PROTEIN SPRAY-DRYING (Full Fac)

In This Tutorial You Will Learn How to

- Define response condition and optimization objectives
- Set up quantitative factors
- Create a full factorial design
- Refine a model by using the Analysis wizard
- Define a design space
- Identify optimal setpoint and robust setpoint by using the Optimization wizard
- Compare predicted process performance at two different setpoints using the Normal Operating Range setting

Background

Spray-drying is a process often used for drugs intended for inhalation. When spray-drying proteins, the main aim is to produce particles of a specified size. In addition, it is important that the protein temperature remains relatively low to avoid unnecessary denaturation. Protein degradation may involve many complicated physical and chemical processes, including denaturation. Therefore, we wish to study protein stability at a molecular level in order to facilitate formulation applications. This example is based on a model protein (D7599) developed by AstraZeneca where protein powders of D7599 were produced by spray-drying.

Objective

The experimental objective of this study was to determine which process parameters influence the quality of the spray-dried product. The data analysis will illustrate the use of the analysis and optimization wizards. Original data source: Cronholm, M., *The Effect of Process Variables on a Spray-dried Protein Intended for Inhalation*, Undergraduate Research Study, Dept. of Pharmaceutics, Uppsala University, Uppsala, Sweden, 1998. Numerical limits of the responses and factors have been moderated to suit the needs of the example.

Data

To characterize the outcome of spray-drying the following five responses were measured:

- Yield – the amount of product produced. This should be maximized.
- Size – particle size. Particles must be in the range 0.5–3.3 µm in order to reach the lower airways.
- Water – water content in the spray-dried protein. This should be minimized and must be below 3.5.
- Outlet temperature – outlet air temperature. This temperature may influence protein degradation and was therefore included. No specific target value was specified for this response.
- HMWP – high molecular weight proteins. Measures the extent of aggregations, i.e., the formation of dimers and oligomers of the protein. This should be as low as possible and always below 1.

Responses

	Name	Abbreviation	Units	Condition	Objective	Min	Target	Max
1	**Yield**	Yie	% by weight	Desired	Maximize			
2	**Size**	Size	µm	Required	Inside	0.5		3.5
3	**Water**	H_2O	% loss of drying	Required	Inside			3.5
4	**Outlet Temp**	OutT	°C	Observed	Predicted			
5	**HMWP**	HMWP	% by area	Required	Minimize		0.2	1
+	Add...							

Spray-drying conditions were varied using a full factorial design in four factors:

- Inlet Temperature – temperature of drying air at the inlet of the equipment. The high and low levels of this factor were set such that degradation would be expected at the high level (220°C) but not at the low level (100°C).
- Atomization gas flow – for this factor the low level (500 l/h) of the atomization gas (nitrogen) was the minimum required to provide sufficient energy for atomization. The high level (800 l/h) was the maximum achievable flow with this spray-dryer.
- Aspiration rate – the aspirator draws air through the instrument and this was varied from 60% to 100% (full capacity).

Feed-flow – indicates the material flow through the equipment. Here, the high level of 5ml/min was the maximum rate which could be used at the low temperature without condensation appearing in the drying chamber; the low level (2 ml/min) was chosen as the slowest practical rate.

Factors

	Name	Abbreviation	Units	Type	Settings	Precision	NOR
1	**Inlet Temperature**	InT	°C	Quantitative	100 to 220	2	2
2	**Atomization Gas Flow**	Ato	liters / hour	Quantitative	500 to 800	5	10
3	**Aspiration Rate**	Asp	%	Quantitative	60 to 100	2	5
4	**Feed Flow**	FF	ml / min	Quantitative	2 to 5	0.05	0.1
+	Add...						

Worksheet

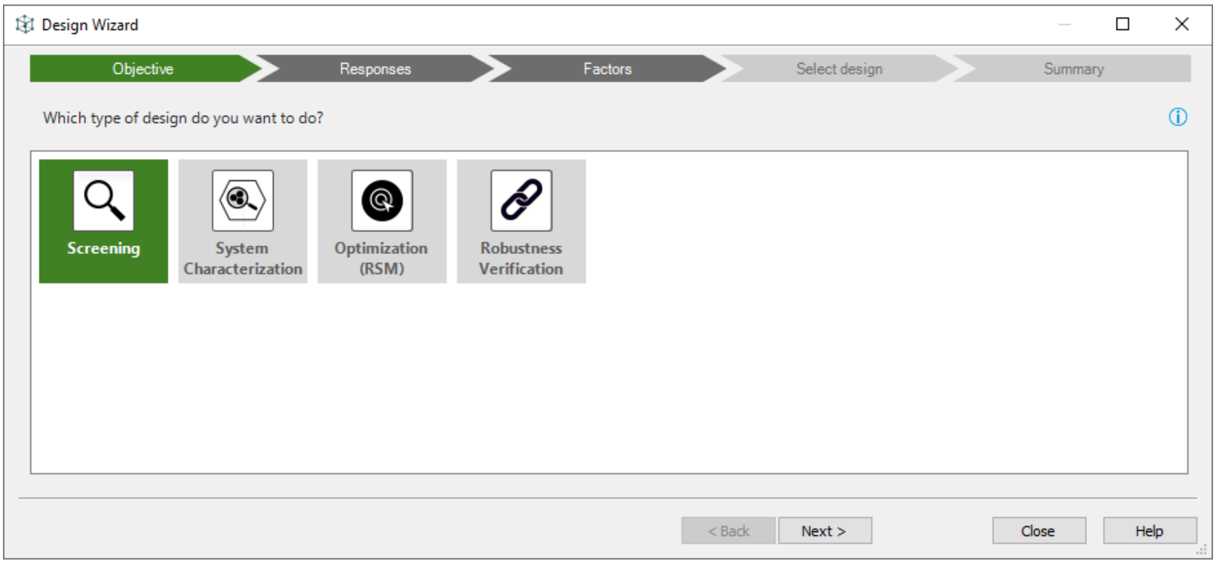
	1	2	3	4	5	6	7	8	9	10	11	12	13
	Exp No	Exp Name	Run Order	Incl/Excl	Inlet Temperature	Atomization Gas Flow	Aspiration Rate	Feed Flow	Yield	Size	Water	Outlet Temp	HMWP
1	1	N1	13	Incl	100	500	60	2	43.5	3.1	4.5	63	0.5
2	2	N2	11	Incl	220	500	60	2	43.5	3.8	3.2	117	0.5
3	3	N3	15	Incl	100	800	60	2	8.5	2.2	4.5	62	0.6
4	4	N4	10	Incl	220	800	60	2	9	2	3.8	114	0.5
5	5	N5	19	Incl	100	500	100	2	51	3.4	3.1	74	0.4
6	6	N6	3	Incl	220	500	100	2	61	4.2	1.9	150	3.3
7	7	N7	9	Incl	100	800	100	2	22.5	1.6	3.3	72	0.3
8	8	N8	5	Incl	220	800	100	2	27	1.6	2.8	142	2.3
9	9	N9	6	Incl	100	500	60	5	48	3.2	4.8	52	0.3
10	10	N10	7	Incl	220	500	60	5	45	3.6	2.3	134	0.8
11	11	N11	8	Incl	100	800	60	5	12.5	2.1	5.1	50	0.3
12	12	N12	2	Incl	220	800	60	5	11	2.3	3.3	120	0.7
13	13	N13	1	Incl	100	500	100	5	61.5	3.2	4	61	0.3
14	14	N14	16	Incl	220	500	100	5	60.5	4.4	2.2	142	1.8
15	15	N15	17	Incl	100	800	100	5	35	1.6	4.2	59	0.3
16	16	N16	12	Incl	220	800	100	5	33	1.75	2.7	138	1.4
17	17	N17	4	Incl	160	650	80	3.5	34.5	2.2	3.4	96	0.5
18	18	N18	14	Incl	160	650	80	3.5	39	2.4	3.7	97	0.4
19	19	N19	18	Incl	160	650	80	3.5	38	2.2	3.2	98	0.5

Setting up the Design

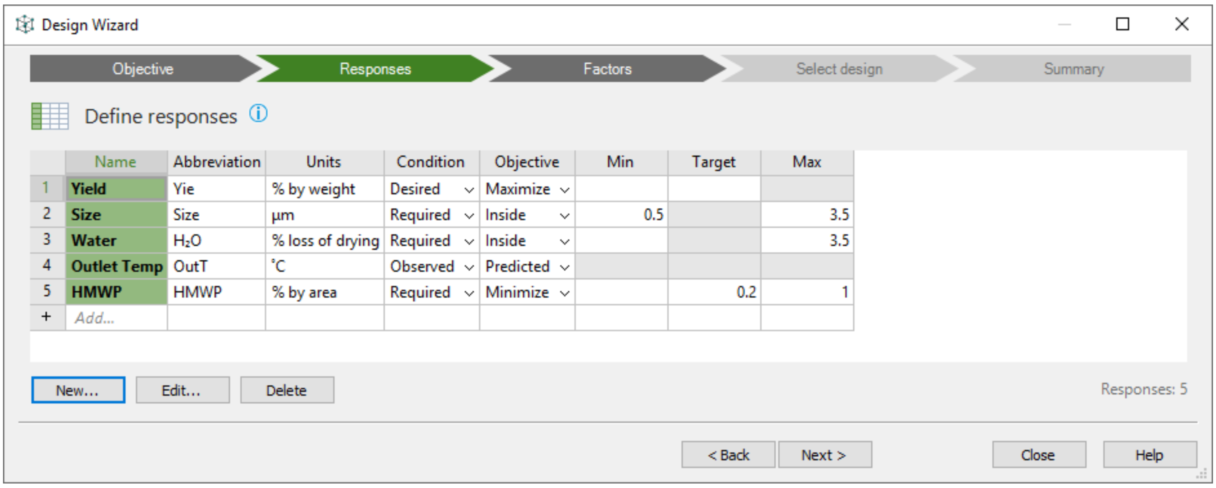
Design wizard

Initiate a new investigation in MODDE.

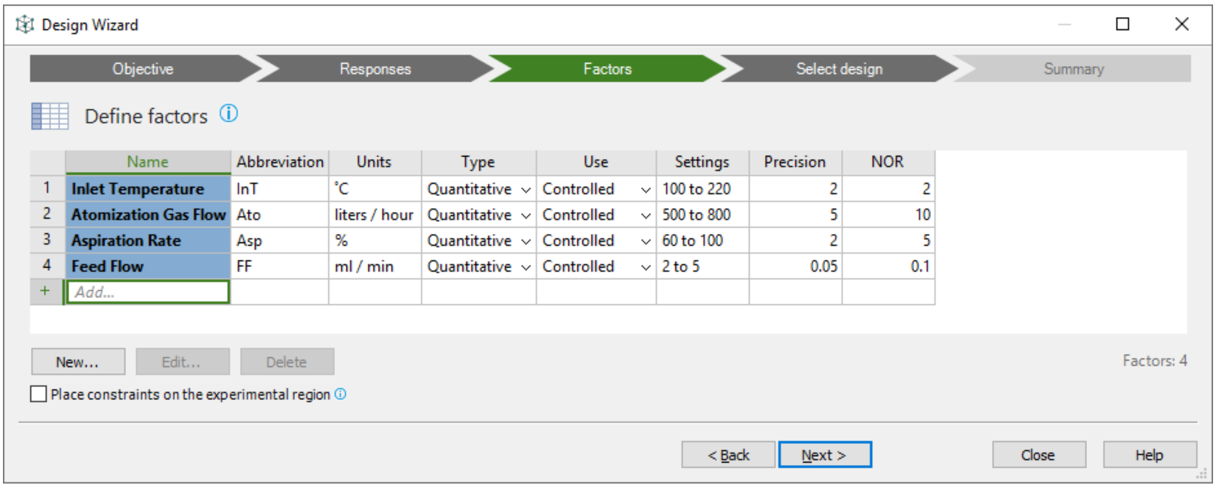
Select File/New/Experimental Design/Screening and step through the Design Wizard as shown below. Click Next.



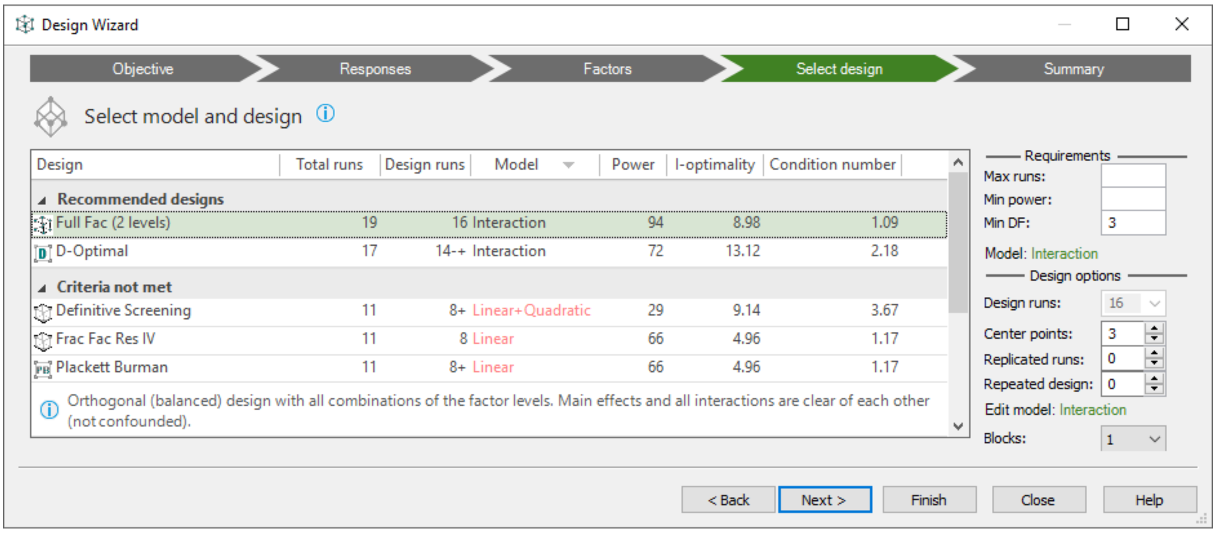
On the responses page, Click New and enter the name of the first response. Set Condition to Desired and Objective to Maximize. Click Add another. Enter the name of the second response. Set Condition to Required and Objective to Inside. Enter the value Min = 0.5 and Max = 3.5. Click Add another. Enter the name of the third response. Set Condition to Required and Objective to Inside. Enter the value Max = 3.5. Click Add another. Enter the name of the fourth response. Set Condition to Observed and Objective to Predicted. Click Add another. Enter the name of the last response. Set Condition to Required and Objective to Minimize. Enter the value Target = 0.2 and Max = 1. Click OK. The five responses have now been defined. Click Next.



On the factors page, Click New and enter the name, abbreviation, unit, settings, precision and NOR for the first factor. Click Add another and fill in the name, abbreviation, unit, settings, precision and NOR for the second factor. Repeat to add the third factor. Repeat to add the fourth factor. Click on OK. The four factors have now been defined. Click Next.



Select the Full Factorial design with 16 design runs. Verify that the number of center points = 3 and Total runs = 19. Click Next.



On the final Summary page you can review your selections and settings, which should look like the screenshot below. Click Finish to exit the design wizard.

Enter the response data or copy them from the end of this document.

Analyze the Data

Analysis Wizard

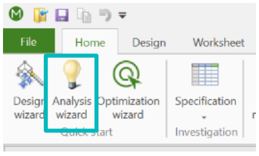
The Analysis wizard can be executed in two modes: Clicking One-Click to automatically transform the data and tune the model if appropriate or stepping through the pages in a manual mode using Next and investigating each page.

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.

One-Click

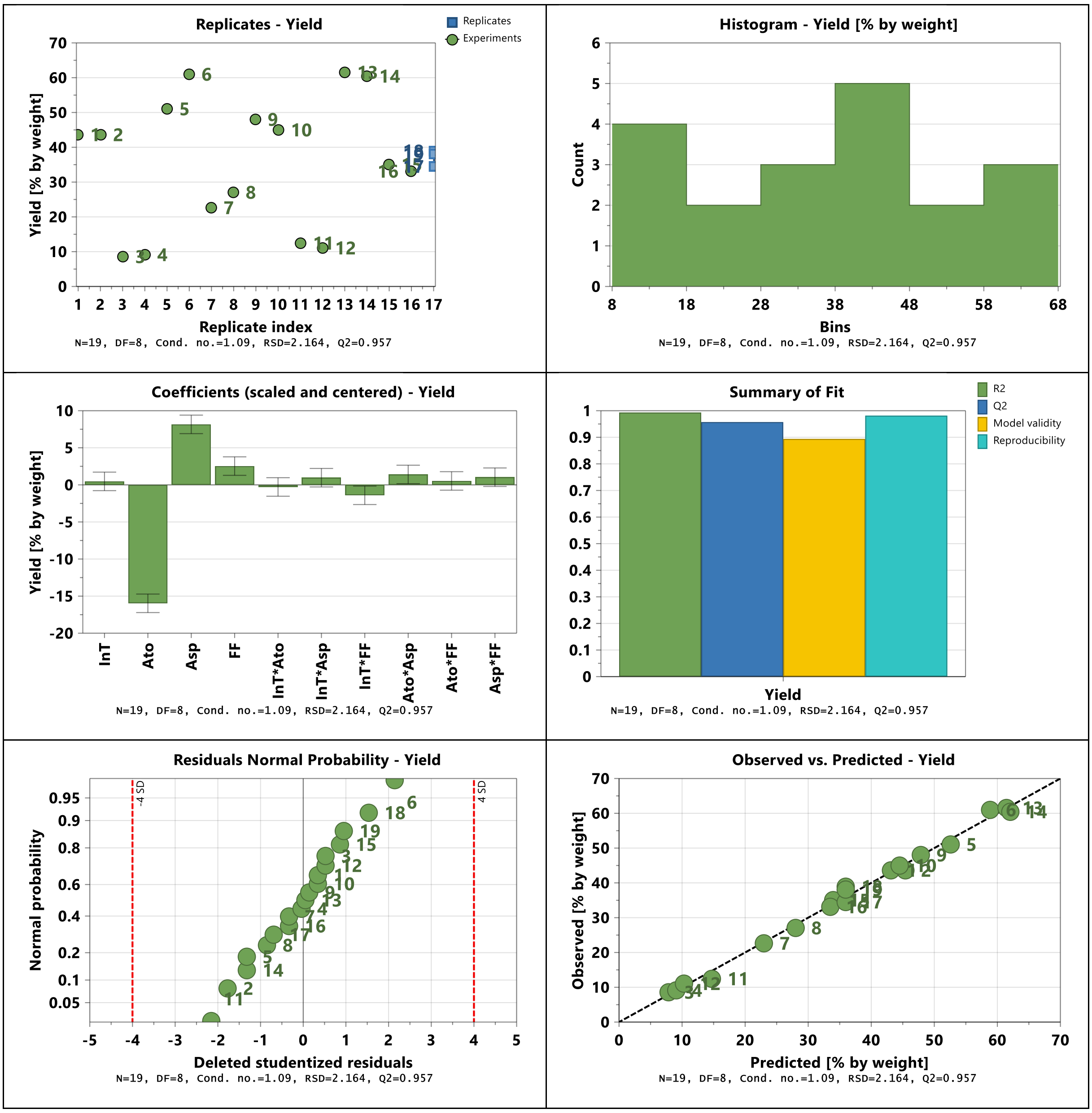
When stepping through the Analysis wizard using One-Click the tests available on each page are automatically performed and transformation of the data and/or tuning the model is done automatically. If there is a warning that cannot be handled automatically, the wizard stops at that page and you can decide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



Start the analysis wizard and work through the responses, guidance can be found in the panel to the right. Evaluate the raw data. Is there any need for data pretreatment, for example transforming the response data? Consider the responses one by one and try to find best possible model. Which factors are important? Are there any non-significant model terms? Are the residuals approximately normally distributed? Refine the model, if necessary.

The analysis wizard was used to work through the analysis steps for all responses.

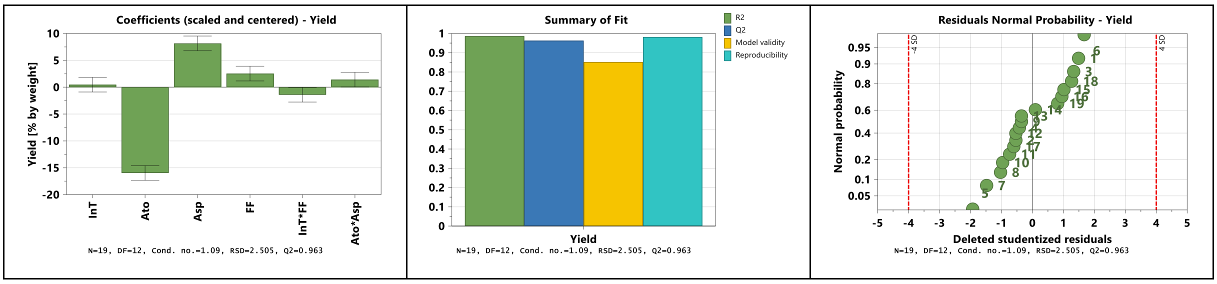
Response 1 (Yield)



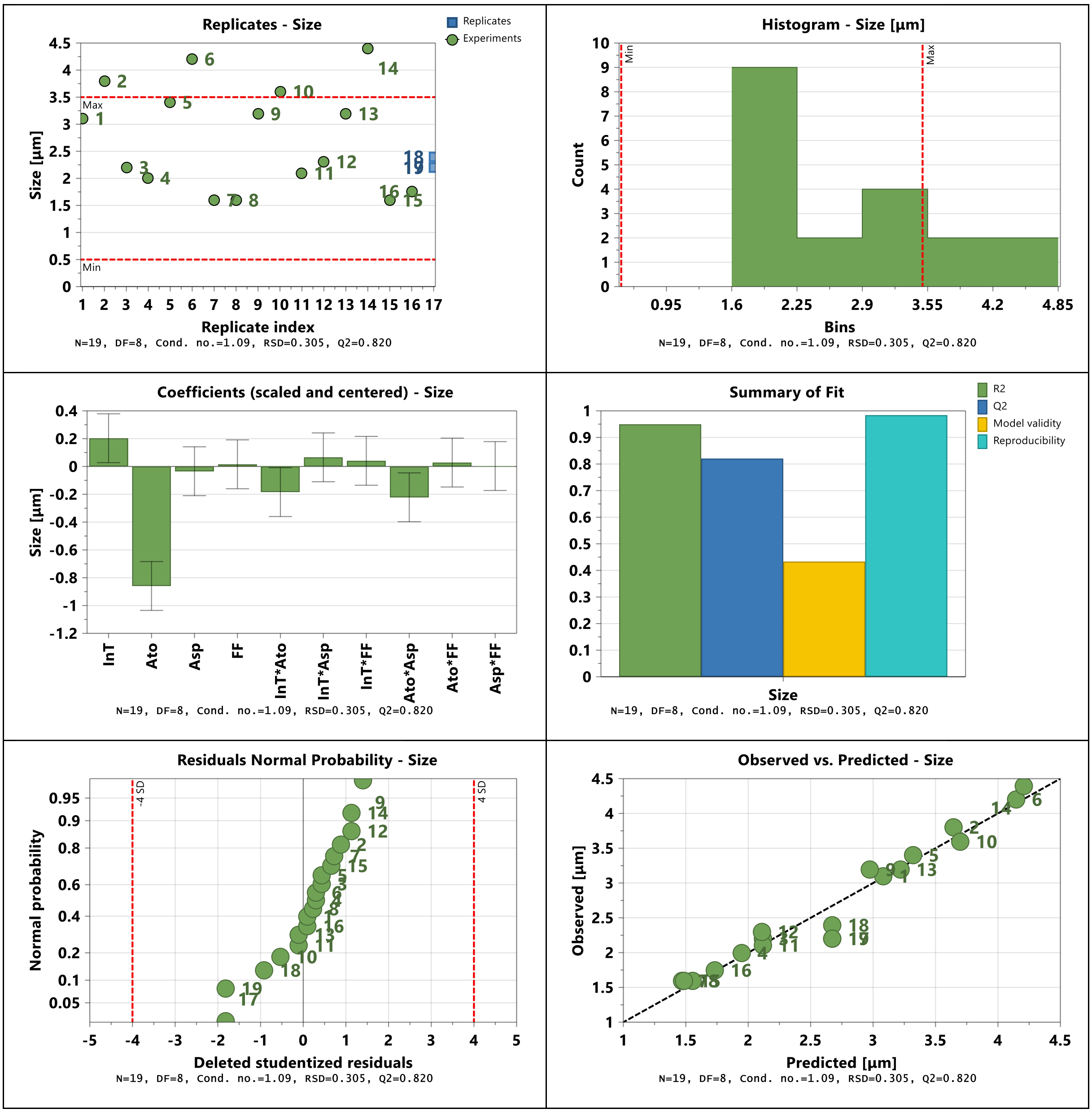
The model is good and the residuals are normally distributed with no outliers, the proof for a valid model. However, the coefficient plot shows that some interaction terms are small, so these can be removed to simplify the model, gain Df, and get the most accurate predictions (good Q2).

In the coefficient plot page the exclude tool was used to remove the small and insignificant interaction terms. Q2 was used as the criterion to determine whether the model improved, starting with the smallest coefficient and excluding them one at a time. This procedure resulted in the following model diagnostics.

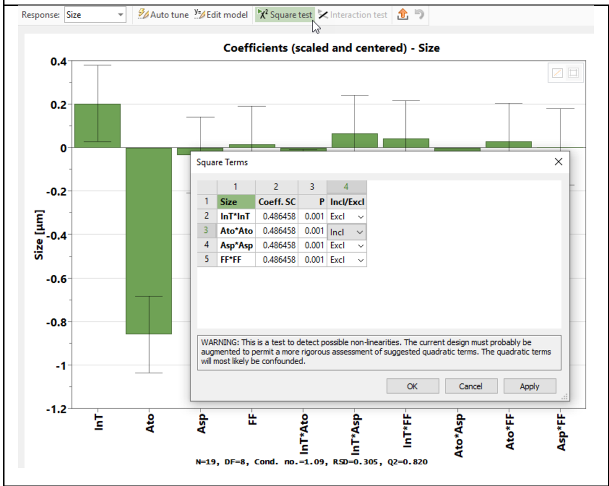
Note: If you are running the autotune option the model will contain eight coefficients. The model seen below has only six coefficients. There is a marginal difference in Q2 of 0,8%. Due to the parsimony principle, the model with six coefficients is the preferred one.

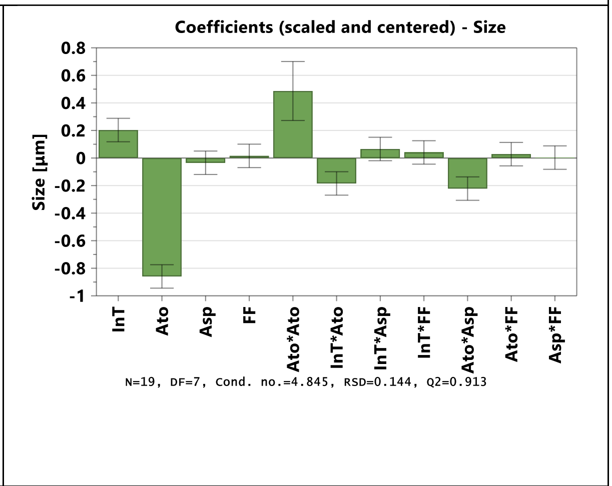


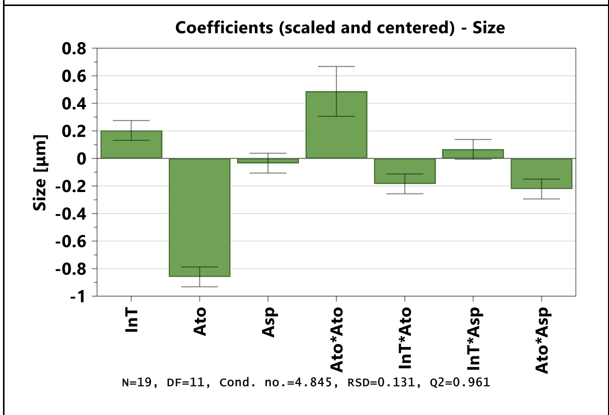
Response 2 (Size)

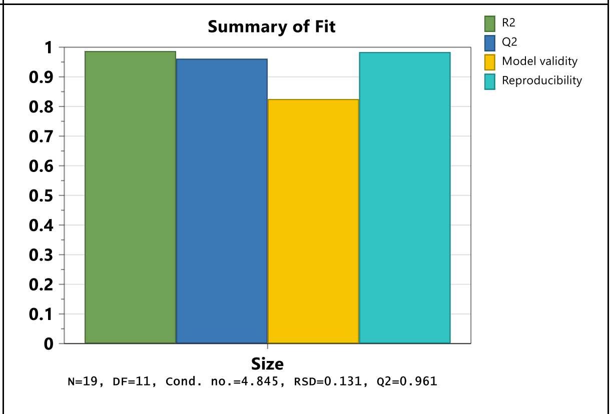


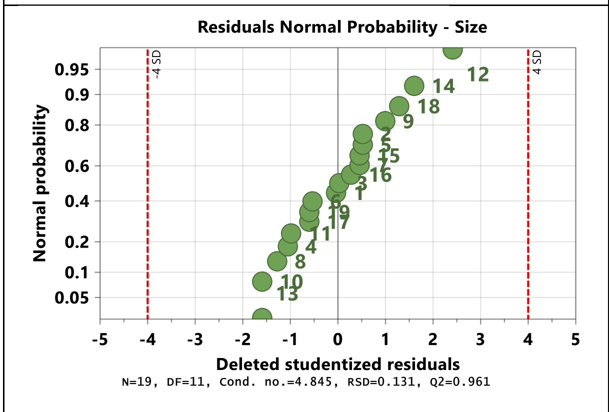
The replicate error is small but the distribution is skewed and a log-transform of the response data might be preferable for the model performance. The skewness test does not recommend a transformation, however.

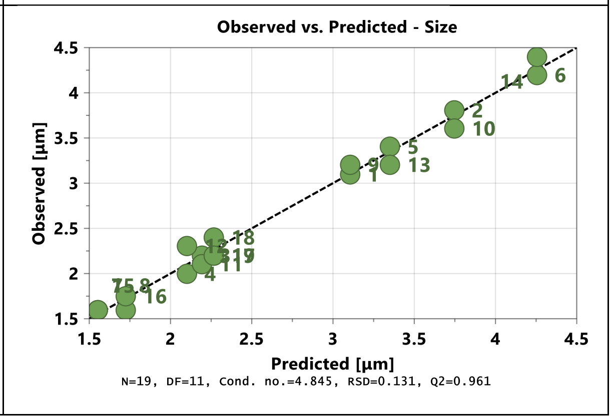








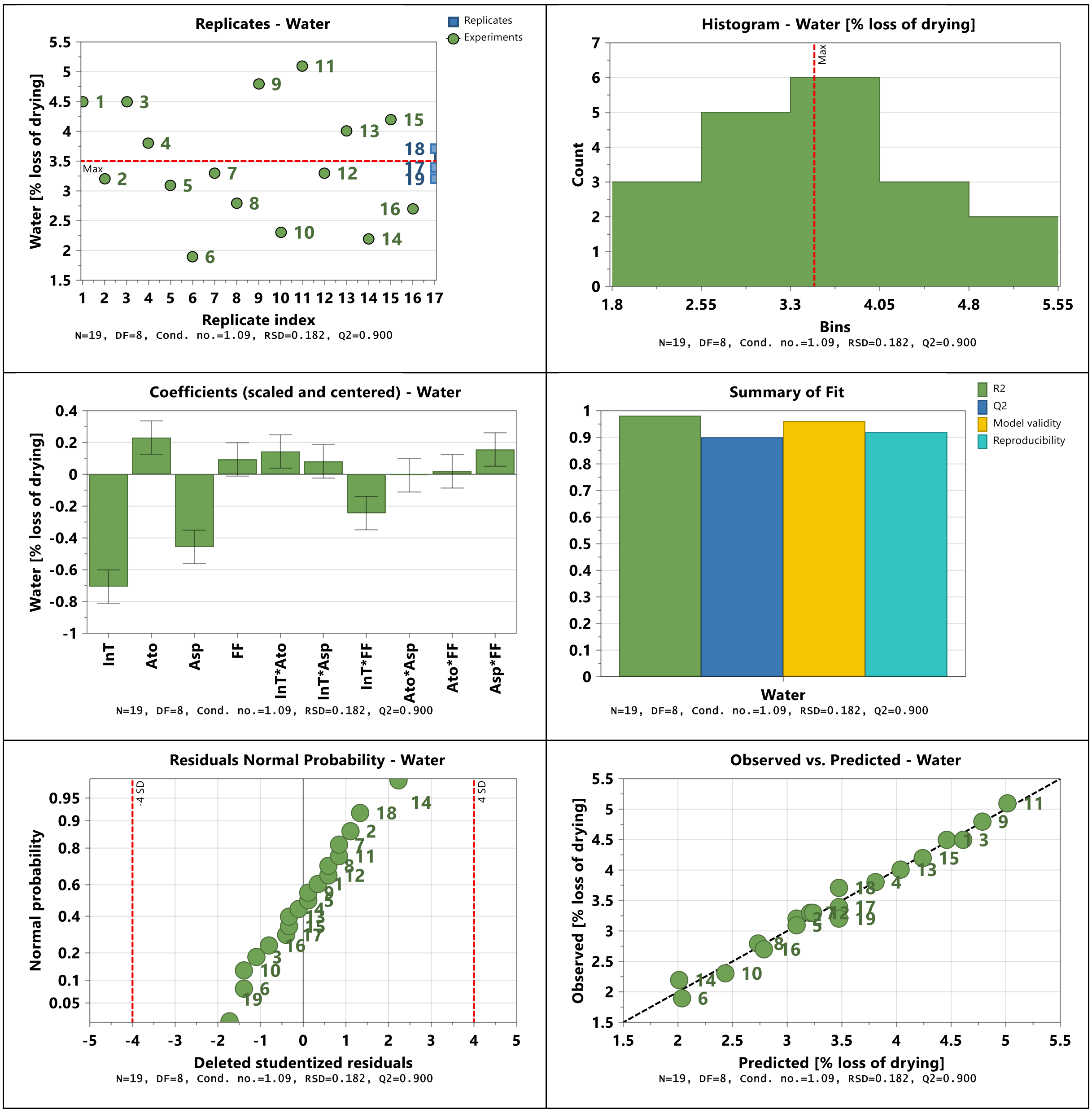




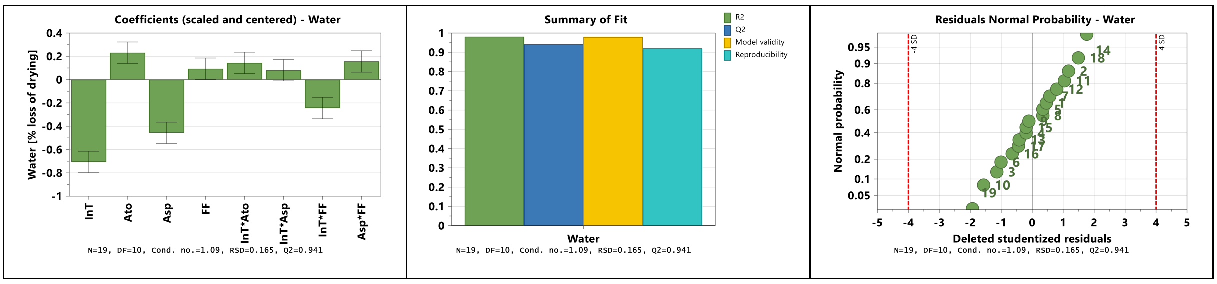
The square test has detected a non-linear phenomenon. The term Ato^2 was added to the model because Ato is the largest main effect. This square term is confounded with the other square terms, and more experiments would be needed to fully resolve the confounded square terms.

The final model after autotuning has good overall modelling statistics and normally distributed residuals.

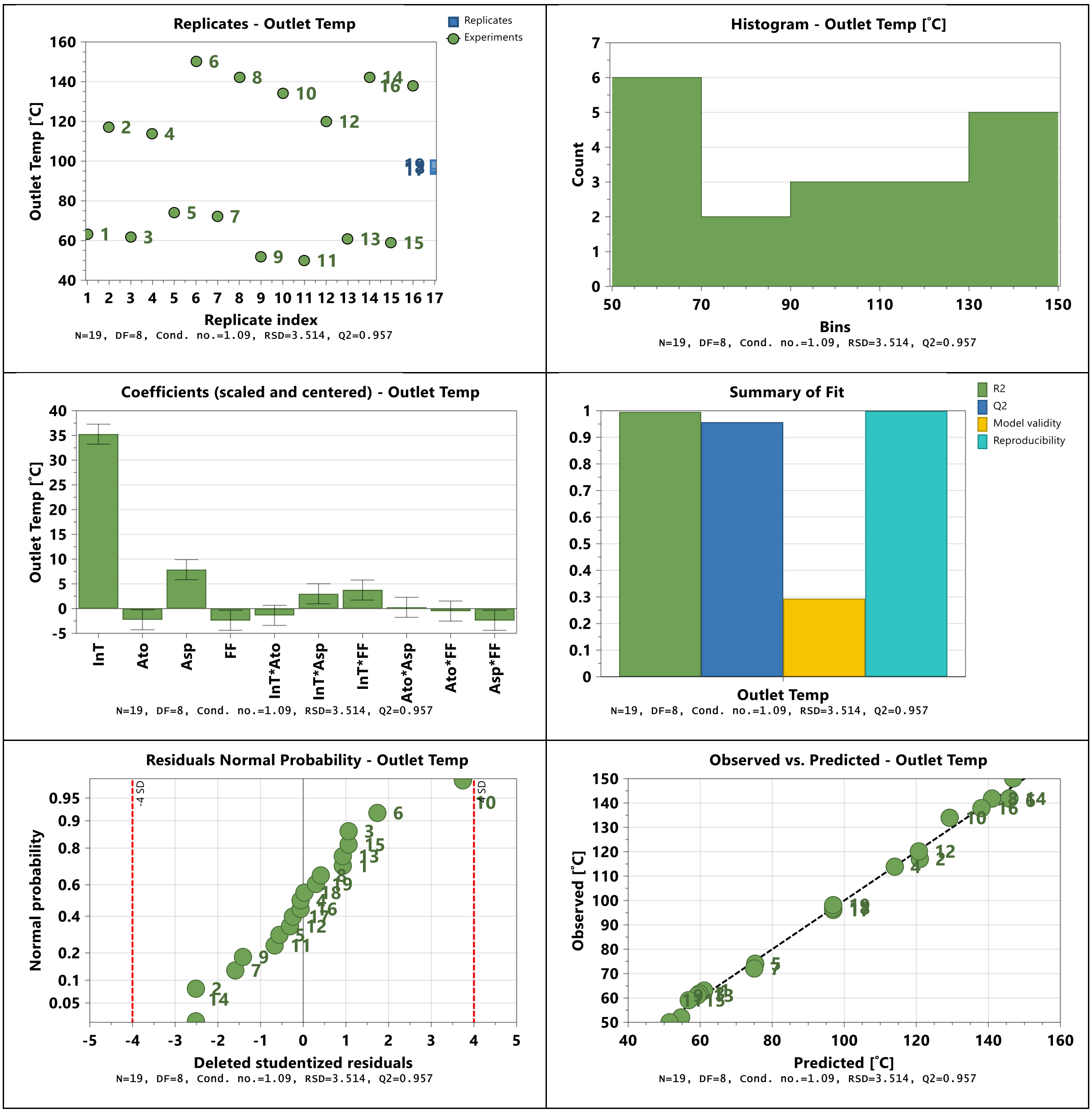
Response 3 (Water)



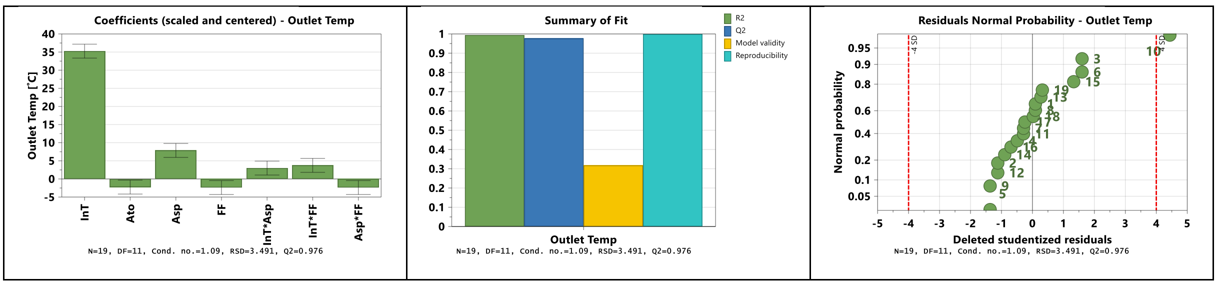
In this case everything looks good. The only thing to do is to remove insignificant terms from the model. Removing the insignificant terms produced the following result.



Response 4 (Outlet Temp)

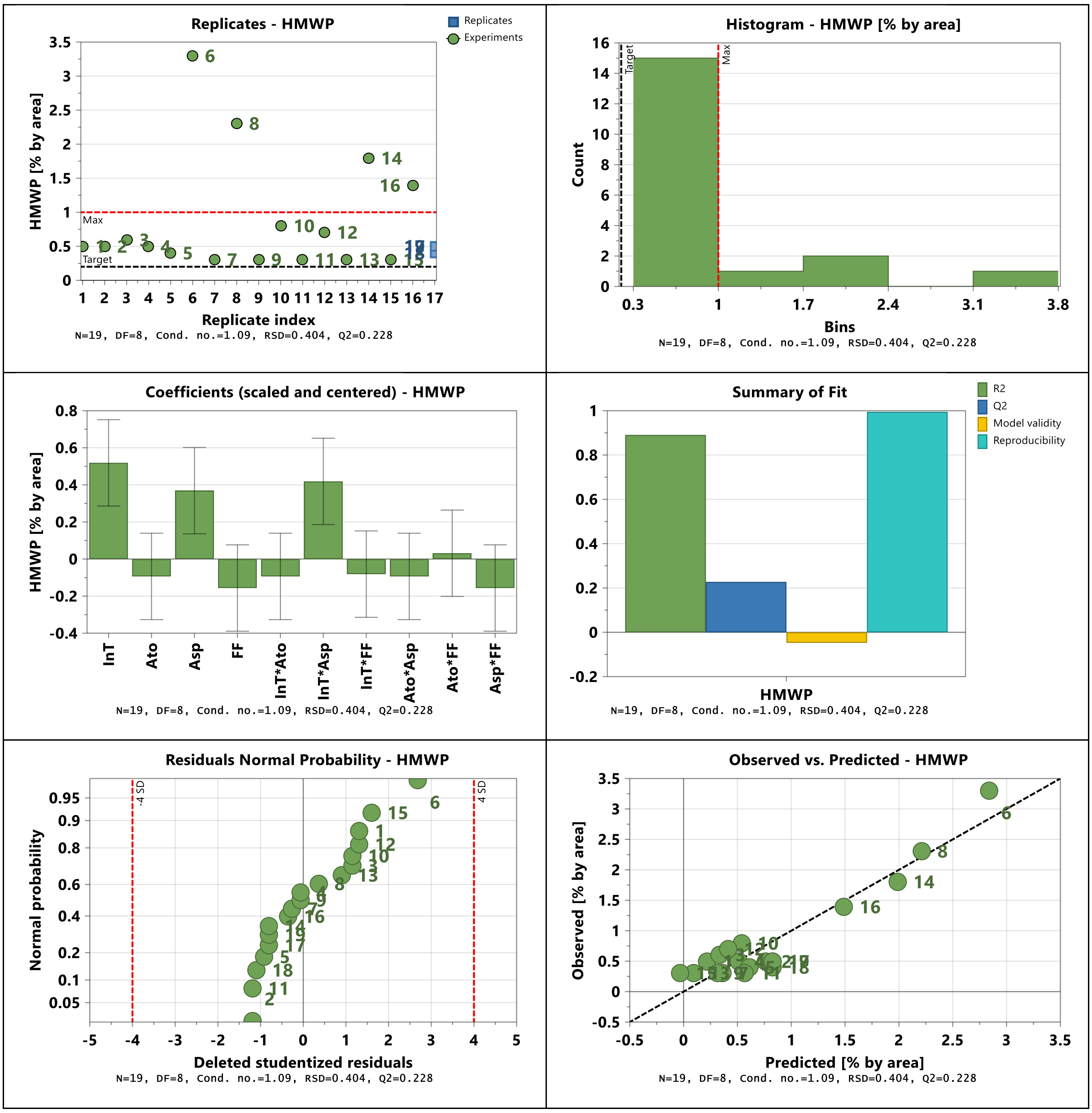


The replicate plot shows that there is limited spread among the replicates. The histogram shows a peculiar distribution and indicates one dominating factor. The residuals are mainly normally distributed but one experiment (10) is on the border to be an outlier and could be checked (typos or other problems). Insignificant terms should be removed from the model. Removing the insignificant terms produced the following diagnostics.

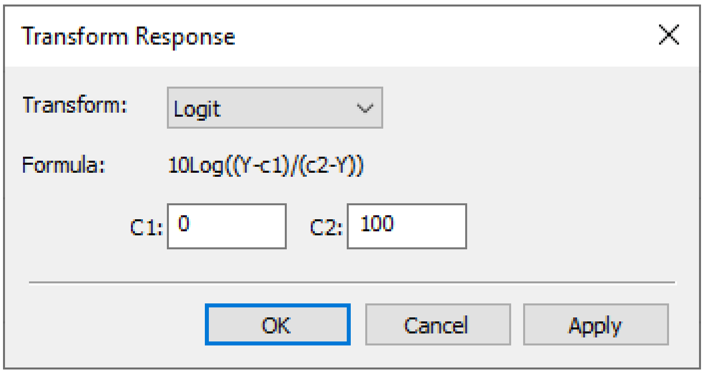


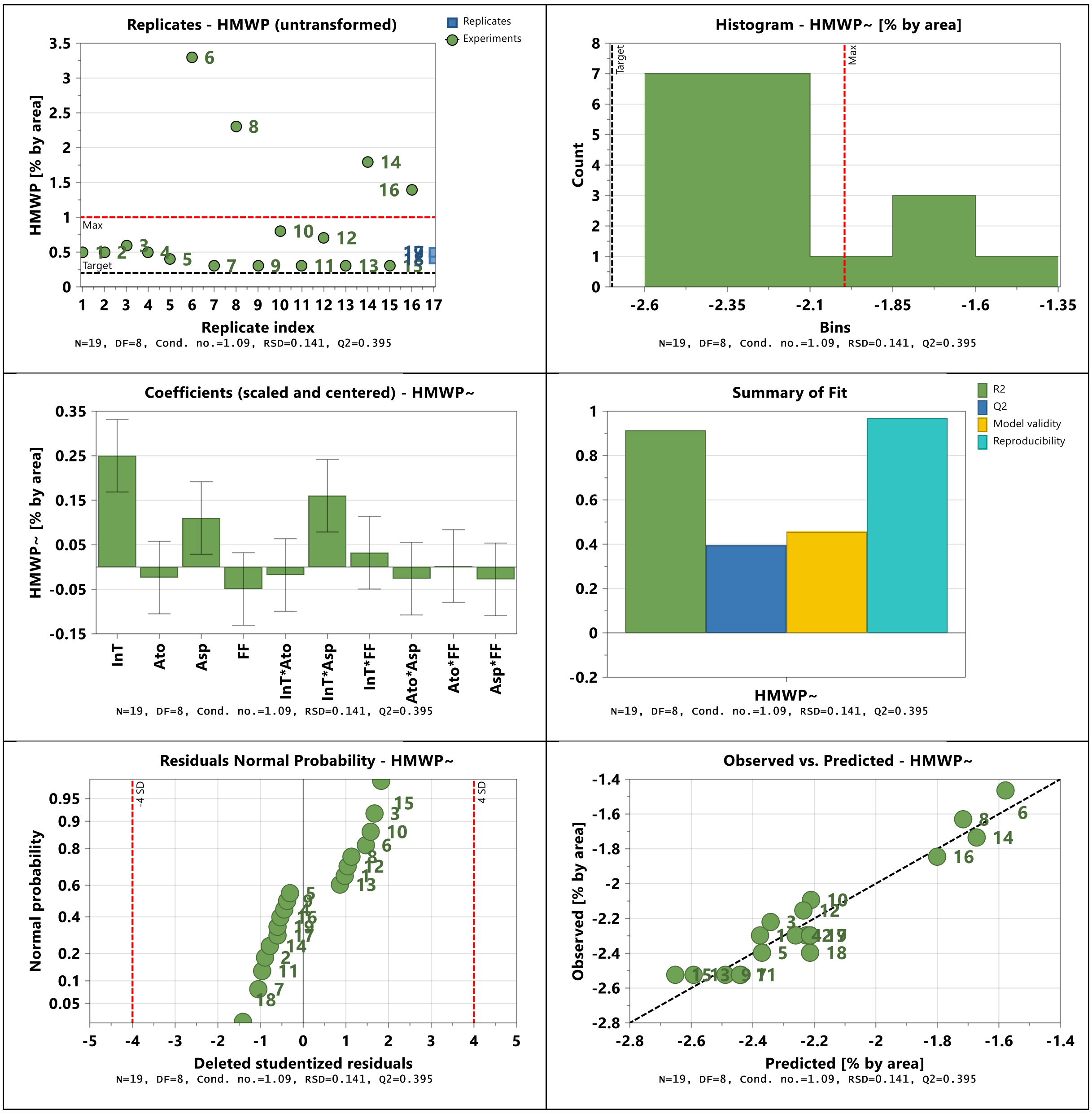
The model validity is a little low and experiment 10 is a statistical outlier. The model is overall very good (high R2 and Q2). Removing this experiment leads to higher validity but will that be a more reliable model? The influence of the deviating run number 10 is quite small due to rather high Df. Recommendation is to always keep all data if no obvious fault is confirmed.

Response 5 (HMWP)

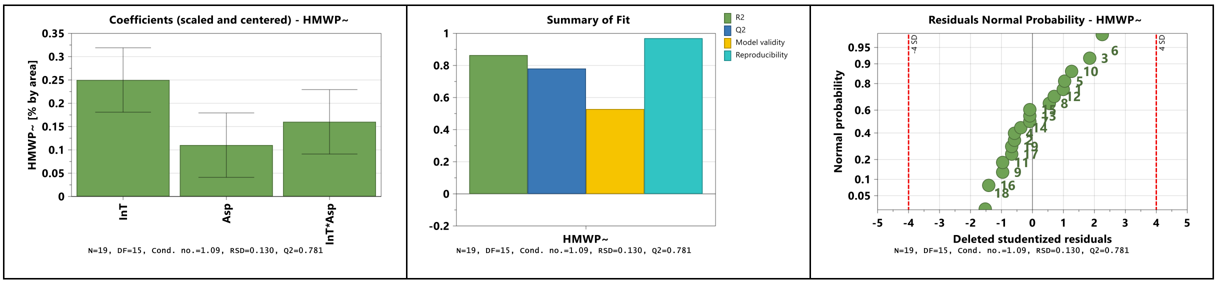


The replicate plot shows some experiments out of specification with some accelerated pattern. The histogram shows a corresponding skewed distribution and the response data should be log transformed. The model is poor (low Q2 and negative model validity). Log-transforming the response data results in the following diagnostics.

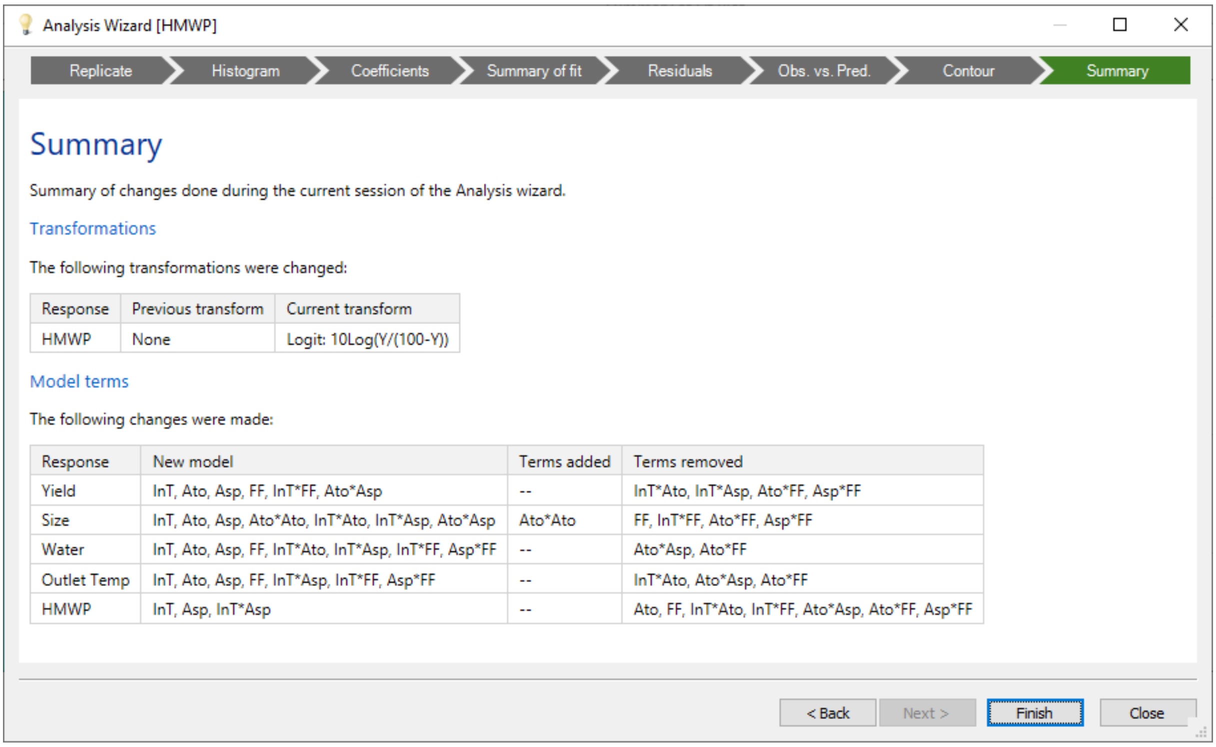




The transformation improves model statistics significantly. The model needs trimming, i.e. insignificant terms need to be excluded. After trimming, the model statistics improved significantly. Note: If you are running the autotune option the model will contain four coefficients. The model seen below has only three coefficients. There is a marginal difference in Q2, less than 1%. Due to the parsimony principle, the model seen below is the preferred one.



Now the work in the Analysis wizard is completed for each response. The final summary page of the wizard documents the model revision steps undertaken for each response. Using these models, we will now try to identify a setpoint at which to operate the system.


Analysis Wizard [HMWP]

Replicate > Histogram > Coefficients > Summary of fit > Residuals > Obs. vs. Pred. > Contour > Summary

Summary

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

Transformations

The following transformations were changed:

Response	Previous transform	Current transform
HMWP	None	Logit: 10Log(Y/(100-Y))

Model terms

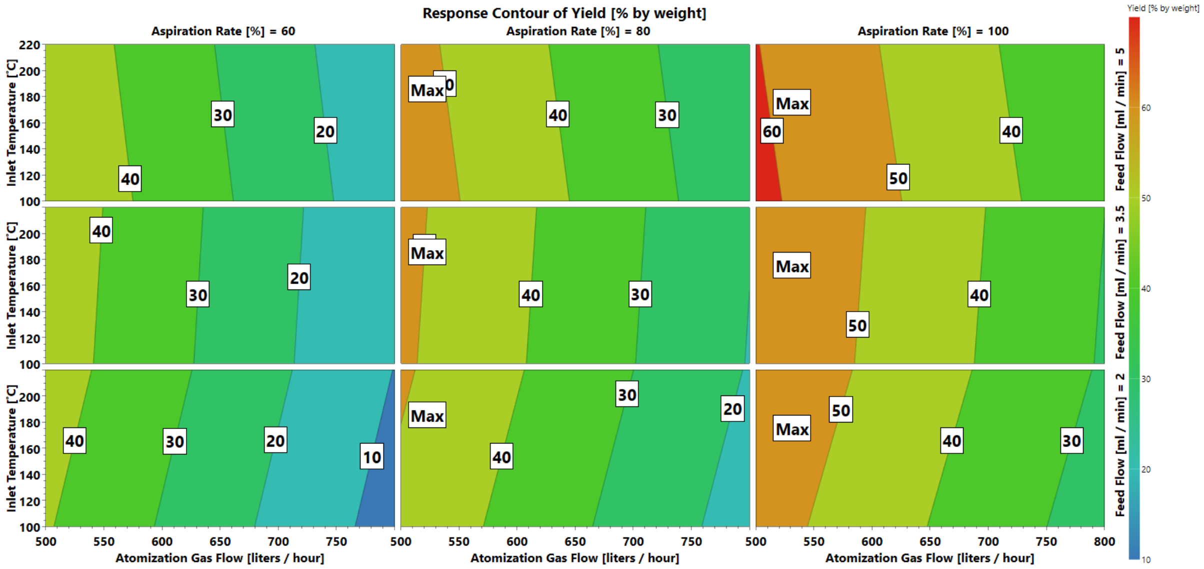
The following changes were made:

Response	New model	Terms added	Terms removed
Yield	InT, Ato, Asp, FF, InT*FF, Ato*Asp	--	InT*Ato, InT*Asp, Ato*FF, Asp*FF
Size	InT, Ato, Asp, Ato*Ato, InT*Ato, InT*Asp, Ato*Asp	Ato*Ato	FF, InT*FF, Ato*FF, Asp*FF
Water	InT, Ato, Asp, FF, InT*Ato, InT*Asp, InT*FF, Asp*FF	--	Ato*Asp, Ato*FF
Outlet Temp	InT, Ato, Asp, FF, InT*Asp, InT*FF, Asp*FF	--	InT*Ato, Ato*Asp, Ato*FF
HMWP	InT, Asp, InT*Asp	--	Ato, FF, InT*Ato, InT*FF, Ato*Asp, Ato*FF, Asp*FF

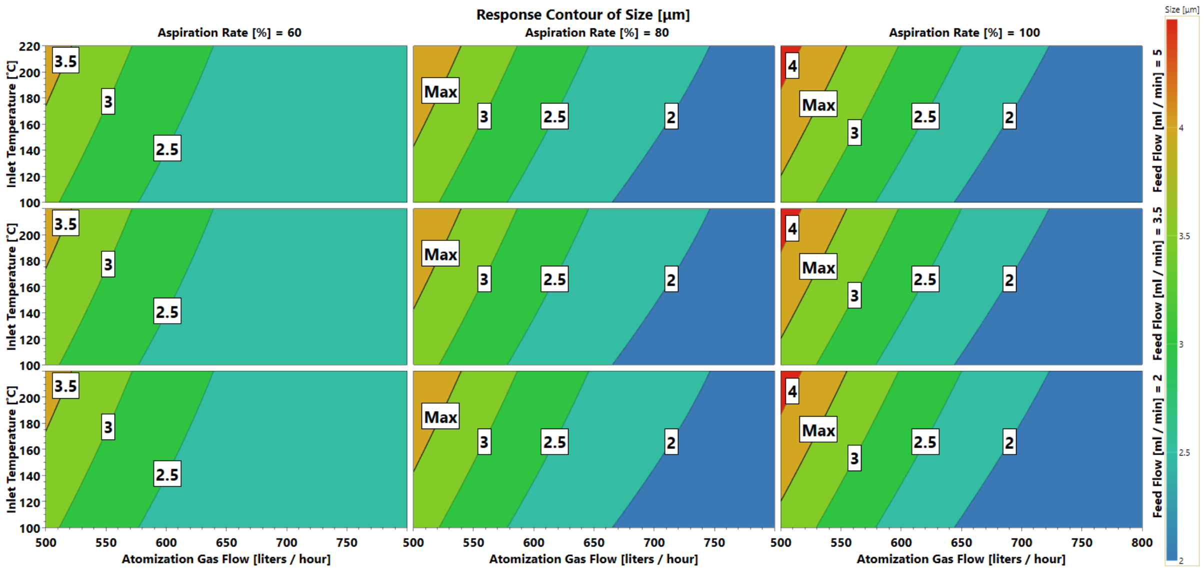
< Back | Next > | Finish | Close

Below, 4D contour plots from the Analysis wizard are given for the individual responses. Such 4D contour plots help in exposing the best condition for each response.

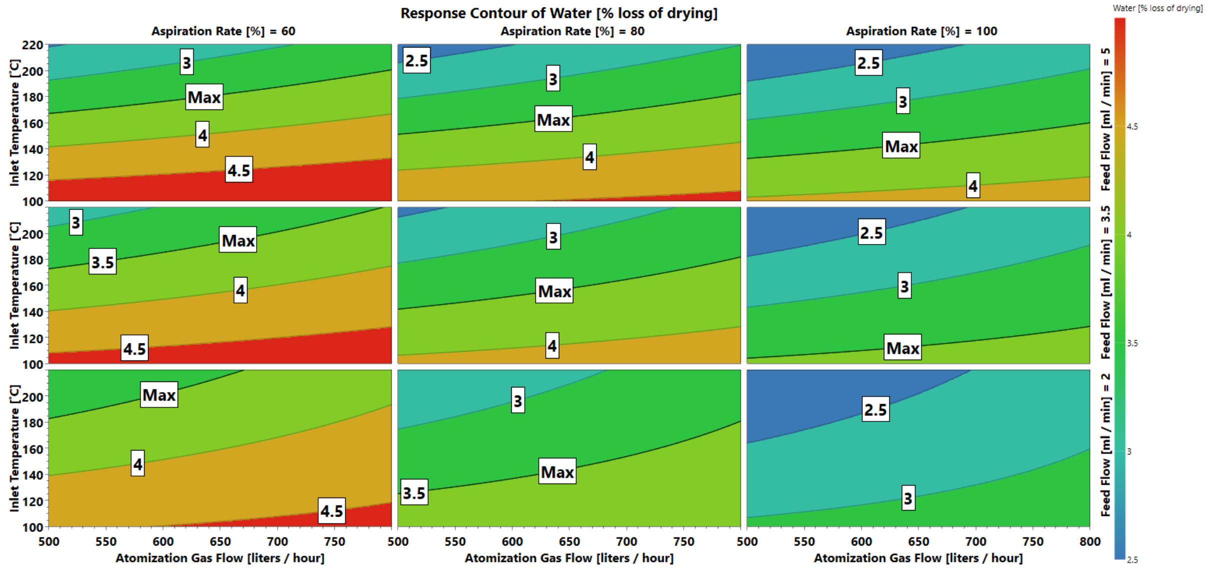
4D contour plot for response Yield.



4D contour plot for response Size.

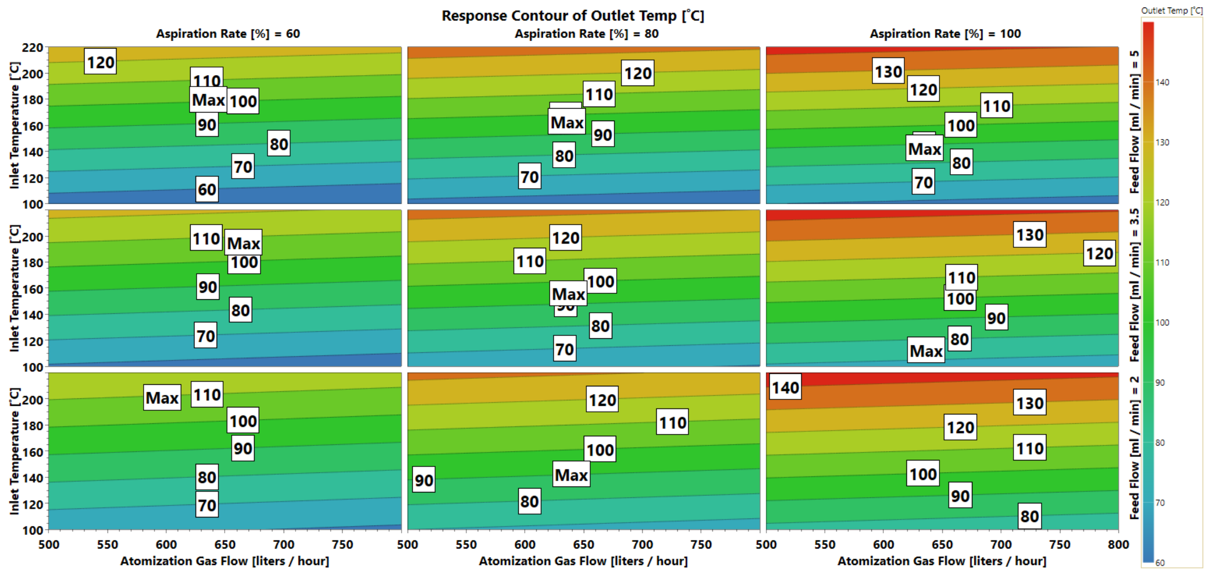


4D contour plot for response Water.

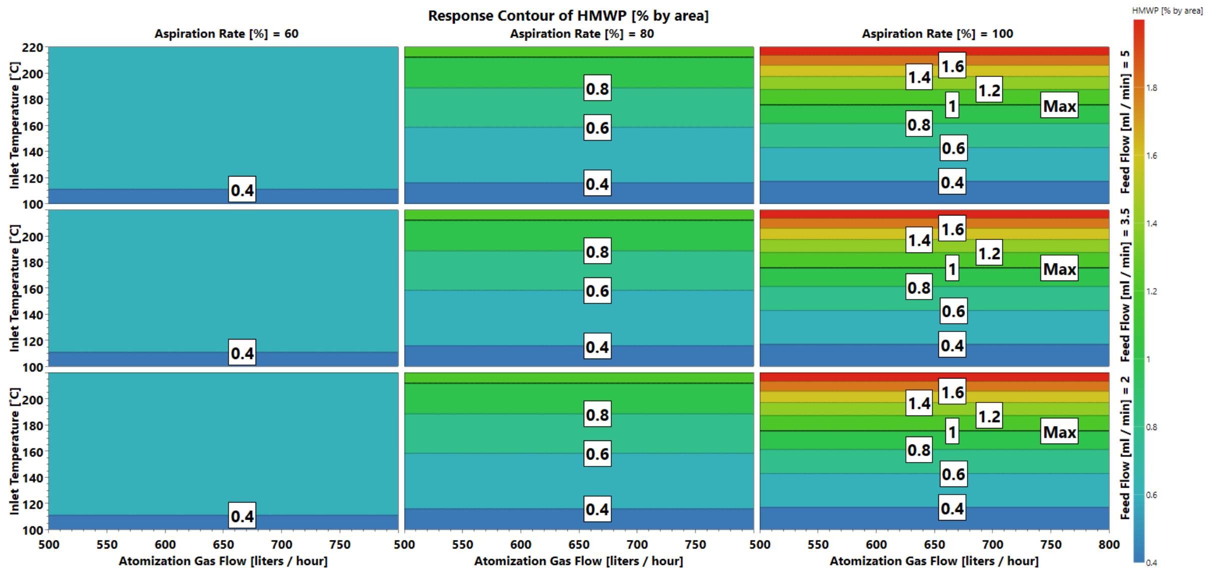


Response Contour of Water [% loss of drying]
Aspiration Rate [%] = 60
Aspiration Rate [%] = 80
Aspiration Rate [%] = 100
Inlet Temperature [°C]
Atomization Gas Flow [liters / hour]
Feed Flow [ml / min] = 5
Feed Flow [ml / min] = 3.5
Feed Flow [ml / min] = 2
Max

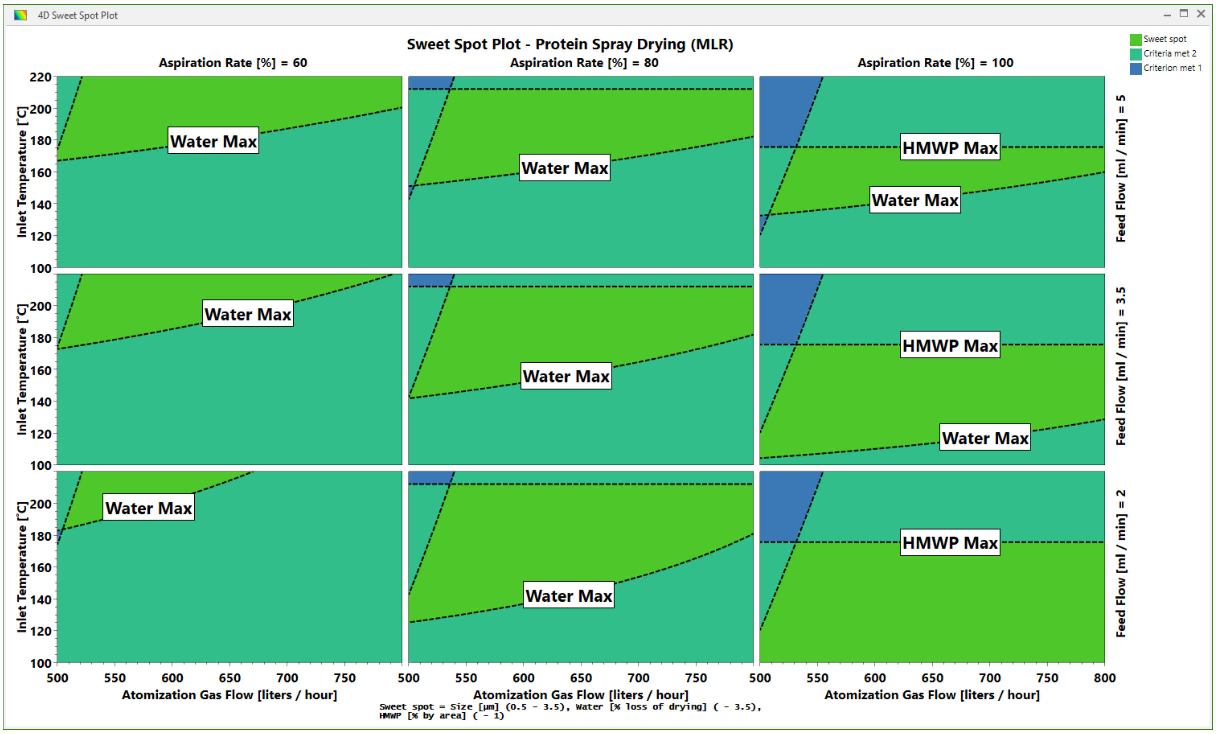
4D contour plot for response Outlet Temp.



4D contour plot for response HMWP.



The 4D Sweet spot plot seen below can be thought of as overlay of the five 4D contour plots and colored according to the different response criteria. Evidently, a sweet spot volume exists.

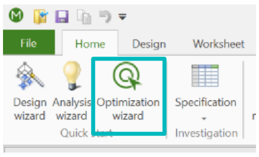


Optimize the Process

Optimization Wizard

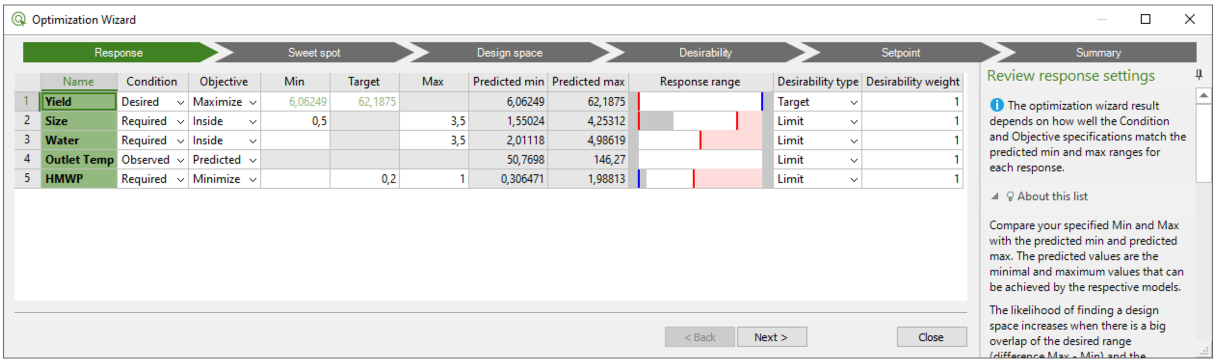
In optimization the first goal is to define the design space and then identify a suitable setpoint that fulfills all required limits and at the same time maximize the yield. With many factors and responses, it might be difficult to find the optimal solution to a problem and in many cases the final result might be a compromise between conflicting demands. The Optimization wizard in MODDE will help with this search for a solution.

Compare the various alternatives for setpoint identification. Discuss the various solutions and compare their possibilities and limitations.

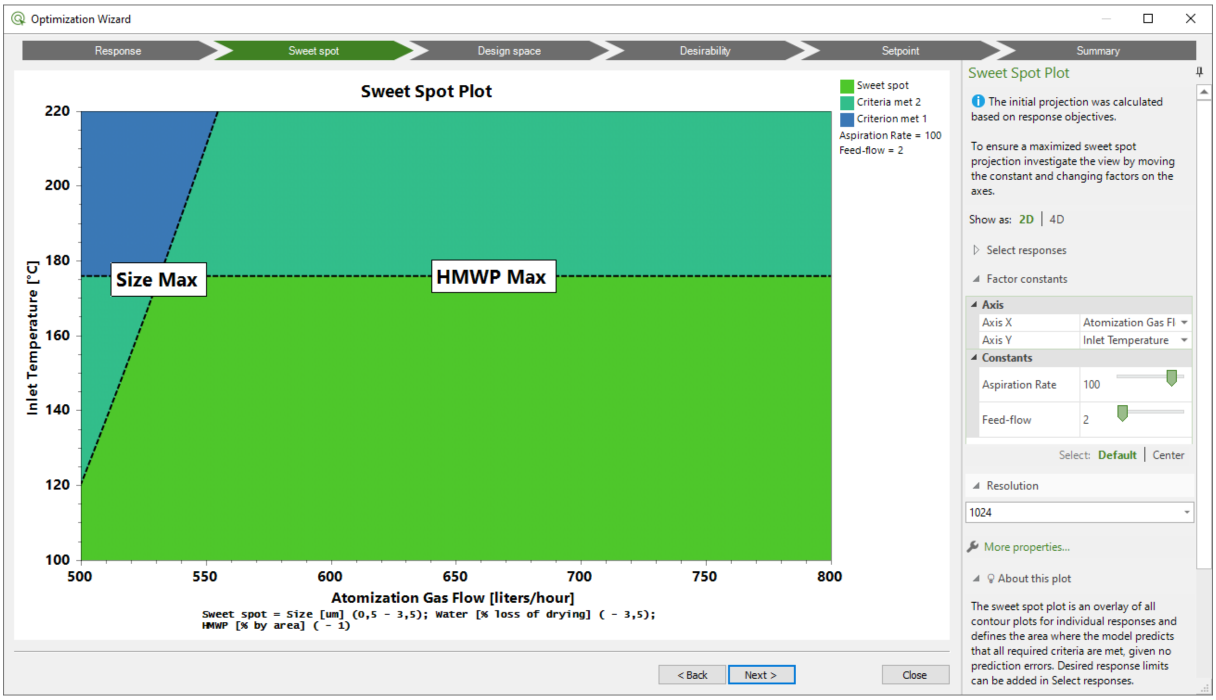


Launch the Optimization wizard and search for a good setpoint at which to operate the system.

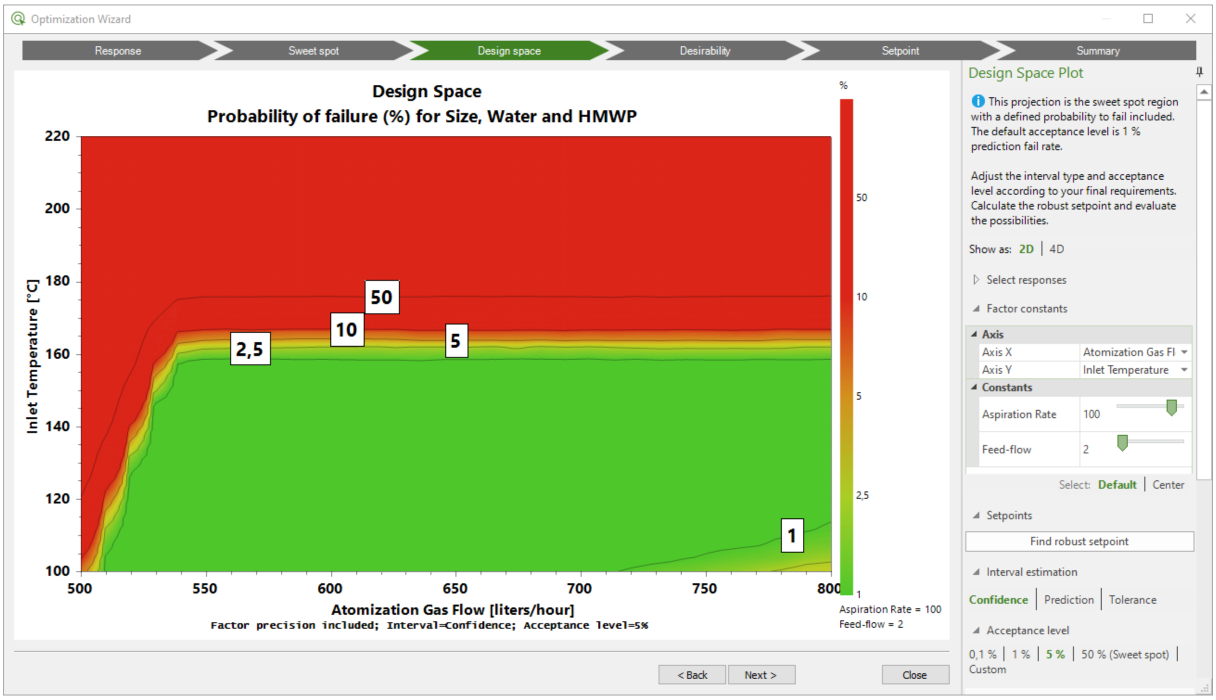
The first page in the Optimization wizard is the Response page, where settings and limits can be reviewed. In this description, the results from running through the Optimization wizard are based on the scenario seen in the screenshot below. Click Next.



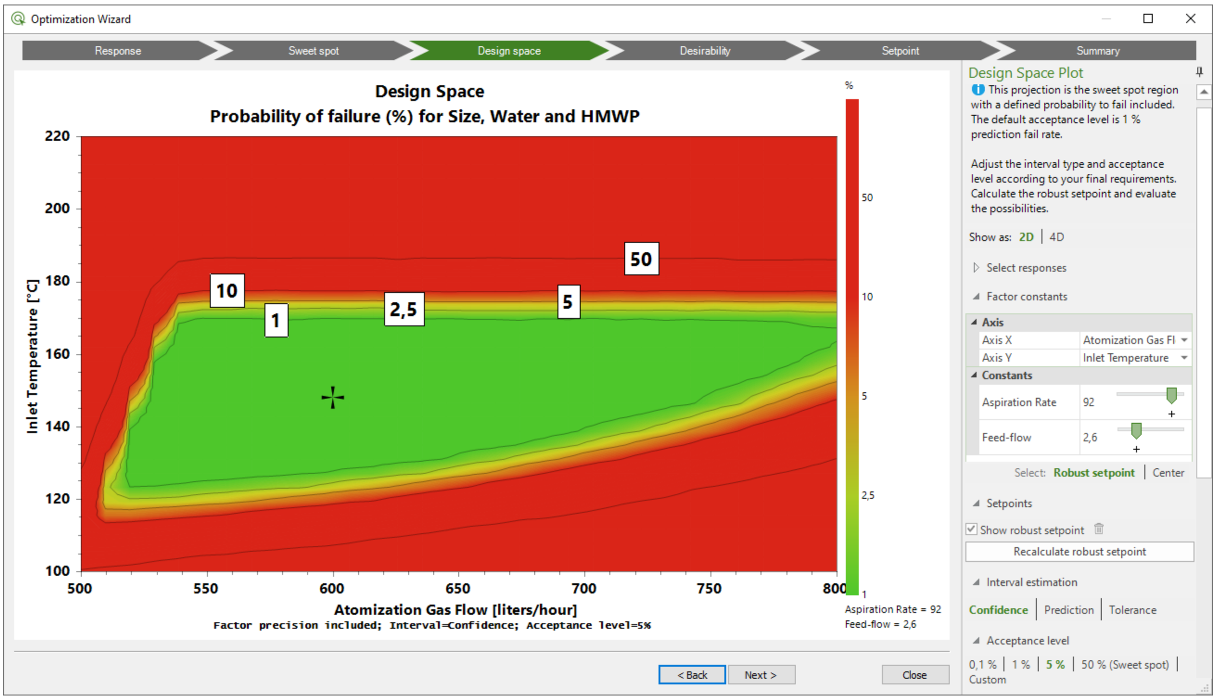
The Sweet spot plot displayed in the wizard is identical to the lower right-hand plot seen in the 4D Sweet spot plot created in Task 2. This is the largest sweet spot area in any of the possible two-dimensional factor subspaces that can be plotted using a 2D-plot. Click Next.



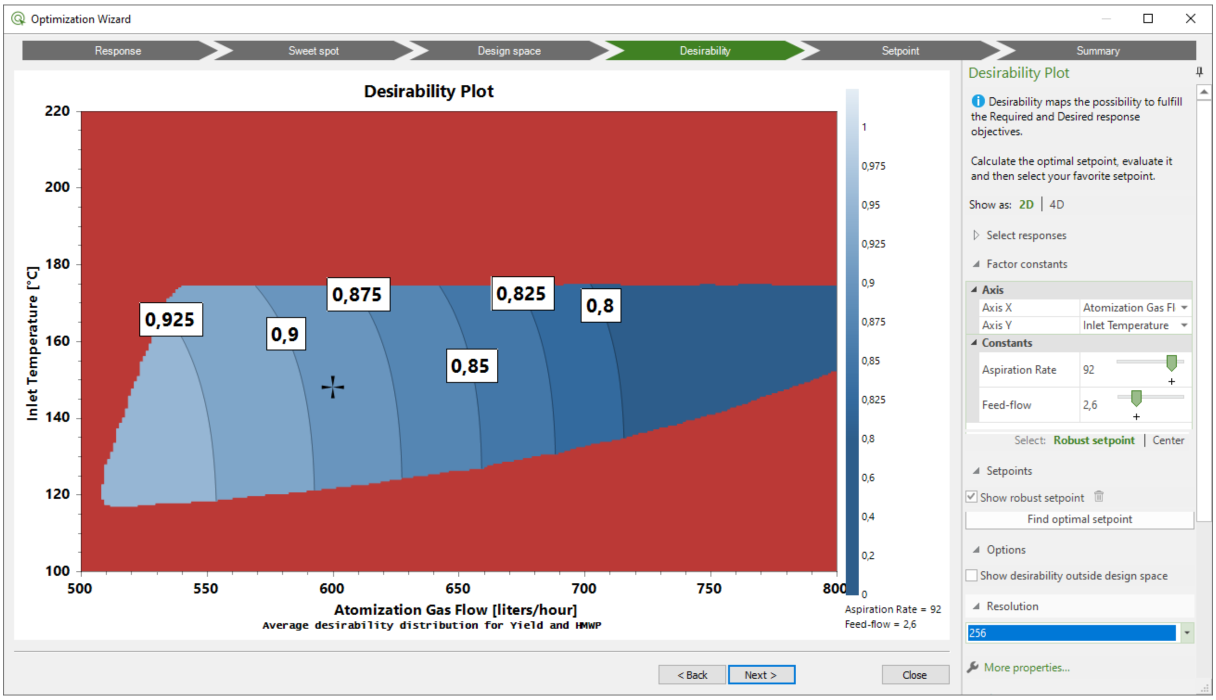
The design space plot indicates a design space can be postulated using a 5% acceptance level. Click on the button called Find robust setpoint to initiate the search for a robust setpoint.



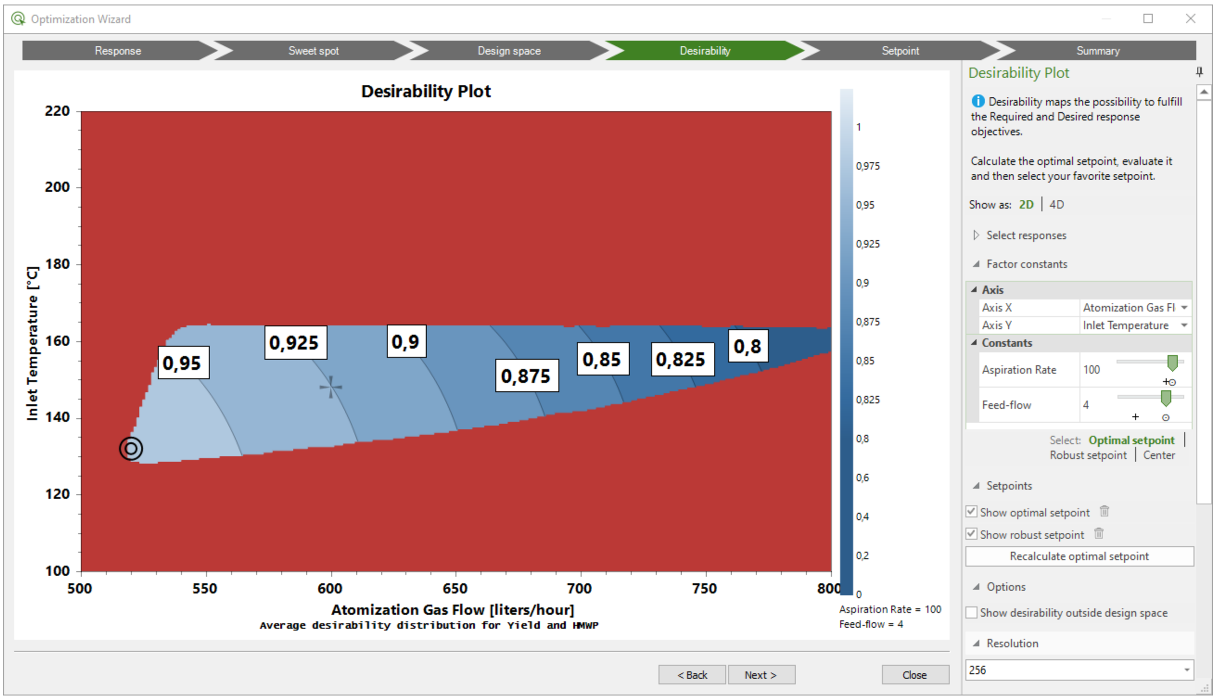
A robust setpoint was detected. Its position is indicated by the cross-hairs symbol. Note that the placement of the robust setpoint depends on the settings of the adjustable parameters (Resolution, etc.) and hence your results might be slightly different from the ones presented in the plot below. Click Next.



The coloring of the desirability plot is influenced by the shape of the design space. The closer to 1.0 the desirability the closer you are to an absolutely optimal solution. Click on Find optimal setpoint, which will be the point displaying the highest desirability.



The position of the Optimal setpoint is indicated by the double-circle symbol. The position of the Robust setpoint, which is not located in the plane shown below, is indicated by a grayed cross-hairs symbol. Click Next.



The next page in the Optimization wizard is the Setpoint page. On this page a setpoint comparison plot is provided. This plot shows simulated response distributions based on existing models and estimated future variation in factor settings. A blue-colored distribution represents simulation around the robust setpoint and a red-colored distribution shows variability at the optimal setpoint.

One important setting influencing the widths of the estimated response distributions is the used Factor distribution (see information pane in the right-hand part of the screenshot). See the below three screenshots for the changes taking place when using the settings None, Factor precision and Normal operating range.

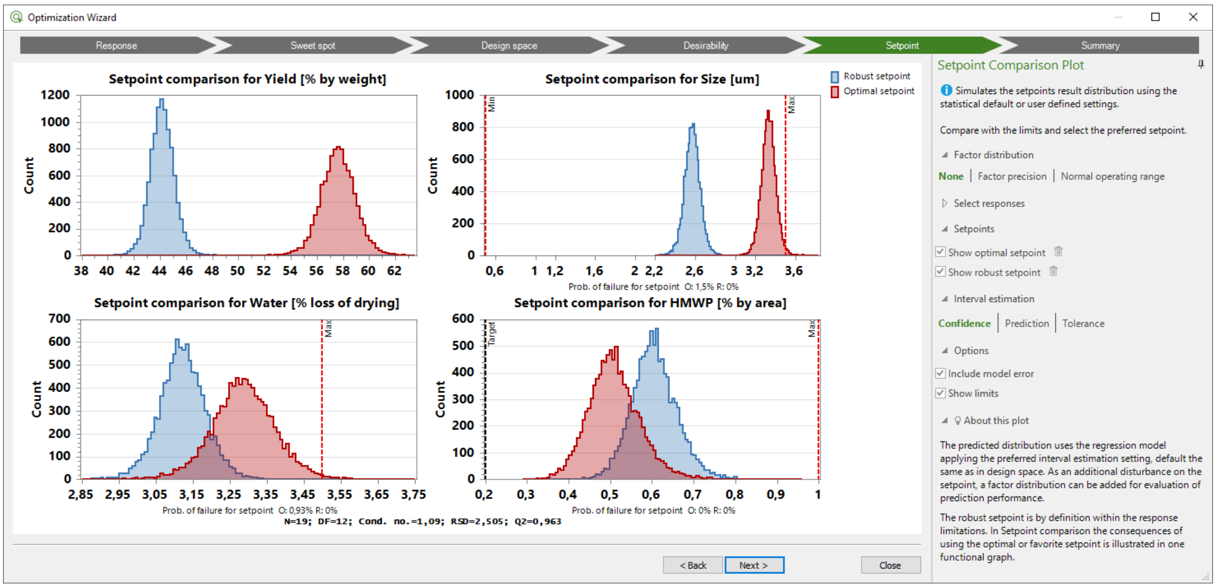
Factor precision

Precision is defined as the uncertainty in measurement of the true value of the factor setting and will influences the size of the design space.

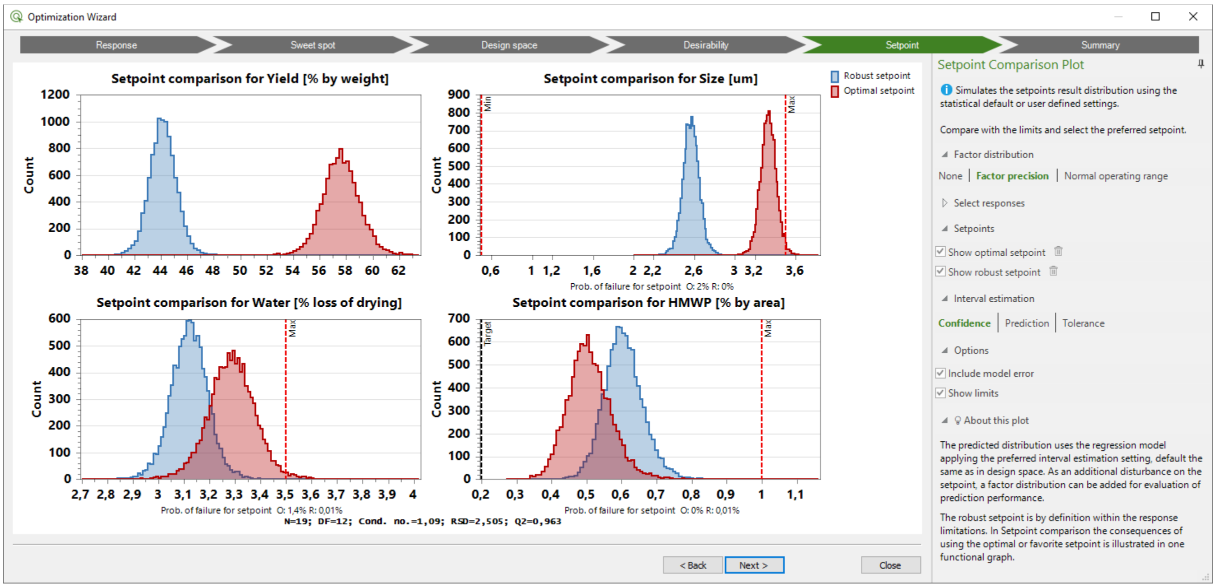
Normal Operation Range, NOR

Region around the setpoint that contains common operational variability. Represents what level of control of the factors that is practically achievable in day-to-day operation. Can be used for simulating future process output.

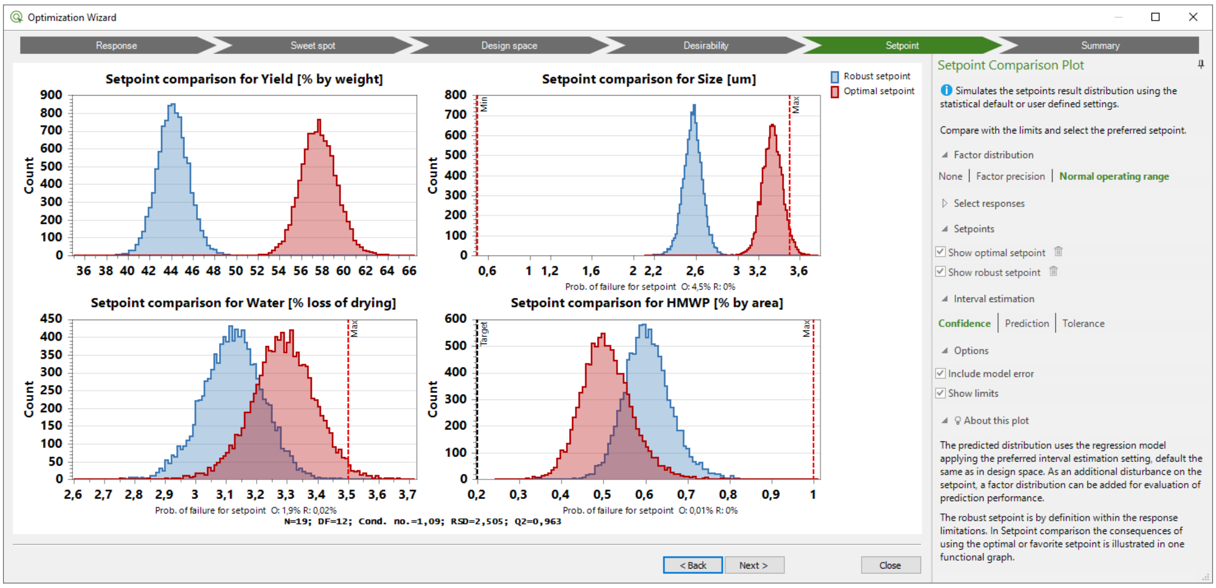
Factor distribution = None.



Factor distribution = Factor precision.



Factor distribution = Normal operating range.



More granular information regarding the different setpoints is provided by the final summary page of the Optimization wizard.

SARTORIUS

Optimization Wizard

Response > Sweet spot > Design space > Desirability > Setpoint > Summary

Response settings

Name	Condition	Objective	Min	Target	Max	Predicted min	Predicted max
Yield	Desired	Maximize				6,06249	62,1875
Size	Required	Inside	0,5		3,5	1,55024	4,25312
Water	Required	Inside			3,5	2,01118	4,98619
Outlet Temp	Observed	Predicted				50,7698	146,27
HMWP	Required	Minimize		0,2	1	0,306471	1,98813

Factor settings

Name	Units	Settings	Precision	NOR
Inlet Temperature	°C	100 to 220	2	2
Atomization Gas Flow	liters/hour	500 to 800	5	10
Aspiration Rate	%	60 to 100	2	5
Feed-flow	ml/min	2 to 5	0,05	0,1

Optimal setpoint

Probability of failure: 3,2% (interval: Confidence)

Response	Value	Unit	Prob.of failure	Cpk
Yield	57,621	% by weight		--
Size	3,335	um	2,1%	0,696
Water	3,288	% loss of drying	1,1%	0,787
Outlet Temp	86,703	°C		--
HMWP	0,505	% by area	0,001%	2,657

Factor	Value	Unit
Inlet Temperature	132,000	°C
Atomization Gas Flow	520,000	liters/hour
Aspiration Rate	100,000	%
Feed-flow	4,000	ml/min

Robust setpoint

Probability of failure: 0,002% (interval: Confidence)

Response	Value	Unit	Prob.of failure	Cpk
Yield	44,154	% by weight		--
Size	2,570	um	0%	3,779
Water	3,122	% loss of drying	0,002%	1,787
Outlet Temp	97,690	°C		--
HMWP	0,603	% by area	0%	2,403

Factor	Value	Unit	Robust low edge	Robust high edge	Hypercube low edge	Hypercube high edge
Inlet Temperature	148,000	°C	124,000	172,000	140,000	164,000
Atomization Gas Flow	600,000	liters/hour	520,000	760,000	540,000	640,000
Aspiration Rate	92,000	%	84,000	100,000	89,333	97,333
Feed-flow	2,600	ml/min	2,000	4,200	2,000	3,000

< Back | Next > | Finish | Close

Conclusions

The Analysis wizard helped in developing strong models for the five responses. The Optimization wizard helped in the search for a suitable setpoint. The Optimization wizard contains a functionality called setpoint comparison by which the consequences of selecting different setpoints can be evaluated.

As a comment to the setpoint comparison plots shown above it can be stated that regardless of setpoint selected the response Water remains the one hardest to meet the goals for, i.e., it has the highest probability of failure. As can be expected, the response distributions at the optimal setpoint are closer to the limits compared with the output for the robust setpoint. The search for the robust setpoint will drive the solution to a point with maximum distance to the border of the design space volume. By selecting the Optimal setpoint a higher yield can be obtained, however, at a slightly higher risk of failure.

Regardless of selected setpoint a further experiment needs to be conducted to verify the results at this point and future work could involve an optimization study centered around these settings.