that the system is robust according to the specifications. The picture next here displays the result of the change in settings.



The proposed settings for ACN are now 25.45 to 26.55 and the estimated distribution for k<sub>2</sub> is 1% hits outside the specification limits.

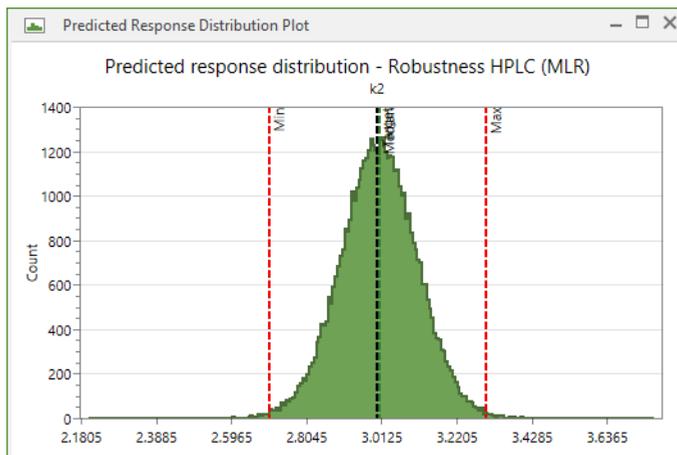
A final step might be to make an adjustment of the factor settings to some practical new specification within the range for ACN, for instance 25.5 to 26.5. The result shown below implies that the critical response k<sub>2</sub> will have 0.66% of future predictions outside the specifications.



### Result statistics for k2

To open the response histogram, click the histogram symbol to the far right in the Setpoint validation window.

In the **Select responses** box, select which responses to display.



To view statistics, create a list from the **Setpoint validation** window (**Tools** tab, or right-click).

1	Factor	Low	Setpoint	High	Distribution	Possible min	Possible max	Experiment	Experiment						
2	AcN	25.5	26	26.5	Normal	25.15	26.9	25	27						
3	pH	3.8	4	4.2	Normal	3.71	4.28001	3.8	4.2						
4	Temp	21	21.5	22	Normal	16.425	26.375	18	25						
5	OSA	0.09	0.1	0.11	Normal	0.086	0.114	0.09	0.11						
6	Column		ColA												
7	Confidence for normal distributions:		95												
8															
9	Response	Prob. of fail	Cpk estimat	Cp estimate	K' estimate	% out of rnr	Average	Median	1st quartile	3rd quartile	Std. dev.	Skewness	Df		
10	k1						2.01157	2.01126	1.9409	2.08267	0.108536	0.00235563	6		
11	k2	0.66%	0.917604	0.930632	0.0139991	0.664	3.0042	3.00428	2.89348	3.07534	0.107454	-0.0173028	6		
12	Rev1	0.004%	2.88641		0.793243	0.004	1.81164	1.81164	1.78993	1.83344	0.0359888	-0.0321665	6		
13	PlateN(2)	0.016%	2.13384		0.0180393	0.016	5963.92	5963.52	5777.9	6140.7	306.789	-0.0232081	6		
14	Model error included in predictions:		Yes												
15	Number of samples:		50000												

A more detailed description of this example is found in the tutorial named "Robustness Testing".

# Generalized subset designs

## appendix

---

### Introduction

When quantitative and qualitative factors are specified to have two levels, the design generation is straightforward. There are then several classical design families to choose from, for instance full factorial designs, fractional factorial designs, and Plackett-Burman designs. However, when quantitative and qualitative factors at three, four, or even more levels are involved in a DOE investigation, a challenge arises in the sense that the number of factor combinations grows substantially, far beyond what is required by e.g. a two-level fractional factorial design. For practical reasons, there may then be a need to limit the number experiments required.

The generalized subset designs, GSD, is a new entry in MODDE providing a possibility to accomplish reduced designs even when handling multiple multilevel factors. This design setup generates a series of reduced designs, subsets, that are logically linked, such that, when combined, all subsets will add up to a full multilevel multifactorial design where all factor combinations are encoded by the global design. Conceptually, the output of GSD is similar to how two-level fractional factorial designs represent complementary reductions of two-level full factorial designs.

The novelty in GSD lies in the fact that the global design is generated as a set of reduced designs that are as perfect complements to each other as possible. However, the individual subsets can be used as standalone fractional factorial designs since the criteria are to generate design sets as orthogonal, equivalent and balanced as possible. Each design set is an integer fraction of all possible combinations, i.e.  $\frac{1}{2}$ ,  $\frac{1}{3}$ ,  $\frac{1}{4}$ , etc. fraction of all factor combinations.

There are a number of applications where GSD is superior:

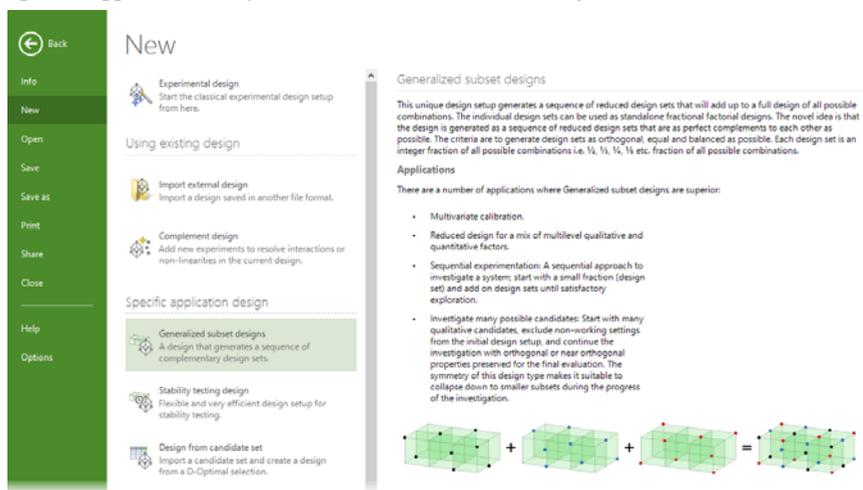
- Stability testing design
- Multivariate calibration
- Reduced design for a mix of multilevel qualitative and quantitative factors
- Sequential experimentation: A sequential approach to investigate a system; start with a small fraction (design set) and add on design sets until satisfactory exploration.
- Investigate many possible candidates: Start with many qualitative candidates, exclude non-working settings from the initial design setup, and continue the investigation with orthogonal or near orthogonal properties preserved for the final evaluation. The symmetry of this design type makes it suitable to collapse down to smaller subsets during the progress of the investigation.

## Generalized Subset Designs example

Generalized Subset Designs, GSD, create a series of reduced design sets that can be used individually or in combination. The objective for each individual design set is to be a balanced unique representative sample of all possible combinations. In this example we have data for all combinations and will compare the result using the complete GSD with the use of one or several design sets (subsets).

The data we use here are from a packaging robustness verification where we want to investigate the storage temperature effect. The expectation is to fulfil the specification of 95% active substance content after the investigated time frame and prove that the factor ranges investigated will not cause any specific degradation that can violate the specification.

In order to launch Generalized subset designs, on the **File** tab, click **New**, and under Specific application design, click **Generalized subset designs**.



## Factor and response setup

In the design wizard all factors to be investigated are defined. Our example deals with the four factors *Temp* (temperature in the storage facility, multilevel quantitative at three levels), *Strength of API* (Active Pharmaceutical Ingredient, quantitative, two levels), *Container* (multilevel quantitative representing four sizes), *Batch of API* (qualitative, three batches). A full factorial design in *Temp*, *Strength*, *Container*, and *Batch* corresponds to 72 (=3\*2\*4\*3) factor combinations.

Design Wizard

Define factors

Factors: 4 (4 process, 0 mixture)

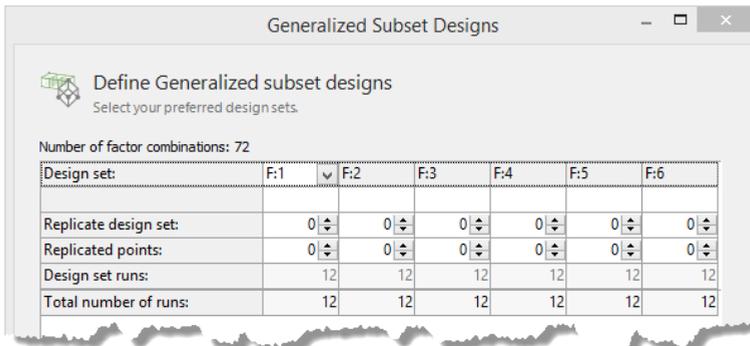
	Name	Abbr.	Units	Type	Use	Settings	Transform	Precision
1	Temp	Temp	°C	Multilevel	Controlled	20; 30; 40	None	0.925
2	Strength	Str	mg	Quantitative	Controlled	10 to 40	None	0.75
3	Container	Con	ml	Multilevel	Controlled	3; 30; 100; 1000	None	24.9
4	Batch	Batch		Qualitative	Controlled	Batch 1; Batch 2; Batch 3		

The next step in the design wizard deals with response definition. For each response it is possible to enter specifications in terms of Min, Target and Max values. Only one response is used in our example, *Content API*, which has a Min specification of 95%. This Min setting refers to the desired amount after storage.



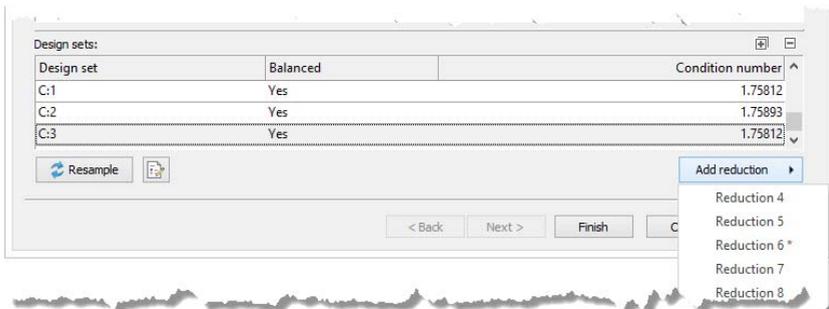
## Define subset designs

When factors and responses have been defined, the next step corresponds to setting up the subset designs. This is done using the **Define generalized subset designs setup** page.



You can interactively change which reduction to use. Moreover, it is possible to generate new reductions and select any set of subset designs deemed appropriate. Clicking **Add reduction** enables creating the reduction of your choice which can then be selected as Design set. The star, e.g. 6\*, signifies that the full design may be parted in balanced design sets.

In our example, we have selected reduction 6 which resulted in balanced equally large design sets. Replicate design set and Replicated points can also be changed at this stage.

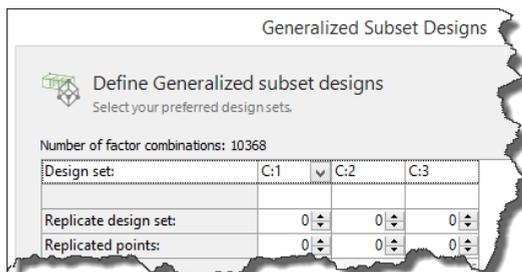


The lower part of the design setup page shows which reductions are eligible and how many runs they contain. By clicking **Expand all** (the plus sign) more details are provided.

For each design set, the augmented list of details informs about whether the design set in question is balanced or not, its condition number and the number of runs it contains.

## Replicates and Center points

In the Number of factor combinations section you can specify number of times to *Replicate design set* and number of *Replicated points/Center points* for each design set.



Note here that when there are only qualitative and/or multilevel factors, the rows labeled as center points ('cp' in the Combination ID) in the Stability testing designs use the first qualitative setting and one of the middle settings for the multilevel factors while replicated point ('rp') for GSD use any level to replicate.

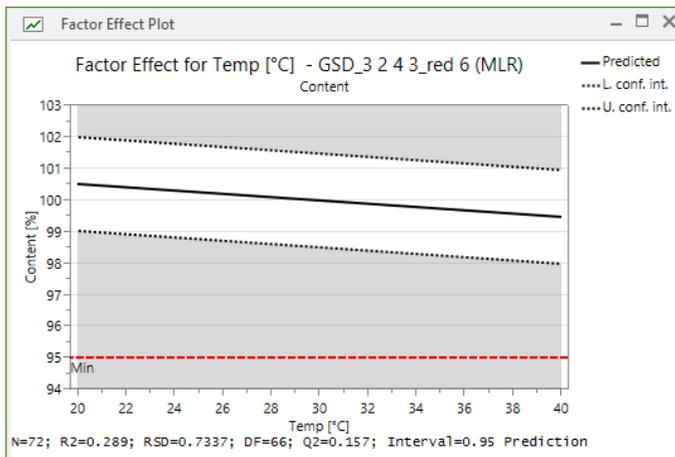
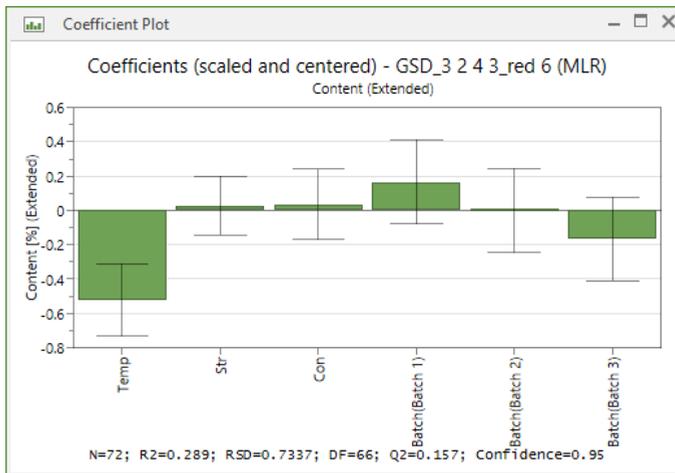
For Stability test designs each design set (A, B, C etc.) includes all available combinations of the factor settings. This means that, if there are no quantitative factors, the center points will be replicates of another experiment for that design set. This also means that, for the other time points the experiment matching the added "center point" are marked 'cp' too.

## Starting with balanced subset

First question: Will it be possible to get a model that can predict the outcome with a reasonable precision using all possible combinations (72)?

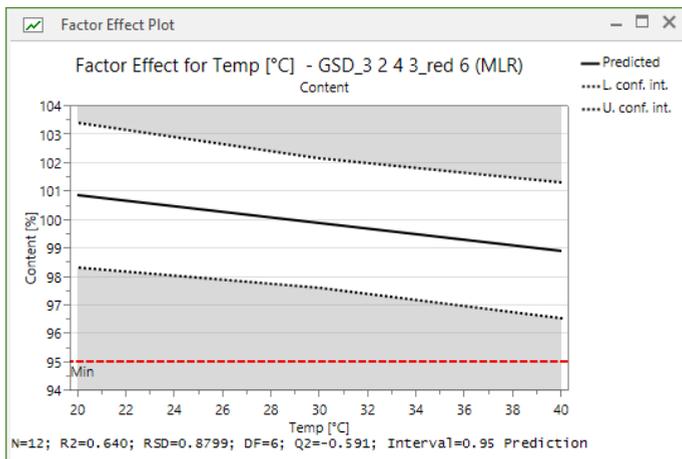
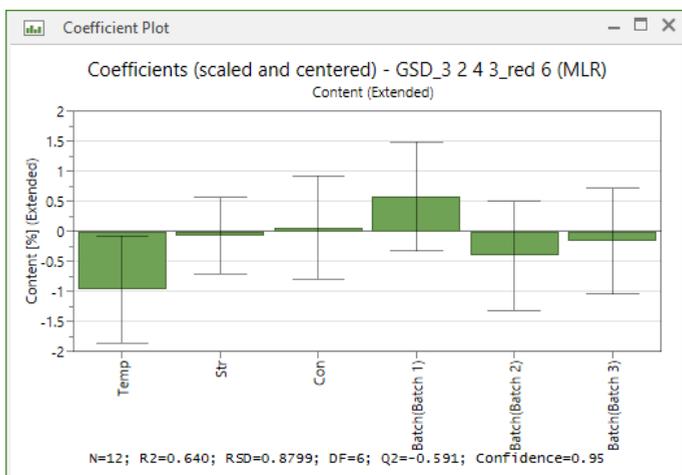
Second question: Is it possible to predict and detect the significant degradation with sufficient precision using only 12 samples of the complete set of 72?

In the first Coefficient and Factor Effect plots here, all 72 combinations were included.



In the next Coefficient and Factor Effect plots only the first design set was included.

*Note: In Generalized subset designs it is common to start with one design set. This means that the qualitative factor Design set will be constant for that factor. MODDE then automatically excludes the factor and when you add design sets you have to manually add the Design set factor to the model.*



Comparing the model with all 72 samples to the selection of 12 gives the same information in the coefficient plot, Temp is a significant factor. The prediction plot can based on those few data still predict that the Content amount will be within specification.

Continuing with another design set, adding another 12 samples, results in plots expressing the same pattern as those above.

## Conclusion

In this example, the evaluation of the full design clearly indicates that the objective of maintaining above 95% API after the robustness test can be met. With a balanced subset selection it will be possible to start the evaluation with rather few sample of all possible combinations and add on subsets until a sufficient and stable interpretation of data can be achieved. Each added subset will keep the balance in data and become the best possible sequential complementary strategy.

---

## Evaluation after excluding setting

Third question: Will it be possible to maintain the original design properties while collapsing the space by excluding a specific setting that will be uninteresting to continue to investigate further?

Scenario; in the case described above we discover after analyzing the first set of data (12 experiments) that batch 3 is very different and not according to expectations (different from batch 1 and 2). We want to continue the investigation by analyzing design set 2 and 3 i.e. adding 12 + 12 more experiments without analyzing the data originating from batch 3.

The easiest way to compare the statistical properties of a design is to compare the condition number, a perfect design will have a condition number of 1.0 a very good design will have a condition number below 5.

Design sets	Original setup cond. no. (no. experiments)	Batch 3 excluded cond. no. (no. experiments)
1	2.32 (12)	3.17 (8)
2	2.07 (12+12)	1.67 (8+8)
3	2.05 (12+12+12)	1.81 (8+8+8)

The properties of the design (condition no.) will remain very good while collapsing the design by excluding observations that includes batch 3. A small increase is observed in the smallest selection due to very few experiments present.

---

## Multivariate calibration, GSD

Multivariate calibration involves predicting sample properties Y (concentrations / levels / qualities / compositions, etc.) from sample measurement characteristics X (often spectral data). After spectroscopic measurements of the samples (with known qualities or concentrations), a regression model is built using the spectral data as X and the sample qualities or concentrations as Y. In the next step, equivalent X-data are measured for additional, “unknown” samples and by inserting these data into the regression model, predictions of the Y-data for the fresh samples is possible. Because of the frequent need for predicting sample qualities or concentrations of one or several analytes in unknown samples, multivariate calibration has found widespread use and acceptance in analytical chemistry, and other analytically oriented sciences. One typical objective of multivariate calibration is often to replace a target or reference method, which may be time-consuming and laborious, with an alternative measurement technique, often spectroscopic in nature, which should be fast, precise, and preferably also non-destructive to the analyzed samples.

One of the most important aspects of multivariate calibration is the consideration of which samples to use for establishing the calibration relationship. Ideally, the calibration set should be representative and diverse; i.e. this set of samples should include all variations in X and Y that are relevant to the problem, and which are expected to occur over time. It is particularly important that the calibration set samples span the range of analyte concentrations and also the relevant concentration ranges of interferences. This can be accomplished by using different approaches, for instance by means of design of experiments (DOE), sampling strategies, and combinations thereof.

When DOE is used to determine the composition of the reference samples at least 5 concentrations levels are needed for each analyte. However, recommendations involving as many as 7 to 9 concentrations levels per analyte are also found in the literature [ref Brereton]. The induced change in the concentration/level/proportion of each sample constituent then defines a separate Y-variable in the multivariate calibration modeling.

The use of DOE in multivariate calibration need not only involve systematic changes in the levels or concentration of the analytes. On the contrary, a way of making the calibration model more robust to batch-to-batch variation in the main constituents (or other undesirable source of variation) consists of imposing simultaneous changes to the levels of the constituents and the settings of occurring process and/or matrix factors.

Because applications of DOE in multivariate calibration easily involve many factors at many levels the novel GSDs represent a viable alternative for reducing the number of experiments required, yet still providing results that are reliable and has a representative set covering all factor settings.

---

## Stability testing design

A pharmaceutical product in storage may change its quality characteristics with time. Hence, it is important to know how well a product retains its quality characteristics over the life span of the product. A product is considered stable as long as its quality characteristics remain within specifications. The shelf life of a product corresponds to the number of days it remains stable at the recommended storage conditions. The process of collecting experimental data for estimating and verifying a product's shelf life is called *stability testing*.

As outlined by the ICH guidelines (1, 2), design of experiments (DOE) is a cornerstone technology in stability testing. This because DOE enables to spread out informative experiments in the critical factors across the time span that is encompassed by the stability test. The normal duration of a stability test is 36 months, with testing time points occurring at 0, 3, 6, 9, 12, 18, 24 and 36 months.

In addition to the obvious Time factor, stability testing often involves between 2-5 additional factors, for instance strength, container size and/or fill. These factors can be quantitative or qualitative in nature and be studied at two or more levels. As with regular DOE protocols, the number of factor combinations grows rapidly with increasing number of levels of the factors. A full study stability test is one in which all factor combinations are tested for every time point. It is easily realized that this may not be a suitable design approach especially when the number of factor combinations is large. Instead reduced designs can be used, in which only a subset of the factor combinations is tested at a given time point.

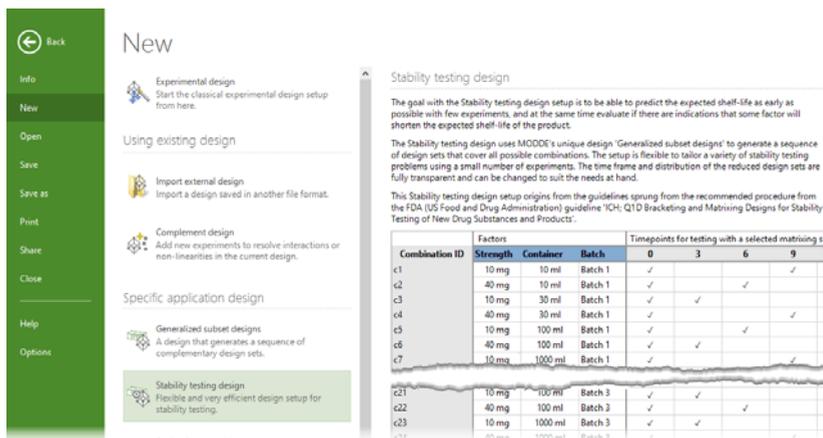
The prevailing approach to stability testing implies that full testing of all factor combinations is done at 0 and 36 months, and that reduced testing is carried out at all other time points (3, 6, 9, 12, 18 & 24 months). Reduced testing means that only subsets of the factor combinations are tested. Various reductions can be used and common reductions are the one-half reduction and the one-third reduction. This approach is known as *matrixing*.

Ideally, the various reductions tested at 3, 6, 9, 12, 18 & 24 months should be complementary, so that all factor combinations indeed are investigated at least once in each type of reduction. Furthermore, it is advantageous if all levels of any qualitative factor are exploited the same number of times in each reduction. This last point is often referred to as balancing. Balancing subsets of factor combinations so that they are perfectly complementary inside each type of reduction is the real challenge in stability testing, and although balancing is worth striving for it is not always straightforward to accomplish.

With the release of MODDE 11, and further developed in MODDE 12, Umetrics introduced a new class of reduced combinatorial designs that are perfectly complementary when superimposed on top of one another. This new design family is described in the Generalized subset designs appendix. In this section we shall describe how the so called generalized subset designs are especially suited to the needs of stability testing, and how they bring a vast array of stability testing scenarios into one common framework. The data set used is illustrative of the main principles for data analysis.

## Setting up a stability test

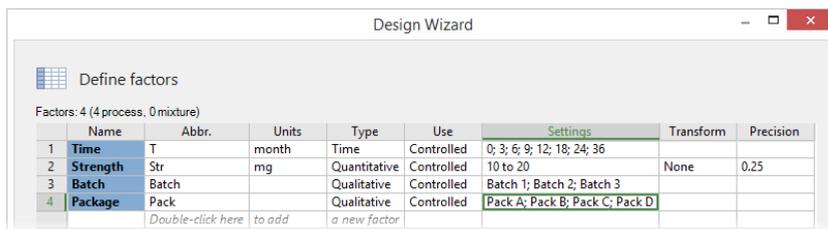
In order to launch a stability testing design, on the **File** tab, click **New**, and under Specific application design, click **Stability testing design**.



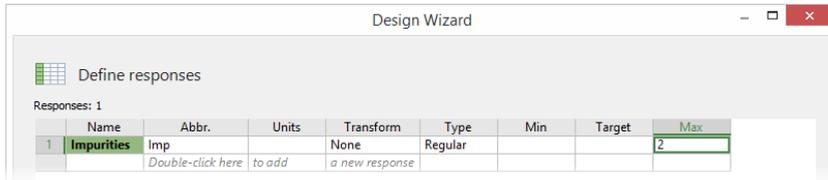
This will open up the Design Wizard, with one factor (Time) pre-defined. As shown below, the default settings for Time are 0, 3, 6, 9, 12, 18, 24 and 36 months. The settings of the Time factor can be changed manually and up to 24 settings (time points) can be entered.



In the design wizard all factors to be investigated are defined. Besides *Time* our example deals with the three factors *Strength of API* (quantitative, two levels), *Batch of API* (qualitative, three batches) and *Primary packaging of tablet* (qualitative, four packaging variants). A full factorial design in *Strength*, *Batch* and *Primary packaging* corresponds to 24 (=2\*3\*4) factor combinations.



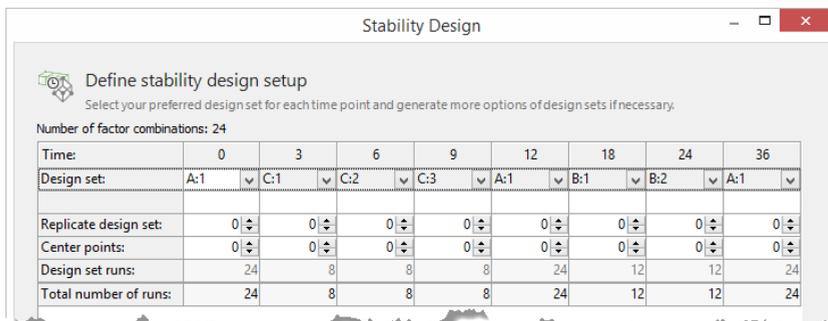
The next step in the design wizard deals with response definition. For each response it is possible to enter specifications in terms of Min, Target and Max values. Only one response is used in our example, the amount of impurities, which has a Max specification of 2%. This Max setting refers to the desired situation at the last studied time point.



## Define stability testing design setup time point design sets

When factors and responses have been defined, the next step corresponds to setting up the design sets. This is done using the **Define stability testing design setup** page. The basis for this page is a timeline running from left to right where all time points previously set for the Time factor are listed as separate columns.

In our example, full testing of all factor combinations (24 runs) is proposed for 0, 12 and 36 months, whereas reduced testing is proposed at all other time points. We can see that two reductions are suggested. Two half-reductions of 12 runs – denoted design sets B:1 and B:2 -- are scheduled for 18 and 24 months, and three one-third reductions of 8 runs -- denoted design sets C:1, C:2 and C:3 -- are scheduled for 3, 6 and 9 months.



You can interactively change which reduction to use at a given time point. Replicate design set and center points can also be changed at this stage.

Moreover, it is also possible to generate new reductions and distribute them across time as deemed appropriate. Other reductions may be added by clicking **Add reduction** in the bottom part of the dialog. Up to 9 reductions are supported provided the average number of experiments per design set is rounded to 4 or more.

Stability Testing Design

---

**Define stability testing design setup**  
 Select your preferred design set for each time point and add reductions if necessary.

Number of factor combinations: 24

Time:	0	3	6	9	12	18	24	36
Design set:	A:1	C:1	C:2	C:3	A:1	B:1	B:2	A:1
Replicate design set:	0	0	0	0	0	0	0	0
Center points:	0	0	0	0	0	0	0	0
Design set runs:	24	8	8	8	24	12	12	24
Total number of runs:	24	8	8	8	24	12	12	24

Design sets:

Design set	Balanced	Condition number
<input checked="" type="checkbox"/> Reduction: A, 1 (runs: 24)		
<input checked="" type="checkbox"/> Reduction: B, 2 (runs: 12, 12)		

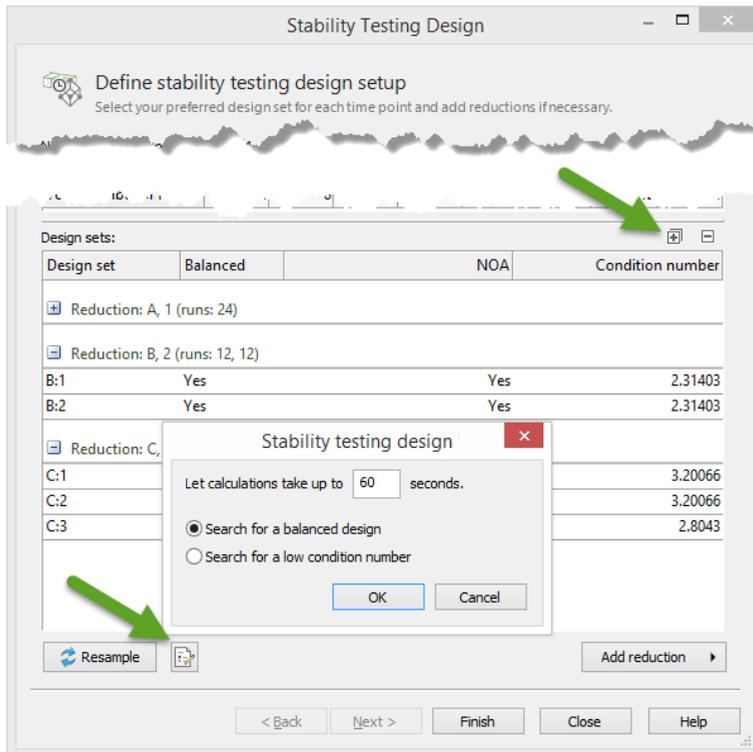
Add reduction

- Reduction 4
- Reduction 5
- Reduction 6

The lower part of the stability testing design setup page shows which reductions are eligible and how many runs they contain. By clicking **Expand all** (the plus sign) more details are provided.

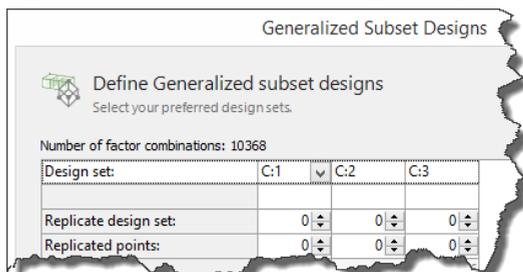
For each design set, the augmented list of details informs about whether the design set in question is balanced or not, its condition number and the number of runs it contains.

If you are not satisfied with the condition number or balance of the design sets, mark the Reduction and click **Resample** at the bottom of the dialog window. In order to let the algorithm take more than the default 60 seconds to search for the best design sets, or shift the aim to search for the lowest condition number, click the Settings-button next to **Resample** and make your change. Then mark the reduction and click **Resample**.



## Replicates and Center points

In the Number of factor combinations section you can specify number of times to *Replicate design set* and number of *Replicated points/Center points* for each design set.



Note here that when there are only qualitative and/or multilevel factors, the rows labeled as center points ('cp' in the Combination ID) in the Stability testing designs use the first qualitative setting and one of the middle settings for the multilevel factors while replicated point ('rp') for GSD use any level to replicate.

For Stability test designs each design set (A, B, C etc.) includes all available combinations of the factor settings. This means that, if there are no quantitative factors, the center points will be replicates of another experiment for that design set. This also means that, for the other time points the experiment matching the added "center point" are marked 'cp' too.

## Stability testing design worksheet

By clicking **Finish** on the stability testing design setup page, the worksheet of the stability testing design is created. In our example, the worksheet comprises 120 runs. An excerpt is shown below. Compared with the worksheet of a conventional DOE protocol there are some minor differences. First of all, there is a new column, the second from left, denoted **Combination ID**. There will be as many Combination IDs as there are factor combinations (excluding the Time factor), i.e. 24 in our example; 25 if there are center points. Secondly, the column denoted **Exp Name** has a modified naming structure for each experimental run, a modified structure which indicates which factor combination and time point applies for each row in the worksheet. Finally, there are two list boxes which enable subset selections of the experiments. Subset selections can be accomplished based on the factor combinations and time points.

1	2	3	4	5	6	7	8	9	10
Exp No	Combination ID	Exp Name	Run Order	Incl/Excl	Time	Strength	Batch	Package	Impurities
1	c1	c1-T0	1	Incl	0	10	Batch 1	Pack A	
2	c2	c2-T0	2	Incl	0	20	Batch 1	Pack A	
3	c3	c3-T0	3	Incl	0	10	Batch 2	Pack A	
4	c4	c4-T0	4	Incl	0	20	Batch 2	Pack A	
5	c5	c5-T0	5	Incl	0	10	Batch 3	Pack A	
6	c6	c6-T0	6	Incl	0	20	Batch 3	Pack A	
7	c7	c7-T0	7	Incl	0	10	Batch 1	Pack B	

In Stability testing designs it is common to start with one design set. This means that the qualitative factor Design set will be constant for that factor. MODDE then automatically excludes the factor and when you add design sets you have to manually add the Design set factor to the model.

## Early stage data analysis – trajectory trending, Stability testing design

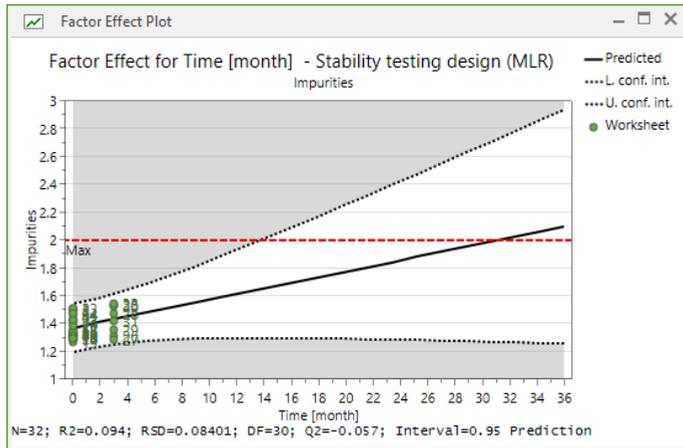
The default data analytical method is MLR, which is a sound choice provided that there are no major disturbances to the geometry of the stability testing design. Data analysis of measured data can be done at any stage of the duration of the stability test. In the early stages, say after 6 or 9 months, the focus lies on trending, i.e., understanding what will likely be the development trajectory of the response with time. This kind of assessment is done with a simple regression model, usually involving only *Time* as the single model term.

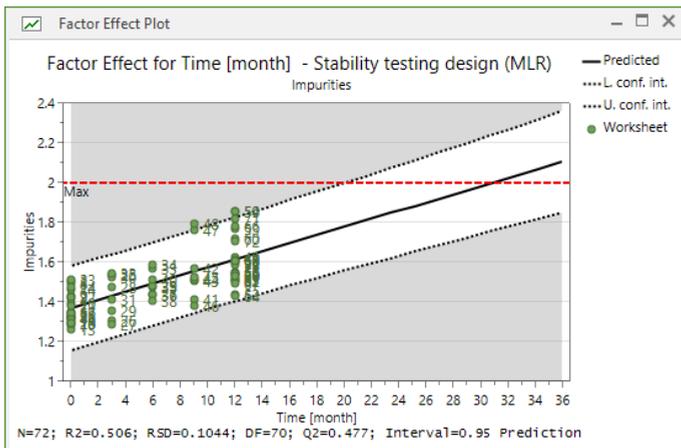
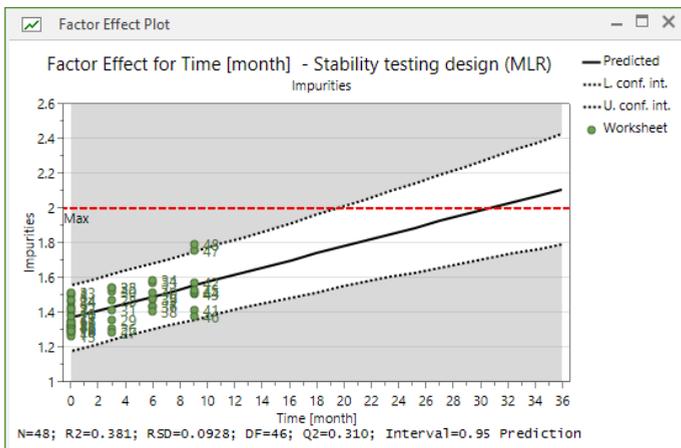
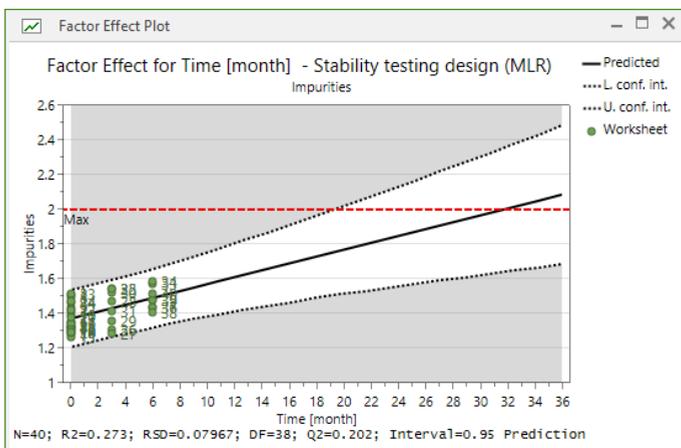
Subset selection of experiments representing different time spans is easily accomplished by using the **Included time points** and **Included combination** boxes of the worksheet. With all experiments included, the calculations automatically include the experiments that have response values and ignores those without.

Worksheet		Included time points: 0; 3; 6		Included combinations: [All]						
		3	4	5	6	7	8	9	10	
		Exp Name	Run Order	Incl/Excl	Time	Strength	Batch	Package	Impurities	
<input checked="" type="checkbox"/>	0									
<input checked="" type="checkbox"/>	3									
<input checked="" type="checkbox"/>	6	c1-T0	1	Incl	0	10	Batch 1	Pack A	1.50313	
<input checked="" type="checkbox"/>	9	c2-T0	2	Incl	0	20	Batch 1	Pack A	1.32298	
<input type="checkbox"/>	12	c3-T0	3	Incl	0	10	Batch 2	Pack A	1.41561	
<input type="checkbox"/>	18	c4-T0	4	Incl	0	20	Batch 2	Pack A	1.31665	
<input type="checkbox"/>	24	c5-T0	5	Incl	0	10	Batch 3	Pack A	1.32241	
<input type="checkbox"/>	36	c6-T0	6	Incl	0	20	Batch 3	Pack A	1.34488	
7	7	c7-T0	7	Incl	0	10	Batch 1	Pack B	1.28007	
8	8	c8-T0	8	Incl	0	20	Batch 1	Pack B	1.49107	

Below, we illustrate how main trend analysis of the Impurities trajectory can be done incrementally after 3, 6, 9 and 12 months. The **Factor Effect Plot**, available on the **Predict** tab, is an indispensable tool in this respect. Gradually, as more and more data points become available, we can see that the precision – in terms of a narrowing interval estimate – in the predicted level of Impurities is improved.

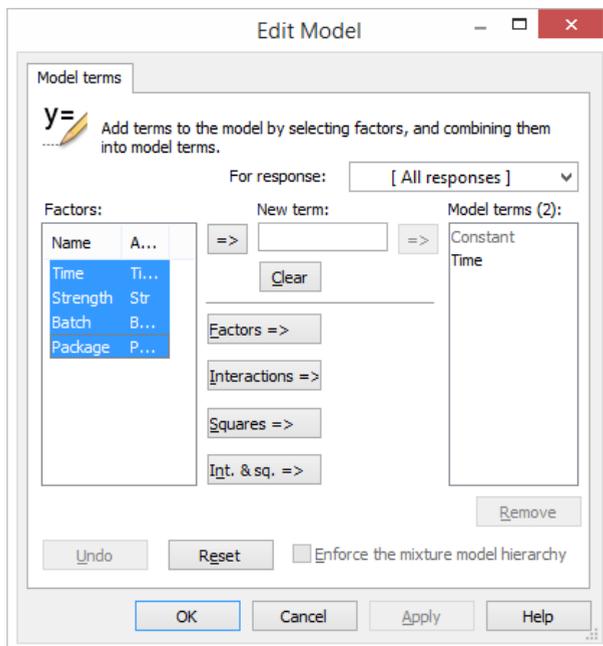
However, judging from this early stage trend analysis, we are at risk of violating the 36 months Max setting of the response. The question is whether any of the other factors display a significant influence on the response? This is addressed in the next section.





## Early stage data analysis – assessment of factor effects

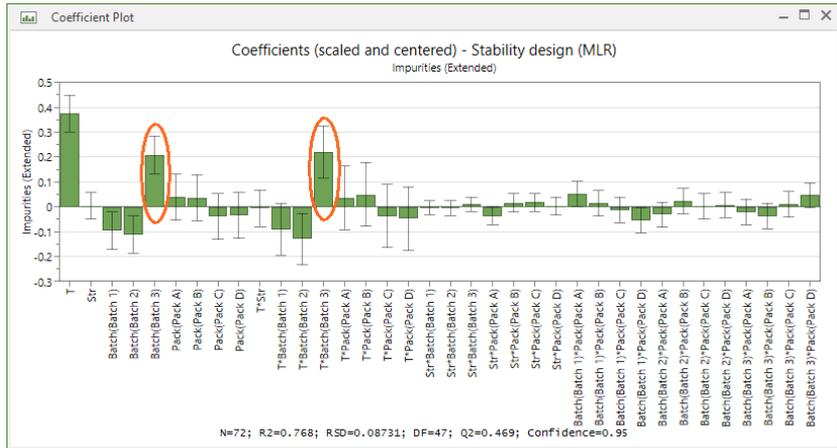
The main conclusion from the previous trend analysis was a worrying increase in the predicted level of Impurities towards the end of the stability test. Because of this undesired future trajectory it is of relevance to investigate the impact of the other three factors on the response variable. In order to accomplish this assessment, the regression model must be modified so that it includes all possible main and interaction effects; on the **Home** tab, click **Edit model**.



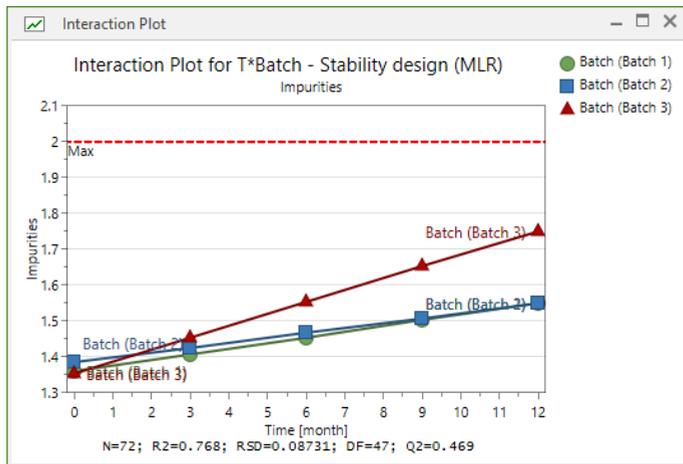
The regression model, based on all 0-12 months' measurements, was updated using the full set of main and interaction effects. A substantial increase in  $R^2$  was the result, from 0.506 (Time factor only) to 0.768 (all main and interaction terms). This increase in the explained variance points to the fact that one or a few of the additional model terms exhibit a significant impact on the response variable. This can be further assessed using the regression coefficient plot.

The regression coefficient plot shows a few noteworthy facts. It is evident that Time is the strongest factor. It has a positive coefficient meaning that the numerical value of the response variable increases with Time. However, there is also a statistically significant influence of the Batch factor; in particular, the third batch exhibits a strong positive influence on Impurities.

Moreover, the coefficient plot reveals a strong interaction effect between Time and Batch 3. The other two factors, Package and Strength, do not influence the response variable.



The **Interaction Plot** is a useful tool to get a graphical appraisal of an interaction. Based on the plot below there is no doubt that Batch 3 has a detrimental impact on Impurities compared with the other two batches.



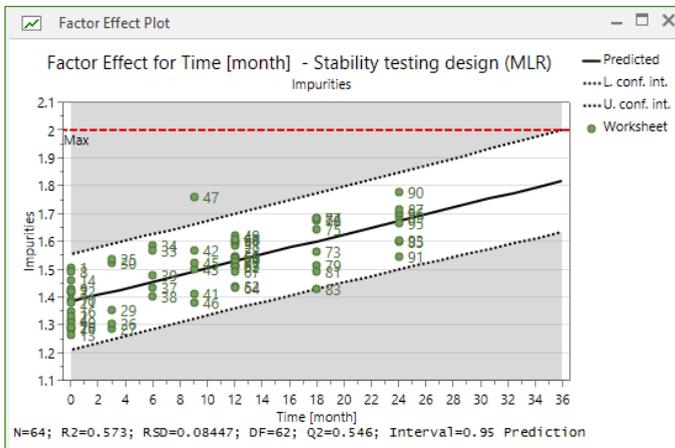
The main conclusion after analyzing the 0-12 months' segment of data is that Batch 3 is associated with a rapid and unacceptable increase in the level of Impurities. As a consequence this batch should be excluded from further consideration in the stability test. It will not comply with the final 36 months' acceptance criterion.

## Late stage data analysis

The early stage data analysis revealed that Batch 3 would not comply with the final acceptance criterion. Hence, all existing samples (0-12 months) and all future samples (18-36 months) including this factor setting need to be excluded from the data analysis. The easiest way to accomplish this is to sort the MODDE worksheet according to the Batch factor. Mark the Batch column, right click and click **Sort descending**. This will give you a worksheet where all runs involving Batch 3 are listed first. Then simply mark all rows containing Batch 3 and in the Incl/Excl column select Excl.

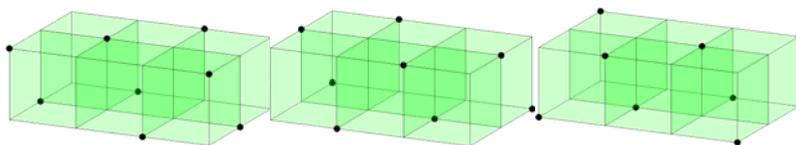
1	2	3	4	5	6	7	8	9	10
Exp No	Combination ID	Exp Name	Run Order	Incl/Excl	Time	Strength	Batch		
1	c1	c1-T0	1	Incl	0	10	Batch 1		
2	c2	c2-T0	2	Incl	0	20	Batch 1		
3	c3	c3-T0	3	Incl	0	10	Batch 2		
4	c4	c4-T0	4	Incl	0	20	Batch 2		
5	c5	c5-T0	5	Incl	0	10	Batch 3		
6	c6	c6-T0	6	Incl	0	20	Batch 3		
7	c7	c7-T0	7	Incl	0	10	Batch 1		
8	c8	c8-T0	8	Incl	0	20	Batch 1		
9	c9	c9-T0	9	Incl	0	10	Batch 2		

Before fitting the model on a larger time span (0-24 months) the model was again edited by reverting to using the Time as the single factor. The resulting Factor Effect plot for Time is seen here. The conclusion is that the 36 months' acceptance criterion most likely will be met.



## Summary and discussion

The new design family for stability testing is based on reduced combinatorial design sets that are perfectly complementary when superimposed on top of one another. The complementarity principle is illustrated by the three design cubes shown here. The design cubes represent a three-factor problem where one factor has two levels, one three levels and one four levels. The full factorial design in these three factors comprises  $2 \times 3 \times 4 = 24$  runs. Each design cube represents a one-third fraction in 8 runs. Mathematically, the three one-third fractions are equivalent and when superimposed on top of one another they perfectly pattern the corresponding full factorial design. This is the gist of the new approach to stability testing in MODDE, *i.e.*, the ability to fraction any stability testing design problem into reductions of various degree (design sets of one-half, one-third, one-fourth, ... reductions) that are balanced and perfectly complementary.



It should be emphasized that what is written and discussed in this appendix relates to the context of work performed in early development. Usually, in early development it is of relevance to explore many factors at many levels, and as early as possible to discover factors and factor settings that clearly encode unfavorable instability in the tested product. One of the interesting aspects of using the novel stability testing designs in MODDE is that all levels of all factors are tested at an early stage even though reduced design sets are employed. And this gives the beneficial asset of allowing the user to detect and remove elements in the experimental scheme that cause instability. Such early removal of uninteresting elements and zooming-in on critical features in the stability testing facilitate optimal use of the available testing resources.

## References

1. Anonymous, Bracketing and matrixing designs for stability testing of new drug substances and products, ICH Guideline Q1D.
2. Anonymous, Evaluations for stability data, ICH Guideline Q1E.

# References

1. R.A. Fisher, "Statistical methods, experimental design, and scientific inference", A re-issue (J. H. Bennett, Ed.), Oxford University Press, Oxford, England, (1990).
2. G.E.P. Box, W. G. Hunter, and J. S. Hunter, "Statistics for experimenters", Wiley, New York, (1978).
3. G.E.P. Box, The collected works, Vol 1. (G.C. Tiao, Ed.), Wadsworth Advanced Books and Software, Belmont, CA, 1985.
4. Morgan, Chemometrics: Experimental Design, ACOL, London, and Wiley, New York, (1991).
5. Wold, "Soft modeling, The basic design and some extensions", In Vol. II of K-G. Jöreskog and H. Wold, Ed.s. *Systems under indirect observation*, Vols. I and II, North-Holland, Amsterdam, (1982).
6. Wold, A. Ruhe, H. Wold and W. J. Dunn III, "The Collinearity Problem in Linear Regression. The Partial Least Squares Approach to Generalized Inverses", *SIAM J. Sci. Stat. Comput.* 5, 735-743, (1984).
7. Höskuldsson, "PLS Regression Methods", *J.Chemometrics*, 2, 211-228, (1988).
8. Wold. "Cross validatory estimation of the number of components in factor and principal components models", *Technometrics* 20, 397, (1978).
9. Draper and Smith, "Applied Regression Analysis", Second Edition, Wiley, New York.
10. Cornell, "Experiments with Mixtures", New York: Wiley, (1990).
11. Cox, "A Note on Polynomial Response Functions for Mixtures", *Biometrika*, 58, 155-159, (1971).
12. Crosier, "Mixture Experiments: geometry and Pseudo components", *Technometrics*, 26, 209-216, (1984).
13. Kettaneh-Wold, "Analysis of mixture data with partial least squares", *Chemometrics and Intelligent laboratory Systems*, 14, 57-69, (1992).
14. Rechtschaffner R.L., Saturated fractions of  $2n$  and  $3n$  factorial designs, *Technometrics*, 1967, Vol.9, N°4, 569-575.
15. Ing-Marie Olsson, Erik Johansson, Martin Berntsson, Lennart Eriksson, Johan Gottfries, and Svante Wold, "Rational DOE-protocols for 96 well plates", *Chemometrics and Intelligent Laboratory Systems*, 2006.
16. Doehlert, D.H., "Uniform shell designs", *Journal of the Royal Statistical Society*, 1970, Serie C, N°19, 231-239.
17. Snee, "Test Statistics for Mixture Models", *Technometrics*, Nov. 1974.

18. Cohen, Jacob (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New-Jersey, 2nd edition.
19. Hoenig, J. M. and Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55, 19-24.
20. Liu , Xiaofeng Steven (2014). *Statistical Power Analysis for the Social and Behavioral Sciences: Basic and Advanced Techniques*. Taylor & Francis, 1st edition.
21. O’Keefe, Daniel J. (2007). Post Hoc Power, Observed Power, A Priori Power, Retrospective Power, Prospective Power, Achieved Power: Sorting Out Appropriate Uses of Statistical Power Analyses. *Communication Methods and Measures*, 1(4), 291-299.
22. Zumbo, Bruno D. and Hubley, Anita M. (1998). A note on the misconceptions concerning prospective and retrospective power. *The Statistician*, 47, 385-388.
23. Cederkvist, H.R., Aastveit, A.H., and Naes, T., The importance of functional marginality in model building – A case study, *Chemometrics and Intelligent Laboratory Systems*, 87, 98-106, 2007.
24. Draper, Norman and Smith, Harry, *Applied Regression Analysis*, Second Edition, Wiley, New York, 1981.
25. Golub, Gene H. and Van Loan, Charles F, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1983.
26. Rao, C. R. (1947), Factorial Experiments Derivable from Combinatorial Arrangements of Arrays, *Journal of the Royal Statistical Society*, Supplement,9, 128–139.
27. Wang, J. C., and Wu, C. F. J. (1992), Nearly Orthogonal Arrays With Mixed Levels and Small Runs, *Technometrics*, 34, 409–422.
28. Hongquan Xu (2002), An Algorithm for Constructing Orthogonal and Nearly-Orthogonal Arrays With Mixed Levels and Small Runs, *Technometrics*, 44, 356–368.
29. Anonymous, Bracketing and matrixing designs for stability testing of new drug substances and products, ICH Guideline Q1D.
30. Anonymous, Evaluations for stability data, ICH Guideline Q1E.

# Index

## 2

2D	
Contour .....	179
Design space .....	106, 187, 229
Sweet spot .....	184, 227

## 3

3D	
Onion plots .....	128
Response surface .....	181
Rotate .....	202
Scatter .....	134, 174
Sweet spot .....	184
Zoom and rotate .....	202

## 4

4D	
Contour .....	105, 180
Design space .....	106, 187, 229
Sweet spot .....	184, 227

## A

Abbreviation	
Factor definition .....	20
Investigation options .....	73
Response definition .....	28
About MODDE .....	5
Absolute limits .....	222, 223
Accelerators .....	7, 82
Acceptance limit .....	80
Activate MODDE .....	5
Add	
Component .....	100
Experiment .....	131
Factor .....	20, 54
Inclusions .....	115
Plot element .....	199
Response .....	28
Terms .....	54, 58, 98
To favorites .....	211
To report .....	211
Add plot element .....	199
Add to favorites .....	211
Add to report .....	211
Alternative setpoints .....	221
Analysis advisor .....	193
Analysis phase .....	2, 16

Analysis wizard	
Coefficients .....	92
Histogram .....	90
Interaction test .....	94, 266, 267
Observed vs. predicted .....	97
Replicates .....	89
Residuals normal probability .....	96
Square test .....	93, 266, 267
Summary of fit .....	95
Tests statistical appendix .....	266, 267
Toolbar .....	88
What is? .....	88
Analyze tab .....	145
ANOVA .....	150, 151, 261
Arrange windows .....	197
Arrow .....	201
Audit trail .....	73, 76, 194
Augmenting designs .....	54
Auto transform .....	266
Auto tune .....	266
Automatic	
Auto update predictions .....	173
Cross-validation significance rules .....	256
Fit .....	100
Update of plots and lists .....	212
Update predictions .....	173
Autoscale modifier .....	30
Axial designs .....	306
Axis .....	204, 205

## B

Background .....	207
Balanced .....	297, 301, 315, 316
Block	
Blocks .....	37
BlockV .....	131
D-Optimal .....	289
Fixed .....	289
Interaction .....	37
Mark .....	201
Option .....	73
Orthogonal .....	286
Random .....	289
RSM .....	288
Screening .....	287
Block interaction .....	37, 39
Blocks .....	37, 39

- Box-Behnken.....301
- Box-Cox .....154, 273
- C**
- Candidate set
- Constrained.....114
  - Create new .....41
  - Edit.....41
  - Imported.....41, 60
  - Layer runs .....43
  - Max size.....70
  - New design .....60
  - Onion .....315
  - Open.....126
  - Regular.....309
  - Size.....41
- Case sensitivity.....17
- CCC.....301
- Blocking.....288
  - Constraints .....219
  - Design.....301
  - Star distance.....38
- CCC constraints.....219
- CCF .....301
- CCO.....301
- Center points .....37
- Central Composite Circumscribed.....301
- Central Composite Face .....301
- Central Composite Orthogonal.....301
- Classical mixture designs .....306
- Close.....68
- Coding qualitative factors.....268
- Coefficients
- Compact format .....77, 264
  - Confidence intervals .....77, 264
  - Extended format.....77, 264
  - Interaction test.....94, 266, 267
  - Normalized.....77, 264
  - Plot.....160
  - PLS orthogonal .....77, 264
  - Scaled and centered.....77, 264
  - Square test.....93, 267
  - Statistical appendix.....77, 264
  - Unscaled .....77, 264
- Color
- ANOVA table .....151
  - Axes .....204
  - by variable .....134
- Contour.....210
- Correlation matrix.....137
- Format plot .....203
- List.....81
- Mini toolbar.....9
- Onion by layer .....128
- Scatter.....134
- Setpoint properties .....237
- Sweet spot.....184
- Worksheet.....131
- Column formatting .....211
- Combinatorial designs.....298
- Compact coefficients.....77, 264
- Compatibility.....314
- Complement design
- Doehlert .....57
  - D-Optimal .....58
  - Estimate squares terms.....56
  - Fold over.....56
  - Inclusions vs. complement.....115
  - New.....54
  - New complement design.....54
  - PBSS to PB.....54
  - Rechtschaffner .....54
  - What is? .....54
  - With inclusions .....116
- Condition number.....138, 141, 269
- Confidence
- Interval for coefficients.....77, 264
  - Interval for predictions.....271
  - Interval type and probability levels..78
  - Level .....73
- Confoundings .....126
- by variable .....134
- Constraints
- Candidate set.....114
  - Define .....113
  - Define graphically.....113
  - Define in spreadsheet.....27
  - Mixture constraint.....305
  - Modify graphically .....114
  - Page .....26
  - Qualitative or multilevel .....28
  - Supported.....27
  - Types of.....28, 112
  - Why .....112
- Context menu .....212
- Contextual tab .....6



What is? .....	329, 330
Design summary .....	122, 123, 124
Design tab .....	109
Design wizard .....	19, 86
Desirability .....	322, 324
Desirability settings .....	322
Determinant .....	308, 311
DF .....	257, 262
DF residual .....	138, 262
Distance to model .....	170
Distance to target .....	283
Distribution	
Factor .....	241
Histogram .....	141
Response .....	242
Setpoint analysis .....	291
Setpoint Factor spreadsheet .....	238
DModY .....	170
Dockable windows .....	193
Doehlert designs .....	291, 303
D-Optimal	
Algorithm .....	310
Balanced .....	40
Blocking .....	289
Candidate set .....	315
Complement .....	58
Criteria .....	42, 47
Design evaluation .....	311
Design runs span .....	40
Design summary .....	122, 123, 124
Design wizard page .....	39
Implementation .....	310
Onion candidate set .....	45
Onion designs .....	314
Regenerate design .....	127
Repetitions .....	40
What is? .....	308
When? .....	309
D-Optimal onion	
Candidate set .....	45, 315
Designs .....	299, 302
From candidate set .....	60
From scores .....	61
Layers max .....	70
Layers page .....	43
Plots .....	128
DPMO .....	293
DPMO default limit .....	80
DS	
Appendix .....	329
Documentation .....	332, 334
Explorer .....	233
From Home tab .....	106
From Optimizer tab .....	229
From Setpoint analysis .....	243
How to find the best .....	338
Monte Carlo simulations .....	337
Options .....	188
Predict tab .....	187
What is? .....	329, 330
DSD .....	301
Dynamic profile .....	226
<b>E</b>	
Edit menu .....	9
Edit model .....	98
Effects	
List .....	167
Main effect .....	165
Normal probability .....	165
Open .....	163
Plot .....	164
Eigenvalue .....	269
Email mip .....	68
Encrypt .....	51
Error bars .....	211
Estimate squares terms .....	56
Evaluate .....	138, 141
Evaluation plot .....	214
Exclude .....	106, 203
Execute folder .....	196
Experiment name .....	131
Experiment number .....	131
Experimental cycle .....	2, 4
Experimental design	
Create .....	53
Objective .....	13
What is? .....	13
Experimental region .....	118
Experiments .....	131
Export	
Favorites configuration .....	196

## MODDE 12

Report to PDF .....	248
To SIMCA .....	68
Extended	
Axial .....	306
Design matrix .....	119, 121
List presentation .....	73, 76, 194
Model list .....	98
External variability .....	135, 286
Extreme vertices .....	122, 123, 124

## F

Factor	
Add .....	19
Advanced .....	24
Contribution .....	285
Definition .....	20, 21, 24
Delete .....	19
Distribution .....	241
Edit .....	19
Effects .....	177, 178
Formulation .....	304
Max number .....	21
Mixture and process .....	304
Mixture definition .....	304
Modify .....	19, 110
Name .....	17
Precision .....	24
Qualitative .....	33, 268
Scaling .....	257
Setting .....	21
Spreadsheet .....	219, 238
Transformation .....	24
Type .....	21
Use for factor .....	23
What is? .....	110
Factor distribution .....	241
Factorial designs .....	297, 298, 301
Favorites	
Add to favorites .....	211
Create folder .....	196
Delete .....	196
Export .....	196
Import .....	196
Open and tile .....	196
Rename .....	196
Restore .....	84, 197
Treat folder as item .....	196
Window .....	196
F-distribution .....	151, 152
File tab .....	49

Filler .....	21, 23
Fit	
Goodness of fit .....	257, 262
Lack of fit .....	152
Methods .....	15, 100, 101
Mixture .....	101
MLR .....	253
Model .....	17, 100
Plots .....	101
PLS .....	254
Summary list .....	149
Summary plot .....	147
Fixed block factor .....	289
Fold over .....	54
Foldover .....	54
Footer .....	207
Format plot	
Axes .....	204
Axis and title font .....	206
Axis X and Y .....	205
Background .....	207
Column .....	211
Contour .....	210
Error bars .....	211
Footer .....	207
Gridlines .....	206
Header .....	207
Labels .....	209
Legend .....	208
Limits and regions .....	208
Mini toolbar .....	203
Open .....	203
Restore .....	200
Save .....	200
Styles .....	211
Tick Marks .....	205
Titles .....	207
Formulation .....	304
Fractional factorial	
Complement .....	54
Designs .....	297
Free-form selection .....	201
Full factorial	
Design .....	297
Three levels .....	301
Full screen .....	197
<b>G</b>	
Gallery .....	6

G-efficiency.....	311
General options	
General tab.....	70
List options.....	81
Restore.....	84
General page	
Factor definition.....	21
MODDE options.....	70
Generalized fractional factorial ...	343, 350
Generalized subset designs.....	343, 350
Generate report	
Add to report.....	211
Report.....	52
Generators.....	117
Goodness of fit.....	257, 262
Graeco-Latin square.....	298
Gridlines.....	206
GSD.....	343, 350

## H

Hat matrix.....	257, 262
Header.....	207
Help	
Help.....	4, 68
Interactive help.....	4
Sartorius Stedim Data Analytics.....	5
Hierarchy.....	257
High limit.....	28, 219
Histogram.....	90, 141
Home tab.....	85
Hypercube.....	233
Hyperlink.....	250
Hyper-triangles.....	291, 303

## I

Image.....	248
Import	
Candidate set.....	60, 315
Design from file.....	54
Favorites configuration.....	196
Inclusions to worksheet.....	115
Scores.....	61
Worksheet to inclusions.....	116
Inclusions	
And design augmentation.....	312

As part of design.....	116
Edit.....	116
Open.....	115
vs. complement design.....	115
Individual response analysis.....	242
Insert in report	
Add to report.....	211
Hyperlink.....	248
Picture.....	248
Template.....	248
Insert rows.....	131
Installation.....	1
Interaction plot.....	167
Interaction test.....	94, 267
Interactive exclude tool.....	106, 203
Interval estimates.....	78
Interval type and probability levels.....	78
Investigation	
Change options.....	73
Compatibility.....	10, 61
Managing.....	10
Mip.....	10
What is?.....	10
Irregular region.....	312

## K

Keyboard shortcuts.....	7, 82
KeyTips.....	7, 82

## L

Labels.....	209
Lack of fit.....	152
Latent structures.....	254
Layers overlap.....	43
L-designs.....	298
Legend.....	208
Licenses.....	5
Limit optimization.....	323
Limits.....	80, 208
Line style.....	9, 211
Linked responses.....	33
List	
Create.....	211
Presentation.....	73

## MODDE 12

Loading plots.....	168, 169
Lock	
Contour levels.....	182
Investigation.....	51
LOF .....	152
Log	
Determinant .....	311
In audit trail.....	76, 194
LogDet.....	311
LogDetNorm .....	311
Low limit.....	28, 219

## M

Main effect plot .....	165
Manage licenses .....	5
Mark .....	201
Maximum runs .....	122, 123, 124
Mid-range.....	25
Mini toolbar.....	9
Mip-file .....	10
Missing data .....	258
Mixture	
And process factors.....	304
Constraint.....	305
Contour .....	179
Cox.....	274, 275
Data in MODDE .....	274
Design space .....	187
Designs .....	304
Experimental region.....	305
Factor definition.....	304
Factors statistical appendix .....	274, 275
Fit.....	101
Irregular region .....	312
Models .....	274, 275
Prediction plots .....	175
Slack variable model.....	274, 275
Sweet spot.....	184
MLR	
Fit methods .....	15
Formula.....	253
MLR scaling	
Mid-range .....	25
Orthogonal .....	25
Responses .....	30
Statistical appendix .....	257
Unit variance.....	25

MODDE	
About .....	5
Starting.....	1
MODDE options	
General tab.....	70
List options .....	81
Restore.....	84

Model	
Distance to .....	170
Edit .....	98
Error.....	294
Fit.....	100
Hierarchy .....	257
Individual.....	100
List.....	98
Q2 .....	259
R2 .....	259
Reproducibility .....	260
Saturated .....	257, 262
Select .....	35, 38
Singular.....	257
Validity .....	260
What is? .....	13

Monte Carlo .....	291
Multilevel factor.....	21
Multiple Linear Regression .....	15, 253
Multiplot	
Overview plot .....	102
Select responses .....	212
Multivariate .....	61

## N

Near orthogonal array.....	315
Network installation.....	5
New	
Application designs.....	59
Complement design .....	54
Design from candidate set.....	60
Design from scores .....	61
D-Optimal .....	127
Experimental design .....	53
External design .....	54
Investigation .....	10
RED-MUP .....	62, 63
Report .....	52
Worksheet from file.....	54
Next component .....	100
NOA .....	315

- Normal probability  
  Effects ..... 165  
  Residuals ..... 156
- Normalized  
  Coefficients ..... 77, 264  
  LogDetNorm ..... 311
- Notes ..... 193
- N-plot  
  Effects ..... 165  
  Residuals ..... 156
- Number format ..... 73
- Number of decimals ..... 25
- N-value ..... 259
- O**
- OA ..... 315
- Objective  
  Select ..... 33  
  Settings ..... 38  
  What is? ..... 118
- Observed vs. predicted ..... 97, 104, 159
- One-Click simple case ..... 88
- Onion  
  Candidate set ..... 315  
  Designs ..... 299, 302  
  Generate ..... 60, 61  
  Layers max ..... 70  
  Plots ..... 128  
  Screening ..... 314
- Open all items ..... 196
- Open investigation ..... 64
- Operators - derived responses ..... 32
- Optimization  
  Blocking ..... 288  
  Design ..... 14  
  Focus ..... 283, 325  
  Limit ..... 280, 323  
  Objective ..... 33  
  Robust ..... 231  
  Target ..... 282, 324
- Optimizer  
  Alternative setpoints ..... 221  
  Conditional tab ..... 225  
  Contour ..... 226  
  Copy to predictions ..... 214  
  Definition ..... 280  
  Design space ..... 229
- Desirability ..... 217, 280, 283
- f(ds) ..... 322
- Factor spreadsheet ..... 219
- Introduction ..... 213, 280, 321
- List ..... 225
- Objective tab ..... 216
- Objectives ..... 214, 322
- Overall distance to target ..... 283
- Properties ..... 222, 223
- Response spreadsheet ..... 216
- Search function ..... 326
- Specifications ..... 325
- Summary ..... 225
- Sweet spot ..... 227
- Tab ..... 225
- Weight ..... 322
- Window ..... 214
- Options  
  Customize ..... 81, 82  
  Investigation Options ..... 73  
  List options ..... 81  
  MLR scaling ..... 25  
  MODDE options ..... 70  
  Restore ..... 84  
  Types ..... 70
- Organization ..... 6
- Orthogonal  
  Array ..... 315  
  Blocking ..... 286  
  Correlation ..... 137  
  Scaling ..... 25
- Orthogonal array ..... 315
- Outliers  
  Distance to model ..... 170  
  Residuals normal probability plot ... 156  
  Score plots ..... 169
- Output window  
  Open automatically ..... 70  
  Show or hide ..... 193  
  What is? ..... 193
- Overall distance to target ..... 283
- Overlay prediction ..... 175
- Overview plot ..... 102
- P**
- Partial Least Squares ..... 15, 253, 254
- Password protect ..... 51
- Paste design ..... 33
- Paste unformatted in report ..... 248

## MODDE 12

PBSS .....	299	Overview .....	102
Picture .....	248	PLS .....	168
Placeholders		Replicates.....	142
View.....	250	Residuals normal probability .....	156
Window.....	251	Scatter .....	134
Plackett Burman		Summary of fit.....	147
Blocking.....	287	Sweet spot.....	184
Designs .....	299	VIP .....	170
PBSS designs .....	299	PLS	
Plackett Burman Super-Saturated.....	299	Fitting with PLS.....	15, 254
Plate-Size.....	62	Orthogonal coefficients.....	77, 264
Plot		PLS coefficients.....	273
Copy.....	65	PLS summary list.....	150
Preview .....	65	Scaling .....	25
Quality .....	65	Scheffé.....	100
Save .....	65	Scores and loadings .....	168, 272
Size .....	65	Summary plots.....	147, 148
Plot formatting		Potential terms.....	310
Axes .....	204	Power estimation .....	39, 316
Axis and title font.....	206	Predict	
Axis X and Y .....	205	Contour .....	179
Background.....	207	Design space .....	187
Column .....	211	Factor effects .....	177, 178
Contour .....	210	List.....	173
Error bars .....	211	Optimizer .....	214
Footer .....	207	Prediction plots .....	175
Gridlines .....	206	Residuals vs. predicted.....	157
Header.....	207	Scatter plot.....	174
Labels.....	209	Setpoint validation .....	190
Legend .....	208	Spreadsheet.....	173
Limits and regions.....	208	Sweet spot.....	184
Mini toolbar .....	203	Tab.....	173
Open.....	203	Prediction	
Restore .....	84, 200	Copy from optimizer.....	214
Save .....	200	Formula.....	271
Styles .....	211	Phase .....	4
Templates.....	200	Plots .....	175
Tick Marks.....	205	Spreadsheet.....	173
Titles .....	207	Using the model.....	17
Plots		With model error.....	294
Box-Cox.....	154	Predictive factor distribution .....	241
Coefficients.....	160	PRESS.....	256, 257, 262
Contour .....	105	PRESS/SSY.....	15
Design space .....	106	Preview.....	65
DModY.....	170	Print	
Effect.....	164	Format.....	67
Histogram .....	141	Quality .....	65
Interaction.....	167	Probability contour plot.....	187, 329
Main effect.....	165	Probability of failure correlation .....	335
Observed vs. predicted.....	104		

Probability of failure limit .....	80
Process and mixture factors .....	304
Process Capability Index .....	292, 293
Process factors .....	279
Product ID .....	5
Program limits .....	70
Properties	
Contour .....	182
Dialog .....	212
Optimizer .....	222, 223
Page .....	212
Report generator .....	252
Setpoint .....	237
Protect investigation .....	50, 51
<b>Q</b>	
Q2 .....	259
Qualitative factors	
Coding at > 2 levels .....	268
Definition .....	21
In derived response .....	33, 268
Quantitative .....	21
Quantitative multilevel .....	21
Quick Access Toolbar .....	9, 10
Quick start	
Analysis wizard .....	88
Design wizard .....	86
<b>R</b>	
R2 .....	259
Random factor .....	289
Raw residuals .....	80, 268
Recalculate scale .....	204
Recent	
Folders .....	64
Investigations .....	64
Rechtschaffner designs .....	299, 302
Recommended designs .....	35
RED-MUP .....	62, 63, 300
Reduced	
Axial .....	306
CCC .....	301
CCF .....	301
Reduced combinatorial designs .....	298
Reference mixture .....	117
References .....	18, 363
Regions .....	208
Register .....	5
Regular responses .....	30
Remove	
Encryption .....	51
Footer .....	207
Header .....	207
Model term .....	203
Placeholder .....	250
Timestamp .....	207
Replicate	
ANOVA .....	151
Design .....	37
Plot .....	89, 142
Tolerance .....	70
Report	
Add to report .....	252
Create .....	52, 245, 246
File .....	247
Formatting .....	250
Home .....	248
Open .....	52, 245, 246
Picture .....	248
Placeholders window .....	251
Properties window .....	252
Start .....	52, 245, 246
Template .....	247
View .....	250
What is? .....	245
Window .....	247
Report generator .....	52, 245, 246, 247
Report writer .....	52, 245, 246, 247
Reproducibility .....	260
Reset	
Factor precision .....	24
Factors columns .....	110
Favorites .....	84
Interface customization .....	10
Messages .....	84
Model .....	98
Optimizer factor settings .....	214
Optimizer response settings .....	214
Plot formatting .....	84, 200
Report template .....	248
Rotation .....	202
Residual standard deviation .....	260

Residuals	
Default .....	155
Gallery .....	104
List .....	158
Normal probability.....	96, 156
Type definition.....	80, 268
Types .....	155
vs. predicted response .....	157
vs. run order .....	157
vs. variable plot.....	158
Response	
Add .....	28
Box.....	212
Definition .....	28
Derived .....	31
Distribution setpoint.....	242
Exclude using modifier .....	30
Linked .....	33
Name.....	17
Profile .....	226
Regular.....	30, 112
Response box .....	212
Select.....	212
Setpoint spreadsheet.....	240
Spreadsheet .....	112
Surface .....	181
Surface modeling .....	14
Transform .....	30
Restore	
Favorites .....	84, 200
Interface customization .....	10
Plot formatting .....	84, 200
Report template.....	247, 248
Ribbon	
Customization .....	10
Description.....	6
Robust optimization	
Design space explorer.....	233
Find.....	231
Results.....	232, 233
Robust setpoint	
Design space explorer.....	233
Find.....	231
Results.....	232, 233
Robustness.....	338
Rotate .....	202
RSD.....	260
RSM	
Blocking.....	288
Design.....	14
Objective.....	33
Run list.....	221
Run order.....	135
<b>S</b>	
Safe region .....	327
Saturated models .....	257, 262
Save	
Audit trail.....	73, 76, 194
Inclusions.....	115
Investigation .....	65
Plot formatting.....	200
Plot or list.....	65
Templates.....	200
Scaling	
Coefficients.....	77, 264
Formula.....	257
MLR .....	25
PLS .....	25
Scatter	
Loading.....	169
Model data .....	134
Prediction.....	174
Score .....	169
Scheffé .....	101
Score plots.....	169
Screen reader.....	203
Screening	
Designs .....	297
Number factors .....	13
ScreenTips.....	6
Search function.....	321
Select	
All .....	7, 82
Default factor .....	73
Fit method.....	100
In plot.....	201
Objective.....	33
Responses .....	212
Selected setpoint.....	220
Send as attachment .....	68
Send by email .....	68
Sensitivity analysis.....	284

Set Run order.....	135	Stability testing design .....	298, 299, 356
Setpoint		Standard error.....	271
Alternative setpoints .....	221	Standard fit.....	100
Design space .....	243	Standardized residuals.....	80, 268
Factor.....	238, 241	Star distance .....	38
Introduction.....	235, 236	Starting MODDE.....	1
Properties .....	237	Starting simplexes .....	284
Response.....	240, 242	Statistical appendix .....	253
Robust.....	231, 232, 233	Status bar.....	193
Selected.....	220	Styles.....	211
Setpoint analysis .....	234	Summary of fit	
Setpoint validation .....	190	Analysis wizard .....	95
Summary.....	242	List.....	149
Settings		Plots .....	101, 145, 147
CCC.....	35	PLS.....	147, 148, 150
Factor.....	21	Surface	
Optimizer.....	214	Onion plots .....	128
Plot.....	203	Response surface .....	181
Restore .....	84	Rotate.....	202
Save .....	200	Sweet spot.....	184
Split objective .....	38	Zoom and rotate.....	202
Share.....	68	Sweet spot .....	105, 184, 227
Favorites configuration .....	196	Symbol style.....	9, 211
Report to PDF .....	248	Syntax derived response .....	32
To SIMCA .....	68		
Shortcut menu .....	212	<b>T</b>	
Show		Tagushi.....	13
Contour levels.....	188	Target	
Expanded RED-MUP.....	70	In optimizer.....	216
Placeholders .....	250	Optimization .....	324
ScreenTips .....	10	Response definition.....	28
Units.....	70	Templates	
Significant term .....	92	Plot.....	200
SIMCA .....	61	Report .....	247, 248
Simplex		Theme.....	70
Mixture .....	305	Threshold .....	137
Optimizer search function.....	284	Tick marks.....	205
Size preview .....	65	Timestamp.....	207
Skewness test.....	138	Titles .....	207
Slack variable .....	274, 275	Tools tab.....	199
Sort.....	133	Tooltip.....	6
Specific application designs .....	59	Total runs .....	37
Specification.....	97	Traditional designs .....	53
Split button .....	6		
Split objective.....	33		
Spreadsheet.....	173, 219, 238		
Square test.....	93, 267		

Transform ..... 24, 30  
 Treat folder as item ..... 196  
 Tukey's and variability tests\_  
     Statistical appendix ..... 266

**U**

Uncentered coefficients ..... 77, 264  
 Unconfound ..... 38  
 Uncontrolled ..... 23  
 Undo ..... 107  
 Unit variance ..... 25  
 Units ..... 70  
 Update  
     Constraint graphically ..... 114  
     Placeholder ..... 250  
     Plots and lists ..... 212  
     Prediction spreadsheet ..... 173  
     Report ..... 248  
 Update of plots and lists ..... 212  
 User interface ..... 10  
 Usp-file ..... 61

**V**

Validity ..... 260  
 Variable importance ..... 170  
 Variable plot ..... 158  
 View tab ..... 193  
 View windows ..... 193  
 VIP ..... 170

**W**

WC plots ..... 272  
 Web ..... 5  
 Weight in optimizer ..... 222, 223  
 Windows  
     Analysis advisor ..... 193  
     Arrange ..... 197  
     Audit trail ..... 76, 194  
     Favorites ..... 195  
     Notes ..... 193  
     Output ..... 193  
 Wizard  
     Analysis wizard ..... 88  
     Contour plot wizard ..... 182  
     Design space wizard ..... 188  
     Design wizard ..... 86  
     Sweet spot wizard ..... 184  
 Worksheet  
     Add experiments ..... 131  
     Adding inclusions ..... 115  
     Colors ..... 131  
     Derived responses ..... 31  
     Design matrix ..... 119, 121  
     Missing values ..... 131  
     Open ..... 131  
     Run Order ..... 135  
     Sort ..... 131  
     Tab ..... 131  
 Write report ..... 52

**Z**

Zoom ..... 9, 202