# HPLC Robustness (Robustness Testing)

## In This Tutorial You Will Learn How to

- Set up a fractional factorial design using the Design Wizard
- Handle qualitative factors
- Verify if a response is robust to small changes in the factors
- Act to convert a non-robust system to become a robust one

## Background and Objective

The aim of robustness testing is to design a process, or a system, so that its performance remains satisfactory even when influential factors are allowed to vary. In other words, we want to investigate the system's sensitivity to *small* changes in certain critical factors. The advantages of this include a wider range of applicability of product, higher quality of product and at the same time a simpler process control. A robustness test is usually carried out before the release of an almost finished product, or analytical system, as a last test to ensure quality. Such a design is usually centered on a 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 present tutorial illustrating robustness testing 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 works reliably, and is reasonably insensitive to varying chromatographic conditions.

In chromatography, the objective is separation of the analytes within a reasonable time. Separation relies on different retention of each analyte on the stationary phase. Thus, the retention of each analyte is important, and this response is described by the capacity factor, k. The degree of separation between two analytes is estimated as the resolution between two adjacent peaks in the chromatogram. A resolution of 1 is considered as the minimum value for separation between neighboring peaks, but for complete baseline separation a resolution of >1.5 is necessary. As the resolution value approaches zero, it becomes more difficult to discern separate peaks.

We will use this example to illustrate 4 different outcomes of robustness testing and how to handle those outcomes;

1. Inside specifications and a significant model
2. Inside specifications and a non-significant model
3. Outside specifications and a significant model
4. Outside specifications and a non-significant model

The goal of this study was to maintain, consistently, a resolution of 1.5 or higher for all chromatographic conditions.

Documentation of robustness can be done with various statistical methods and in this tutorial we will demonstrate such features available in MODDE®.

## Example Dataset

The investigators explored five factors: (1) amount of acetonitrile in the mobile phase; (2) pH of the mobile phase; (3) temperature; (4) amount of the OSA counter-ion in the mobile phase; and (5) stationary phase batch (column). Note that the last factor is qualitative.

They then mapped the influence of these factors onto the chromatographic behavior of two chemical analytes.

Response specification:

Responses

| | Name | Abbreviation | Units | Condition | Objective | Min | Target | Max |
|---|---|---|---|---|---|---|---|---|
| 1 | **k1** | k1 | min | Observed | Predicted | | | |
| 2 | **k2** | k2 | min | Required | Target | 2.7 | 3 | 3.3 |
| 3 | **Res1** | Res | | Required | Maximize | 1.5 | 3 | |
| 4 | **PlateN** | Pla | | Required | Maximize | 4000 | 6000 | |
| + | *Add...* | | | | | | | |

Factor specification:

Factors

| | Name | Abbreviation | Units | Type | Settings | NOR |
|---|---|---|---|---|---|---|
| 1 | **AcN** | AcN | % | Quantitative | 25 to 27 | 1 |
| 2 | **pH** | pH | | Quantitative | 3.8 to 4.2 | 0.2 |
| 3 | **Temp** | Temp | °C | Quantitative | 18 to 25 | 3.5 |
| 4 | **OSA** | OSA | mM | Quantitative | 0.09 to 0.11 | 0.01 |
| 5 | **Column** | Col | | Qualitative | Column A, Column B | |
| + | *Add...* | | | | | |

Robustness testing is generally conducted using low resolution designs supporting linear models; the variability around a chosen setpoint is rather small and we can assume linearity in the investigated region. In this case the design selected was a Fractional Factorial Res III design with four center points.

Worksheet:

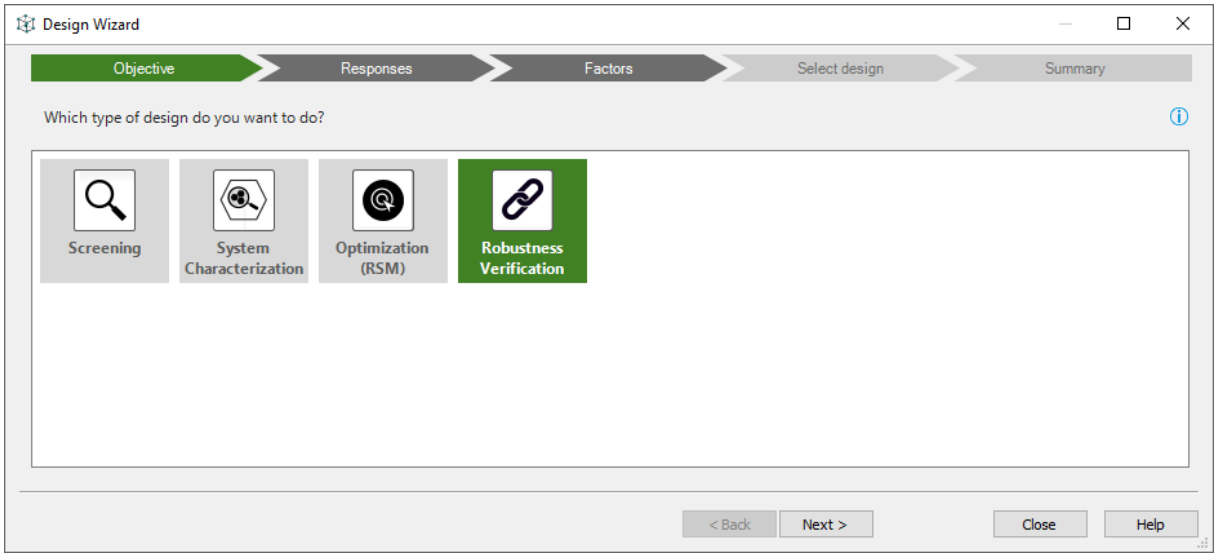
Worksheet

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Exp No | Exp Name | Run Order | Incl/Excl | AcN | pH | Temp | OSA | Column | k1 | k2 | Res1 | PlateN |
| 1 | 1 | N1 | 9 | Incl | 25 | 3.8 | 18 | 0.11 | Column B | 2.2906 | 3.3421 | 1.87 | 6310 |
| 2 | 2 | N2 | 7 | Incl | 27 | 3.8 | 18 | 0.09 | Column A | 1.7547 | 2.6802 | 1.75 | 5902 |
| 3 | 3 | N3 | 8 | Incl | 25 | 4.2 | 18 | 0.09 | Column B | 2.3933 | 3.4705 | 1.89 | 5991 |
| 4 | 4 | N4 | 10 | Incl | 27 | 4.2 | 18 | 0.11 | Column A | 1.823 | 2.8013 | 1.8 | 5783 |
| 5 | 5 | N5 | 6 | Incl | 25 | 3.8 | 25 | 0.11 | Column A | 2.1456 | 3.1599 | 1.83 | 6412 |
| 6 | 6 | N6 | 5 | Incl | 27 | 3.8 | 25 | 0.09 | Column B | 1.5031 | 2.4845 | 1.8 | 5702 |
| 7 | 7 | N7 | 1 | Incl | 25 | 4.2 | 25 | 0.09 | Column A | 2.2289 | 3.2715 | 1.86 | 5542 |
| 8 | 8 | N8 | 11 | Incl | 27 | 4.2 | 25 | 0.11 | Column B | 1.5994 | 2.6193 | 1.84 | 6136 |
| 9 | 9 | N9 | 4 | Incl | 26 | 4 | 22 | 0.1 | Column A | 2.0661 | 3.0592 | 1.81 | 6231 |
| 10 | 10 | N10 | 2 | Incl | 26 | 4 | 22 | 0.1 | Column A | 2.0253 | 3.0285 | 1.82 | 5909 |
| 11 | 11 | N11 | 3 | Incl | 26 | 4 | 22 | 0.1 | Column B | 2.0243 | 2.9903 | 1.79 | 6190 |
| 12 | 12 | N12 | 12 | Incl | 26 | 4 | 22 | 0.1 | Column B | 2.0131 | 3.0068 | 1.81 | 5992 |

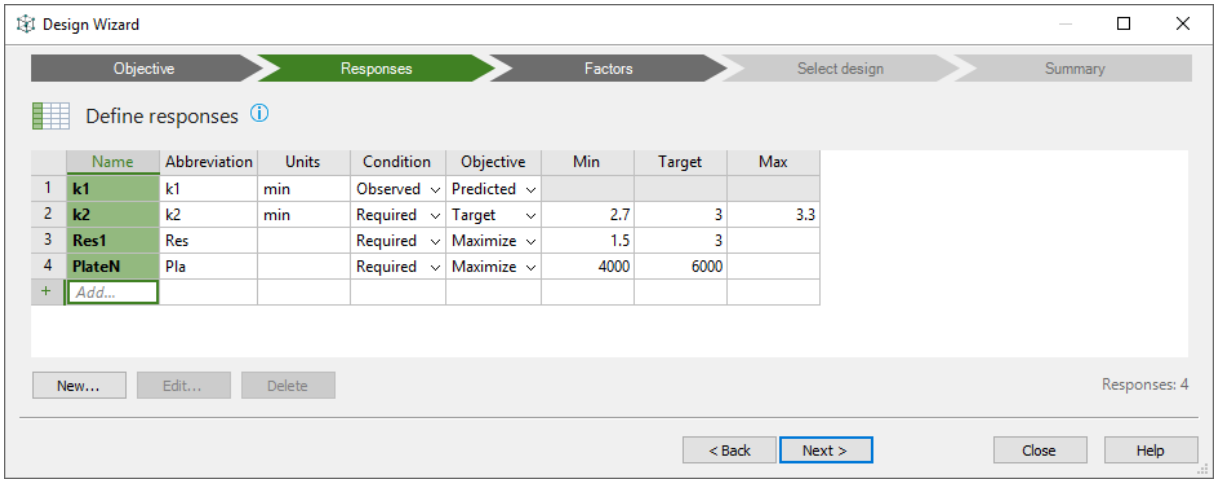
Important: Note that the center-point for the four first factors (experiments 9-12) is divided into two replicates using Column A and Column B respectively.

## Setting Up the Experimental Protocol

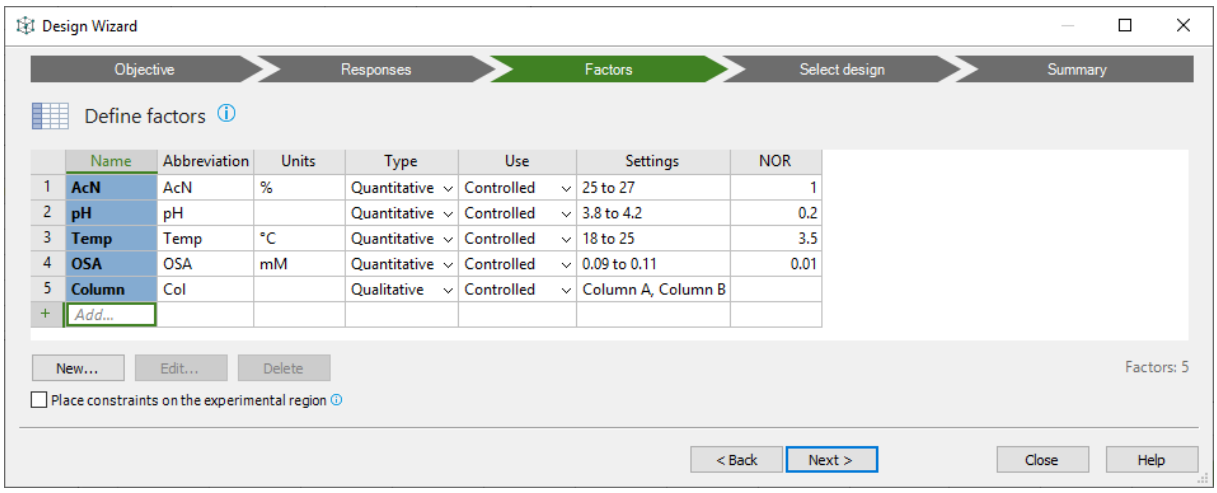
Use the Design Wizard to define a new investigation in MODDE® with five factors and four responses. Select File/New/Experimental Design/Robustness Verification and step through the Design Wizard as shown below. Click Next.



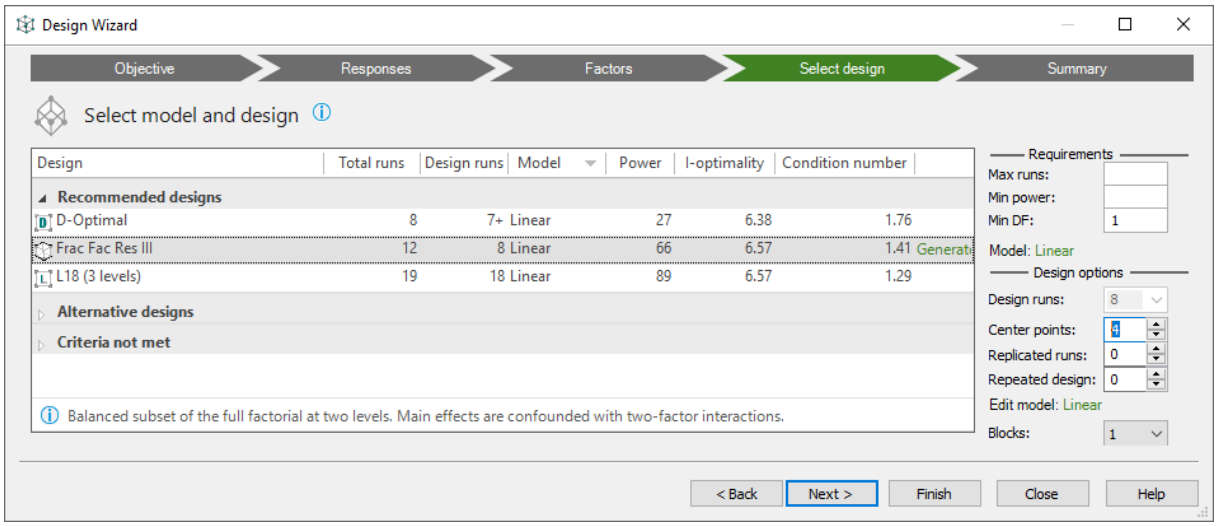

On the responses page, define the four responses according to the information given below. Click Next.



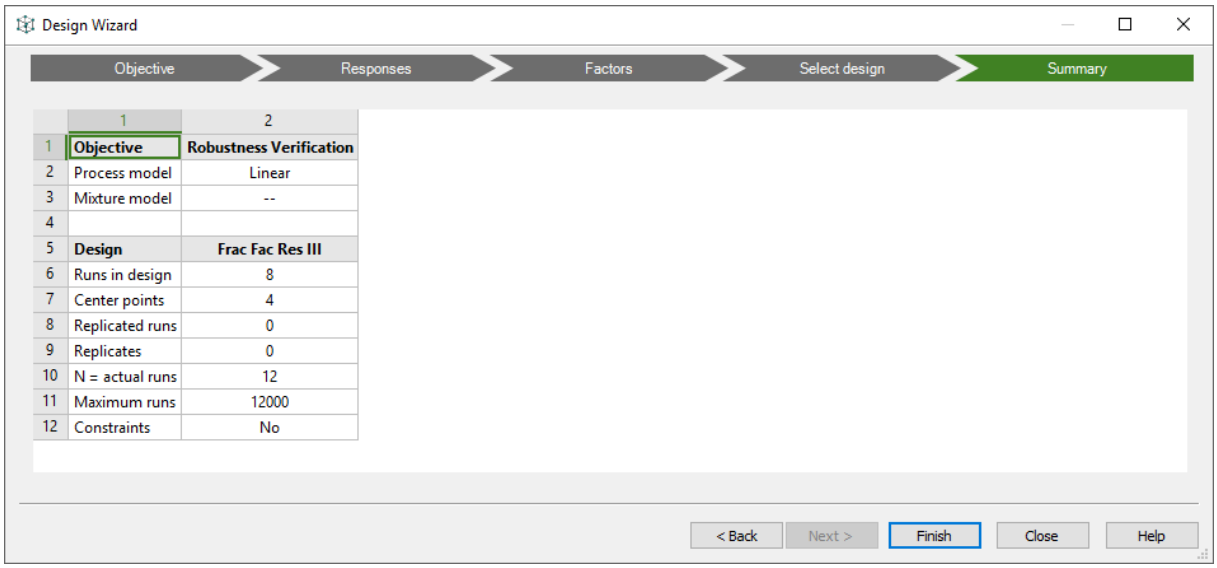

On the factors page, define the five factors according to the information given below. Observe that the last factor is a qualitative factor in two settings. Observe the NOR information. Click Next.




Select the Fractional Factorial design with eight design runs. Verify that the number of center points = 4 and Total runs = 12. Click Next.

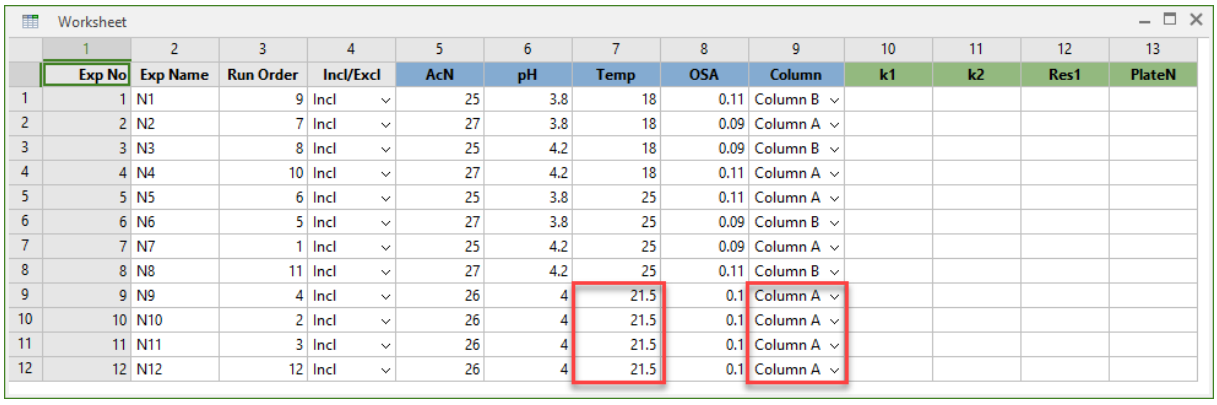


On the final Summary page you can review your selections and settings, which should look like the screenshot below. Click Finish to exit the design wizard.




| | 1 | 2 |
|---|---|---|
| 1 | **Objective** | **Robustness Verification** |
| 2 | Process model | Linear |
| 3 | Mixture model | -- |
| 4 | | |
| 5 | **Design** | **Frac Fac Res III** |
| 6 | Runs in design | 8 |
| 7 | Center points | 4 |
| 8 | Replicated runs | 0 |
| 9 | Replicates | 0 |
| 10 | N = actual runs | 12 |
| 11 | Maximum runs | 12000 |
| 12 | Constraints | No |

The resulting worksheet places all four replicates on the first setting of the qualitative factor. In reality two replicates were run for each setting of the qualitative factor. This can easily be changed in the worksheet using the drop down arrows. Also observe the small change in the centerpoint setting for the Temperature factor, 22 and not 21.5 °C. Correct this manually. As a last step you can copy the response data from the file Raw data for DOE computer exercises.XLS.




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Exp No** | **Exp Name** | **Run Order** | **Incl/Excl** | **AcN** | **pH** | **Temp** | **OSA** | **Column** | **k1** | **k2** | **Res1** | **PlateN** |
| 1 | 1 | N1 | 9 | Incl | 25 | 3.8 | 18 | 0.11 | Column B | | | | |
| 2 | 2 | N2 | 7 | Incl | 27 | 3.8 | 18 | 0.09 | Column A | | | | |
| 3 | 3 | N3 | 8 | Incl | 25 | 4.2 | 18 | 0.09 | Column B | | | | |
| 4 | 4 | N4 | 10 | Incl | 27 | 4.2 | 18 | 0.11 | Column A | | | | |
| 5 | 5 | N5 | 6 | Incl | 25 | 3.8 | 25 | 0.11 | Column A | | | | |
| 6 | 6 | N6 | 5 | Incl | 27 | 3.8 | 25 | 0.09 | Column B | | | | |
| 7 | 7 | N7 | 1 | Incl | 25 | 4.2 | 25 | 0.09 | Column A | | | | |
| 8 | 8 | N8 | 11 | Incl | 27 | 4.2 | 25 | 0.11 | Column B | | | | |
| 9 | 9 | N9 | 4 | Incl | 26 | 4 | 21.5 | 0.1 | Column A | | | | |
| 10 | 10 | N10 | 2 | Incl | 26 | 4 | 21.5 | 0.1 | Column A | | | | |
| 11 | 11 | N11 | 3 | Incl | 26 | 4 | 21.5 | 0.1 | Column A | | | | |
| 12 | 12 | N12 | 12 | Incl | 26 | 4 | 21.5 | 0.1 | Column A | | | | |

After copying and pasting the data the updated worksheet should look like the screenshot below. Now you are ready to start analyzing the data.

Worksheet

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Exp No | Exp Name | Run Order | Incl/Excl | AcN | pH | Temp | OSA | Column | k1 | k2 | Res1 | PlateN |
| 1 | 1 | N1 | 9 | Incl | 25 | 3.8 | 18 | 0.11 | Column B | 2.2906 | 3.3421 | 1.87 | 6310 |
| 2 | 2 | N2 | 7 | Incl | 27 | 3.8 | 18 | 0.09 | Column A | 1.7547 | 2.6802 | 1.75 | 5902 |
| 3 | 3 | N3 | 8 | Incl | 25 | 4.2 | 18 | 0.09 | Column B | 2.3933 | 3.4705 | 1.89 | 5991 |
| 4 | 4 | N4 | 10 | Incl | 27 | 4.2 | 18 | 0.11 | Column A | 1.823 | 2.8013 | 1.8 | 5783 |
| 5 | 5 | N5 | 6 | Incl | 25 | 3.8 | 25 | 0.11 | Column A | 2.1456 | 3.1599 | 1.83 | 6412 |
| 6 | 6 | N6 | 5 | Incl | 27 | 3.8 | 25 | 0.09 | Column B | 1.5031 | 2.4845 | 1.8 | 5702 |
| 7 | 7 | N7 | 1 | Incl | 25 | 4.2 | 25 | 0.09 | Column A | 2.2289 | 3.2715 | 1.86 | 5542 |
| 8 | 8 | N8 | 11 | Incl | 27 | 4.2 | 25 | 0.11 | Column B | 1.5994 | 2.6193 | 1.84 | 6136 |
| 9 | 9 | N9 | 4 | Incl | 26 | 4 | 22 | 0.1 | Column A | 2.0661 | 3.0592 | 1.81 | 6231 |
| 10 | 10 | N10 | 2 | Incl | 26 | 4 | 22 | 0.1 | Column A | 2.0253 | 3.0285 | 1.82 | 5909 |
| 11 | 11 | N11 | 3 | Incl | 26 | 4 | 22 | 0.1 | Column B | 2.0243 | 2.9903 | 1.79 | 6190 |
| 12 | 12 | N12 | 12 | Incl | 26 | 4 | 22 | 0.1 | Column B | 2.0131 | 3.0068 | 1.81 | 5992 |

## Analyzing the Data

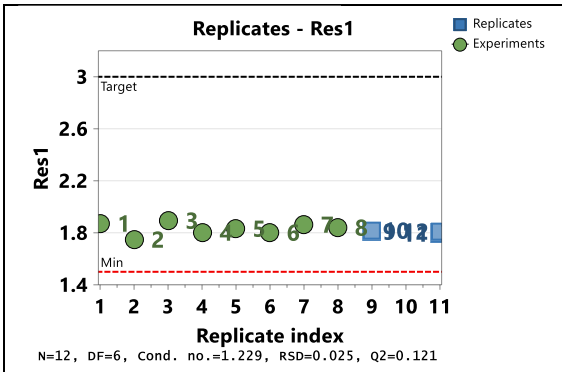
Use the Analysis wizard to analyse the data. Omit the first response in the evaluation as it has no critical specification (no Min or Max values in the response specification).

Use the 4 limiting cases discussed on page 1 to explore the robustness of responses Res1 and PlateN with respect to;

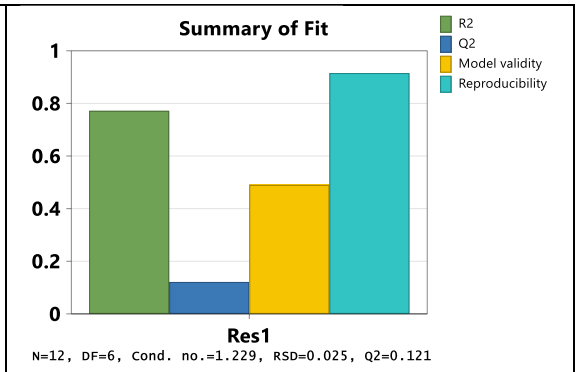
- Robustness – how can we determine whether the model is robust?
- Significance – what signifies a significant model?
- Specifications – are the response values within Min/Max specifications?
- Variability – how much can we expect the response to vary at most? Normally?

First Case – Inside Specification and Significant ModelThe first case is *inside the specification and a significant model*. The HPLC application contains one example of this case, the Res1 response. We assume, on the basis of the initial raw data assessment, that this response is robust, because all the measured values are inside the specification, that is, above 1.5. Actually, as seen in the replicate plot, the measured values are all above 1.75.

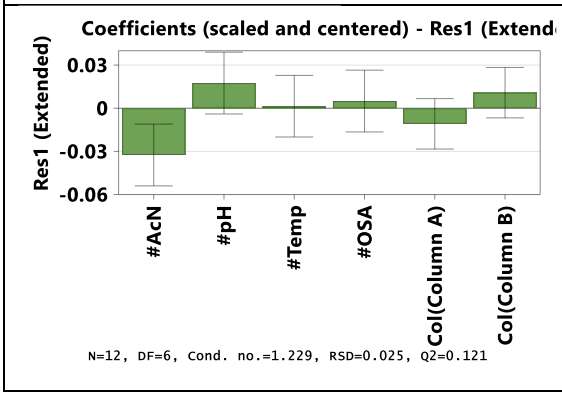
The question of whether the model is significant, however, is more debatable. It is possible to interpret the regression model as representing a weakly significant regression equation. We will do so in this section for the sake of illustration. The classification of the model as significant is based on a joint assessment of the low, but positive, $Q^2$, seen in the summary plot, and the significant linear term for acetonitrile, seen in the coefficient plot. Hence, Res1 may be regarded as an illustration of the first case.

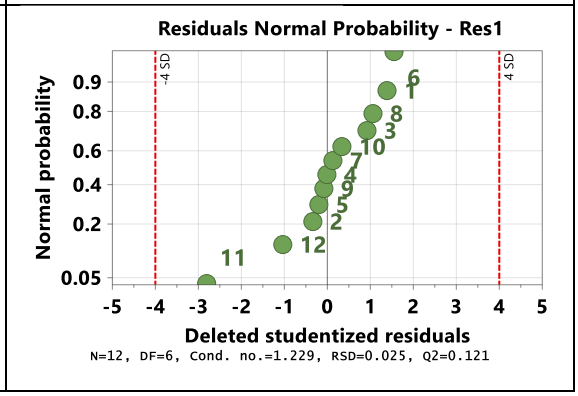



Replication plot for Res 1, all experiments are above the specification limit.




Summary of fit plot, Q2 indicates a weak but significant model. No lack of fit.



| | |
|---|---|
| Regression coefficients of the model for Res1; there is one significant factor, acetonitrile. | The residuals are normally distributed. No outliers. |

The question is what variability in Res1 can be expected when all factors are allowed to vary in the region that has been investigated. An understanding of such variability can accomplished in two ways using model predictions; (1) use the model and predict the worst case or (2) use the model and simulate what the result will be with normally distributed random disturbances on all the factors in the region that has been investigated. The first approach assumes that we are trying to run the HPLC at its setpoint, but at the same time we recognize that all factors may occasionally be at their extreme values (which is highly unlikely with trained personnel). The second approach, where it is more likely that we run close to the setpoint, rather than at the extreme values, is the more realistic one.

(1) By using the tool *Predict/Spreadsheet* the worst case scenario for Res1 can be predicted by using the factor combination low AcN, high pH, high Temp, high OSA, and ColB, and the other extreme experiment by high AcN, low pH, low Temp, low OSA, and ColA. The prediction list will give predictions including the 95% confidence interval showing that Res1 will be in the range from 1.7 to 1.94. Well over the limit of 1.5.

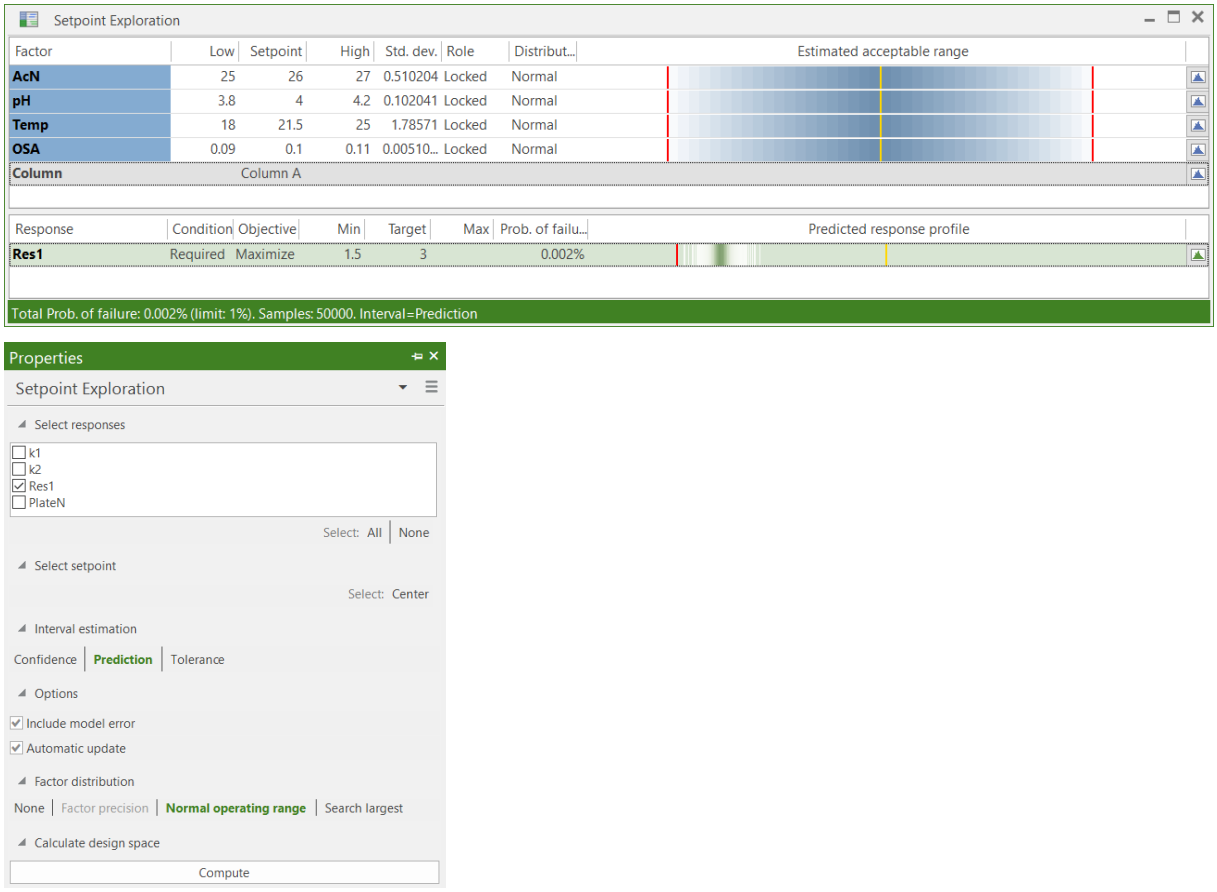
Prediction Spreadsheet

| | 1 | 2 | 3 | 4 | 5 | 6 7 8 9 1 1 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|
| | **AcN** | **pH** | **Temp** | **OSA** | **Column** | **lr l lr l** | **Res1** | **Lower** | **Upper** |
| 1 | 25 | 4.2 | 25 | 0.11 | Column B | 3 3 3 6 9 | 1.88968 | 1.84053 | 1.93884 |
| 2 | 27 | 3.8 | 18 | 0.09 | Column A | 2 3 9 9 8 | 1.75518 | 1.70515 | 1.80521 |

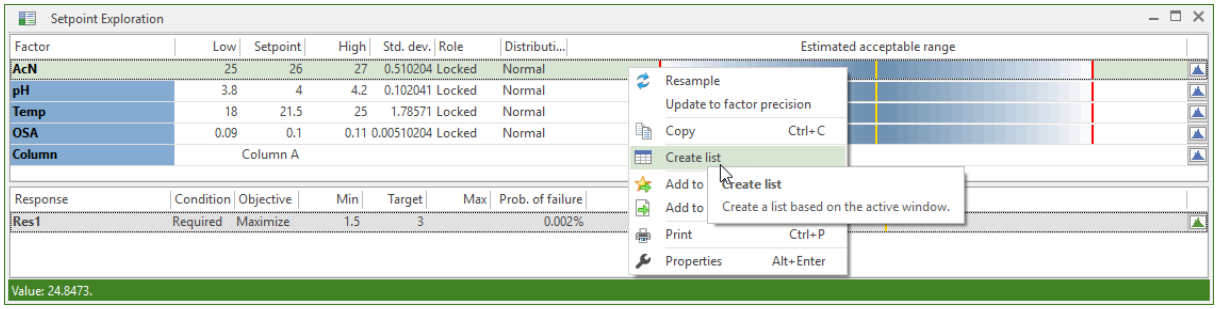
(2) A more realistic approach is to use the model and simulate random disturbances within the range of operation for all factors. By using the tool Setpoint exploration (Predict tab) the robustness is tested with a large number of random disturbances (Monte Carlo simulations) in a user-specified region, in this case the Normal Operating Region (NOR). In the screenshot shown below the factor part shows the extent of the NOR and with a specific selection for the qualitative factor Column, ColA. ColA is chosen as the worst case indicated by the model (Column A is predicted to correspond to the lowest Res1).

The result is shown as a distribution of random samples including model prediction errors. The result is well within the specification limits. The result can be obtained as general statistics, like average and standard deviation, as well as capability index (Cp or CpK) or DPMO or probability of failure.

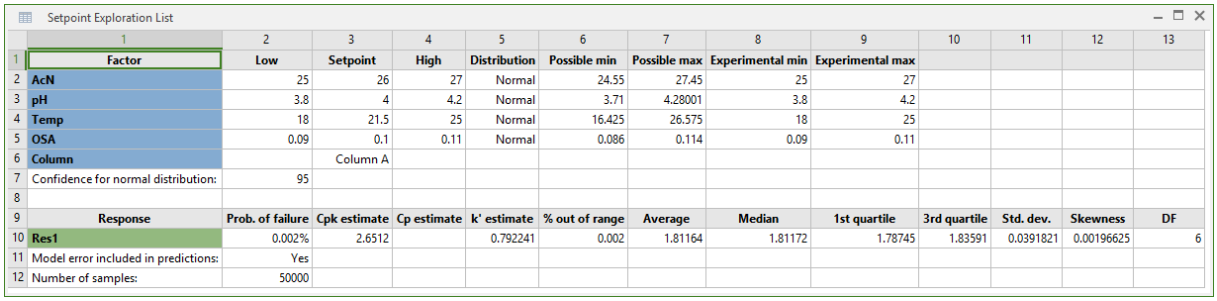
Note: In order to mimic the results shown below you have to open the Properties pane and make sure that the Prediction interval is used and the factor distribution is set to Normal operating range.




Right-click and select Create List.

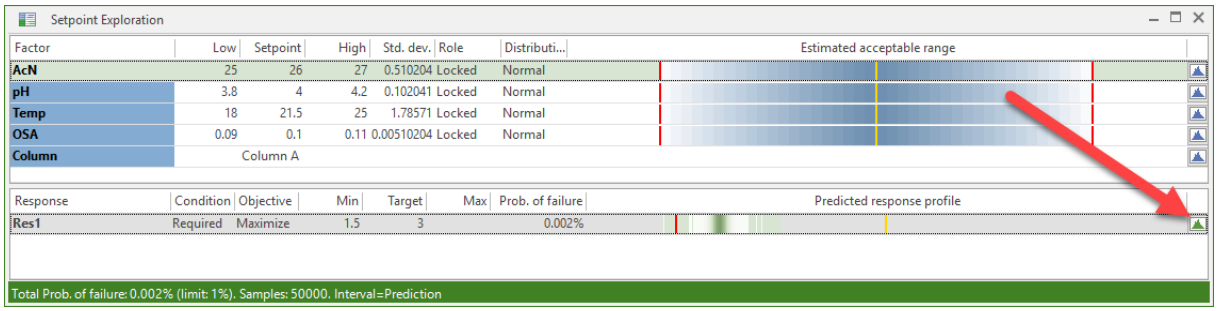


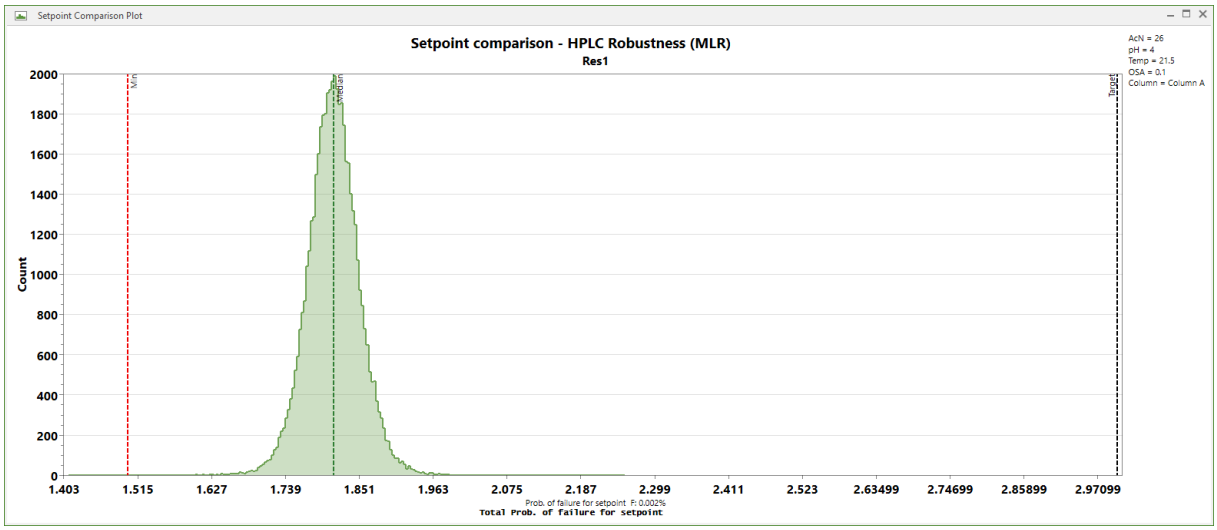

This action opens up the Setpoint Exploration List, which contains more information and summary statistics. The overall result is a distribution of Res1, well within the specifications, where probability of failure is close to 0 (0.002%).



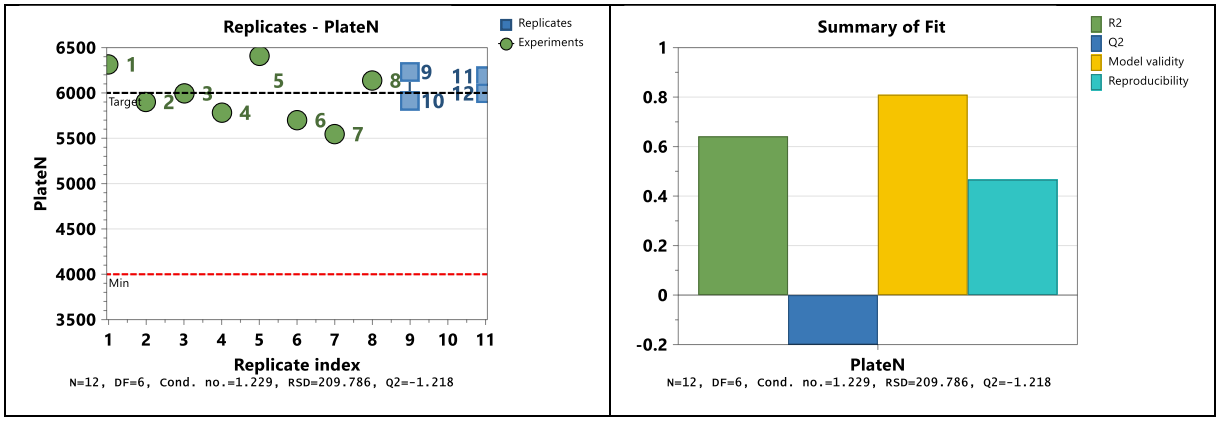

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Factor | Low | Setpoint | High | Distribution | Possible min | Possible max | Experimental min | Experimental max | | | | |
| 2 | AcN | 25 | 26 | 27 | Normal | 24.55 | 27.45 | 25 | 27 | | | | |
| 3 | pH | 3.8 | 4 | 4.2 | Normal | 3.71 | 4.28001 | 3.8 | 4.2 | | | | |
| 4 | Temp | 18 | 21.5 | 25 | Normal | 16.425 | 26.575 | 18 | 25 | | | | |
| 5 | OSA | 0.09 | 0.1 | 0.11 | Normal | 0.086 | 0.114 | 0.09 | 0.11 | | | | |
| 6 | Column | | Column A | | | | | | | | | | |
| 7 | Confidence for normal distribution: | 95 | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | |
| 9 | Response | Prob. of failure | Cpk estimate | Cp estimate | k' estimate | % out of range | Average | Median | 1st quartile | 3rd quartile | Std. dev. | Skewness | DF |
| 10 | Res1 | 0.002% | 2.6512 | | 0.792241 | 0.002 | 1.81164 | 1.81172 | 1.78745 | 1.83591 | 0.0391821 | 0.00196625 | 6 |
| 11 | Model error included in predictions: | Yes | | | | | | | | | | | |
| 12 | Number of samples: | 50000 | | | | | | | | | | | |

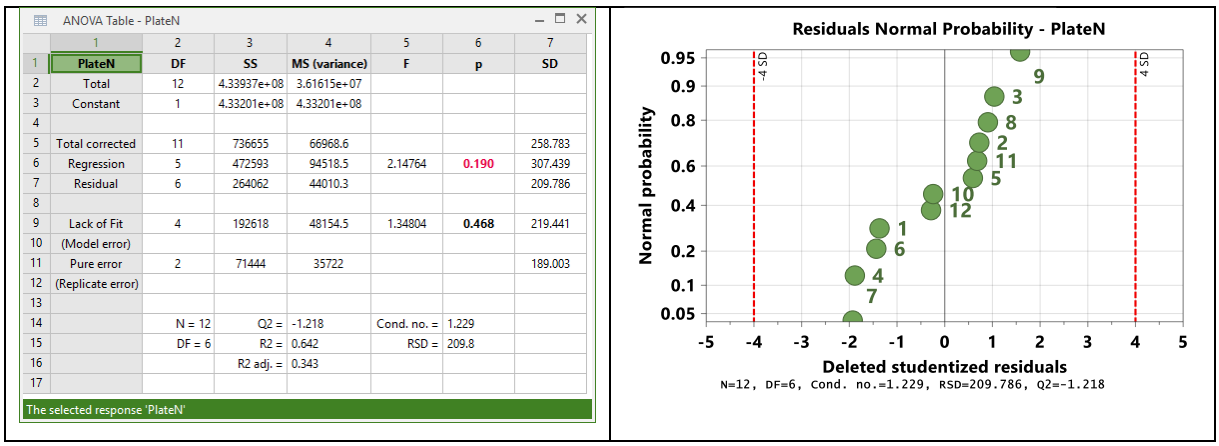
The distribution of the predicted values of Res1 can be visualized by clicking on the button called Open setpoint comparison plot.



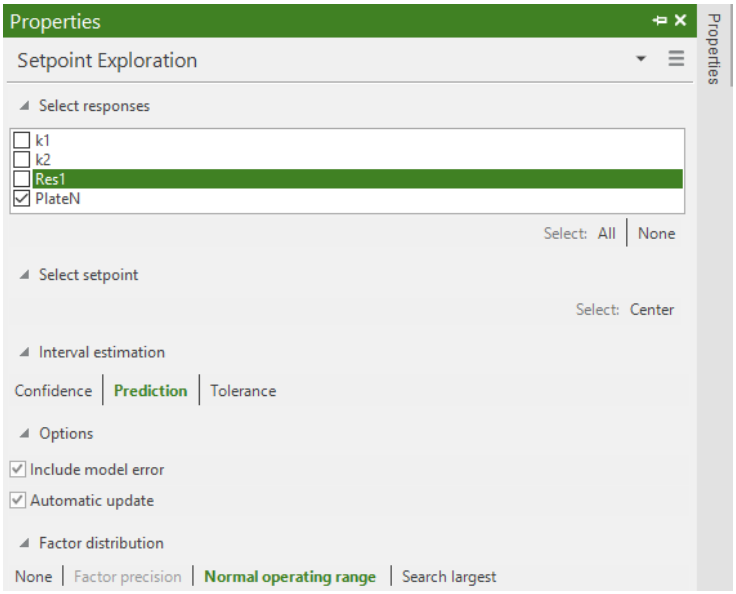



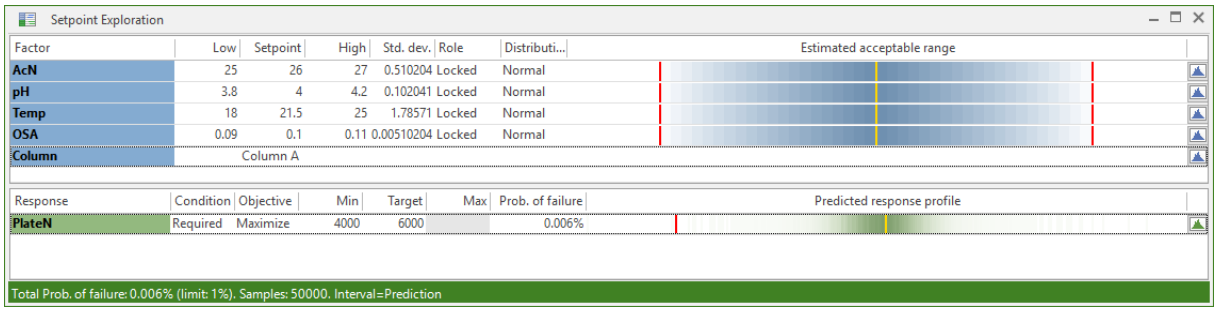

Second Case – Inside Specification and Non-Significant ModelThe second case is *inside the specification and a non-significant model*. This is the ideal outcome of a robustness test. We use the PlateN response as an illustration. We know that the measured values of this response are all inside the specification, and the regression model obtained was non-significant. In general, to assess model significance, two diagnostic tools are more appropriate than any others. The first tool consists of the $R^2$ and $Q^2$ parameters. The second important modeling tool is the analysis of variance (ANOVA), and particularly the upper F-test, which is a test of the significance of the regression model. We can see in the ANOVA Table that the PlateN model is not significant, because the p-value of 0.19 exceeds 0.05.






The *Setpoint Exploration* confirms directly that plate number is above the specification limit for all factor settings within the investigated region. Probability of failure = 0,006% shows that the HPLC method is robust.

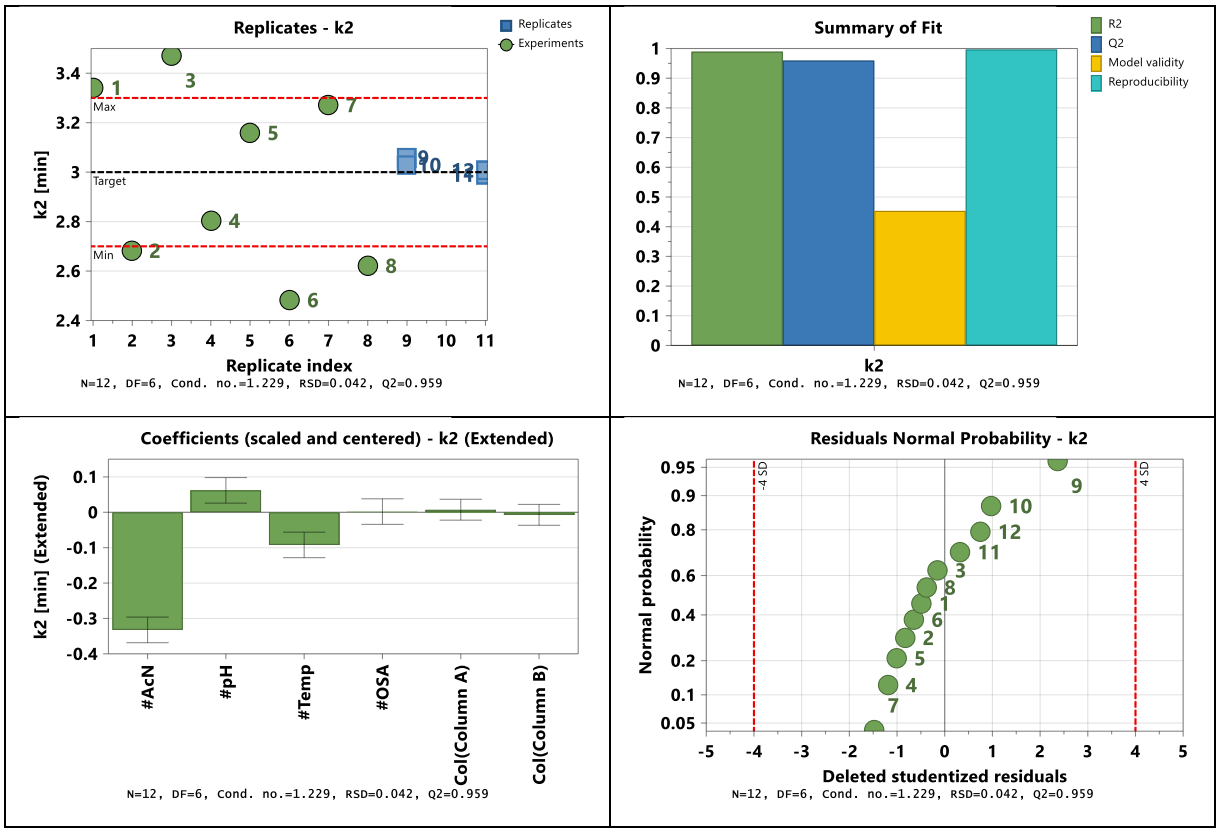



## Third Case – Outside Specification and Significant Model

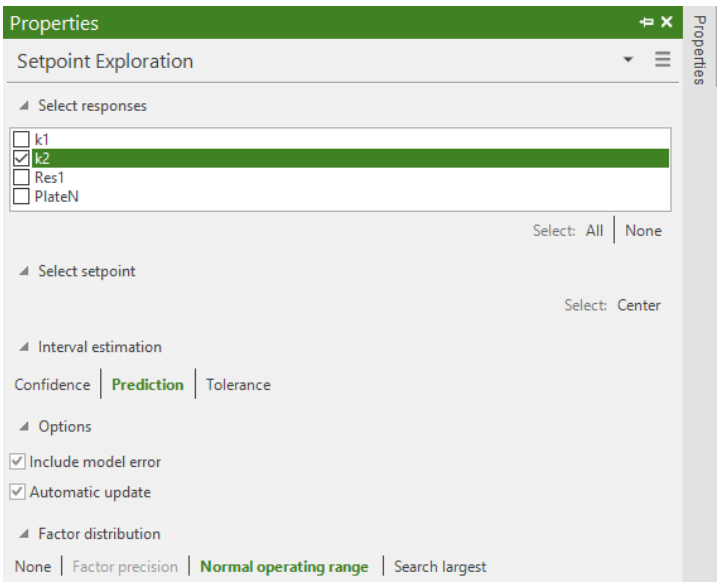
The third case is *outside the specification and a significant model*. This case occurs whenever a significant regression model is calculated but the raw response data themselves do not meet the goals of the problem formulation. We will use the second response, $k_2$, of the HPLC data to illustrate this case. Specification for a capacity factor is uncommon in the pharmaceutical industry, but one is available here for the sake of illustration.

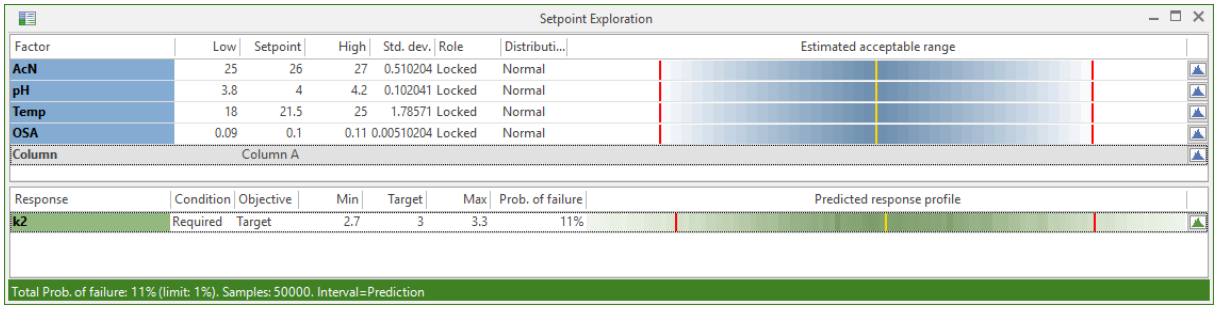
We start by assessing the statistical behavior of the $k_2$ regression model. This behavior is evident from the summary of fit plot below, which indicates the sensitivity of $k_2$ to small factor changes. In order to understand what is causing this susceptibility to changes in the factors, it is necessary to examine the regression coefficients.




We can see that it is mainly acetonitrile, pH and temperature which affect $k_2$. Now we can use the model and its prediction of $k_2$ and do modifications on how much the region of accepted variability in each factor has to be reduced in order to get inside the specification limits for $k_2$. By using the *Predict/Setpoint Exploration* tool the estimate of a valid design space can be done automatically or user controlled. If we start by letting all variables vary within the experimental region, from low to high setting, we get an estimate of how $k_2$ will vary.

By using the regression model and Monte Carlo simulations with normal distribution of random factor settings in the specified region, we get an estimate of $k_2$ predictions for a real situation. The following picture is a visualization of variability range for the factors and the predicted $k_2$ distribution. The distribution exceeds the limits with 11% of the predictions outside specification.

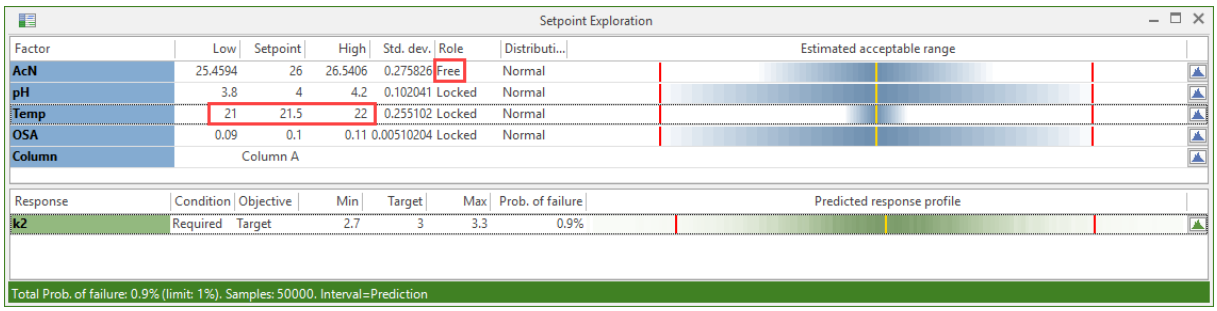



How to get inside specifications for $k_2$?

As the distribution is well centered in the acceptance region a good way to reduce the predicted variability for $k_2$ is to reduce the factor variability symmetrically. Since the prediction is in the center of the acceptance region, the reduction can be done around the average factor setting. There are two options; the first is to reduce the range for factors that are easy to control. The obvious candidate in this case is temperature. With standard equipment for temperature stabilization the range can be reduced to +/- 0.5 degrees. The second option is to reduce the most significant factor (AcN) that will have the largest effect on the reduction of variability of $k_2$. The reduction of AcN can be done automatically by a search function or manually. In this case a combination of these options will be used.

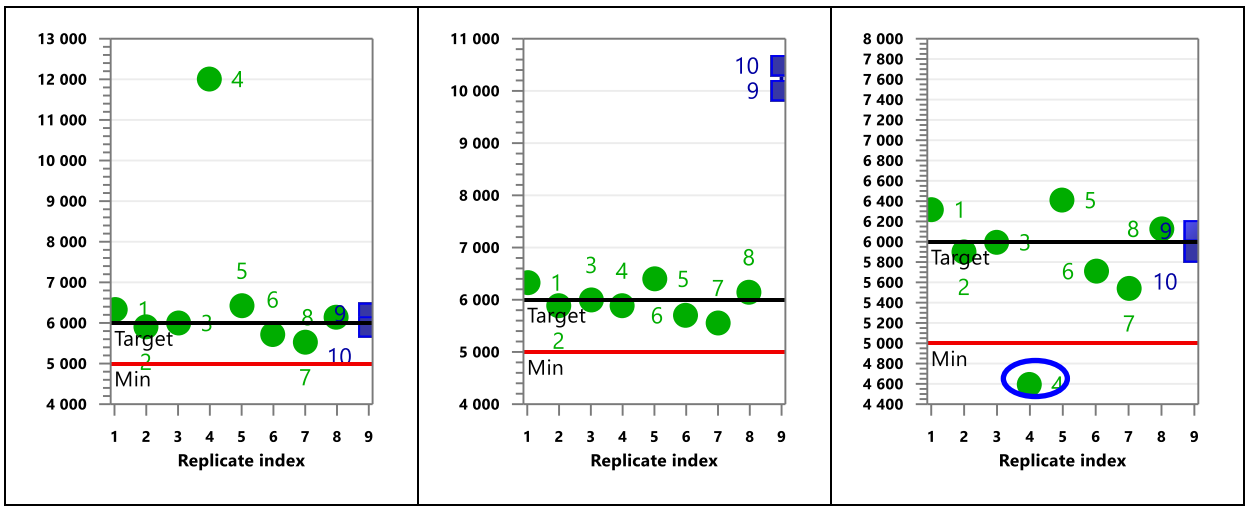
We alter the factor specifications by imposing a fixed reduction on Temperature (from 21.0 to 22.0 degrees) and setting AcN to free. Now MODDE® will search for the largest range for AcN where the distribution for $k_2$ will be predicted according to the specifications, given that all other factor settings are set according to the specifications. The result shown in the following picture is that AcN can vary from 25,5 – 26,5 with the other factor specifications adhered to and the prediction of $k_2$ will then be inside the given limits.




We note that the given limits on AcN are time-consuming to achieve. Therefore, it should be considered if it is possible to relax the upper limit on $k_2$ a little and thereby increase the time for each analytical run but saving overall time. For example, with an upper $k_2$ limit of 4 the limits for AcN would be widened to become 25 to 26.3. Furthermore, should it be possible to decrease the lower $k_2$ limit even larger ranges for AcN would be acceptable.

Fourth Case – Outside Specifications and Non-Significant ModelThe fourth case is *outside the specification with a non-significant model*. This case may be the result when the derived regression model is poor, and there are anomalies in the data. Such anomalies are important to uncover, because their presence will influence the modeling. An informative graphical tool for identifying whether this case is relevant is the replicate plot.

The left-hand figure below shows an example in which one strong outlier is present; this would exclude all possibilities of robustness. The second figure depicts another case where all the replicated center-points have much higher response values than the other runs. This pattern hints at curvature and implies a lack of robustness. A third common situation, which partly resembles the first case, is when one experiment deviates from the rest and also falls outside some predefined robustness limits. This is shown in the last figure.




Evidently, there can be several underlying explanations to this case, and we have just shown a few. Therefore, we consider this case as the most complex one. In summary, we have described four cases of robustness testing, and it is important to realize that robustness testing results are not statically locked to these four outcomes. In principle, there is a gradual transition from one case to another, and hence an infinite number of outcomes are conceivable.

## Conclusions

The application of DOE in the robustness testing of the HPLC system was very successful. With this approach it was possible to infer the robustness of the Res1 response. From a tutorial point of view, the HPLC application is good for several reasons. It represents a realistic case in which all the necessary steps for verifying the robustness of an analytical system are illustrated. Furthermore, this dataset also allowed us to study how tools for evaluation of raw data, tools for regression analysis, and tools for model use, were put into practice in a real situation. It should be clear from this application, that the modeling step is of crucial importance in robustness testing, as it is linked to an understanding of the nature of the robustness or non-robustness. We also used the HPLC study for discussing four cases of robustness testing.