SARTORIUS

Simplifying Progress

# Design of Experiments (DOE) for the Beginner

Lennart Eriksson, Ph.D., Assoc. Prof.
Senior Lecturer and Principal Data Scientist

# Born in Data Analytics

- Company founded in 1987 by Professor Svante Wold, in Umeå, Sweden
  - Originator of Chemometrics and the SIMCA® Methodology
- Patented technologies in Design of Experiments and Multivariate Data Analysis
- We help our customers bring high-quality products to market faster
- Part of Sartorius Stedim Biotech since April 2017
- Products like MODDE®, SIMCA® and SIMCA®-online
- Global strength with local presence



SARTORIUS

# Business Growth Through the Entire Product Lifecycle

**Active Dashboard**
Interactive performance insight

**Control Advisor**
Avoid problems before they arise

**MODDE®**
Get it right from the start

**SIMCA®**
Turn data into growth

**SIMCA®-online**
Ensured manufacturing success

## Umetrics® Suite

**Education/training and consulting**
Transferring expertise to you

# Contents

# Why DOE Is Used and Common Applications

- DOE is used to
  - gain knowledge,
  - increase understanding,
  - estimate proper operating conditions (“design space estimation”) of a system/process/product
- DOE applies to problem areas such as:
  - Development of new products and processes
  - Enhancement of existing products and processes
  - Optimization of quality and performance of a product
  - Optimization of an existing manufacturing procedure
  - Screening of important factors
  - Minimization of production cost
  - Robustness testing of products and processes

# A Small Example – the COST Approach

- System
  - A chemical reaction

- Goal
  - Find conditions for optimal yield

- Factors affecting the system
  - Volume between 500 and 700 ml
  - pH between 2,5 and 5

- Response
  - Yield of desired product



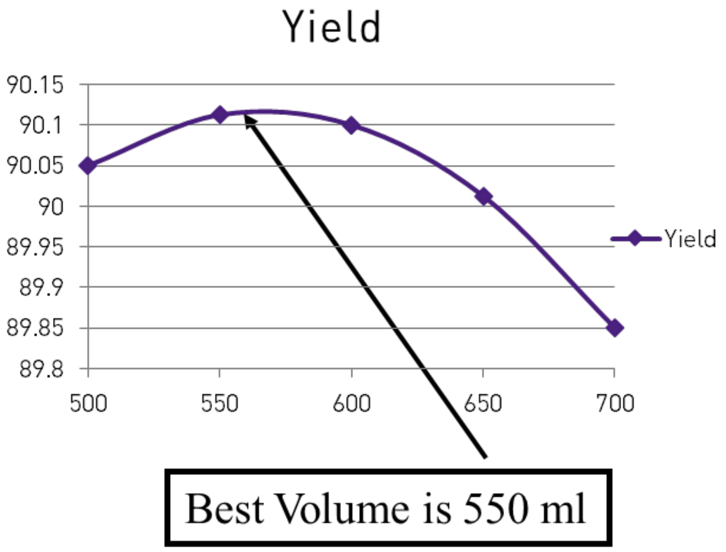
SARTORIUS

# COST Approach – Vary the First Factor

- Investigate Volume
  - Keep pH constant at 3
  - Vary Volume between 500 and 700 ml
  - Measure Yield

| Volume | Yield |
|---|---|
| 500 | 90.05 |
| 550 | 90.11 |
| 600 | 90.10 |
| 650 | 90.01 |
| 700 | 89.85 |

Ref: Box, Hunter, Hunter; *Statistics for experimenters*



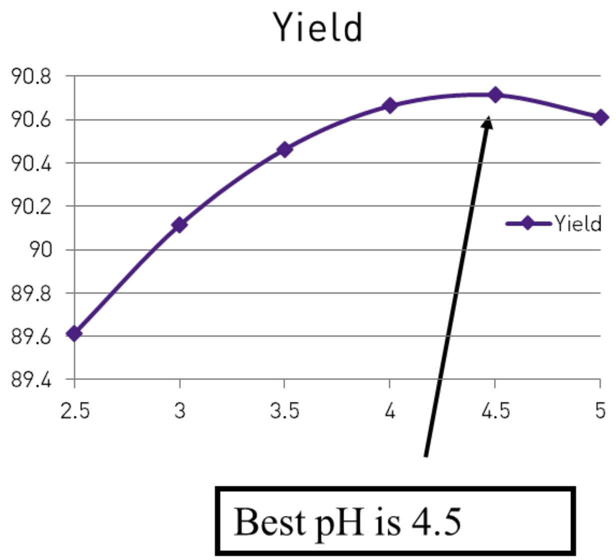



SARTORIUS

# COST Approach – Vary the Second Factor

- Investigate pH
  - Keep Volume at 550
  - Vary pH between 2,5 and 5
  - Measure Yield

| pH | Yield |
|---|---|
| 2.5 | 89.61 |
| 3.0 | 90.11 |
| 3.5 | 90.46 |
| 4.0 | 90.66 |
| 4.5 | 90.71 |
| 5.0 | 90.61 |

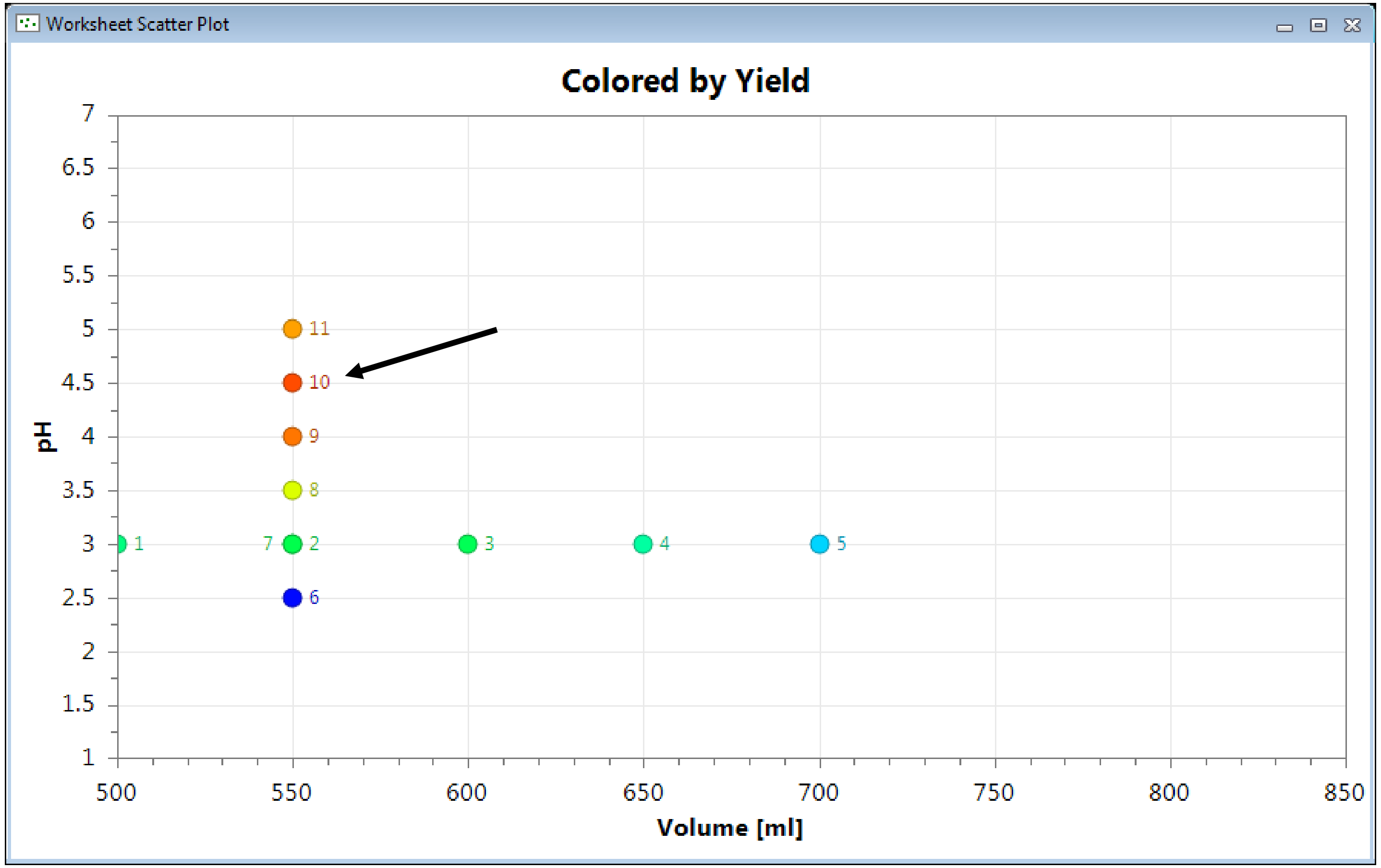


# COST Approach – The Experiments

Is this really the optimal point?

Are there other directions, giving higher yield?

Optimal number of runs?

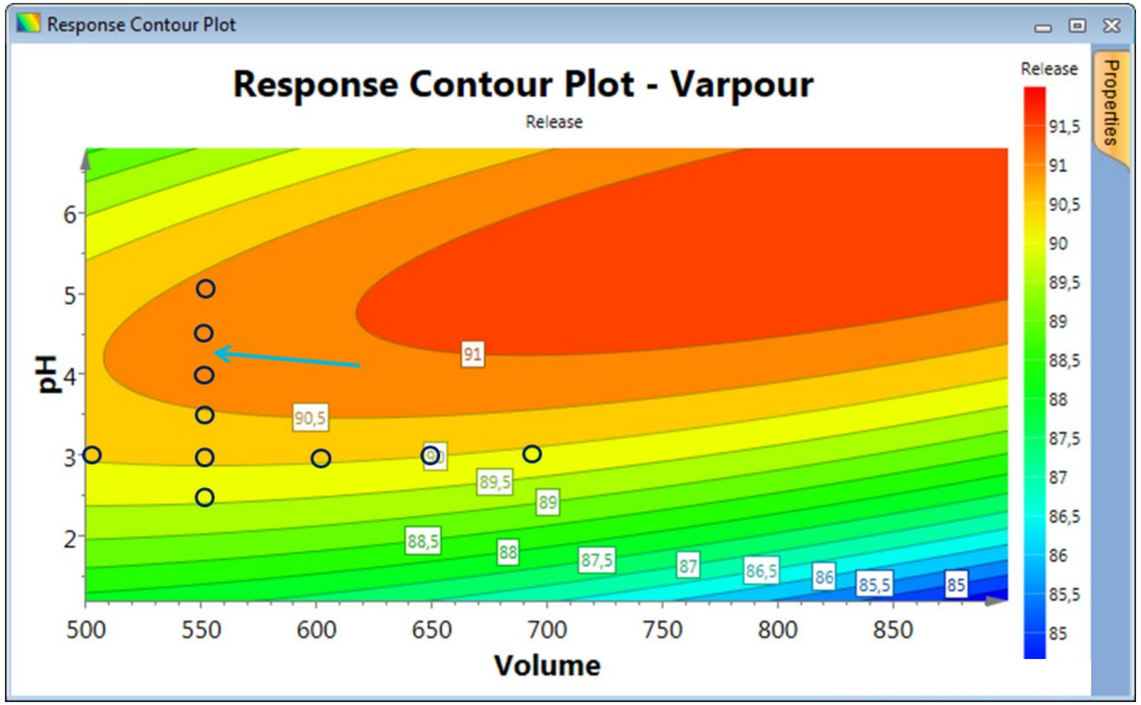





SARTORIUS

# COST Approach – In the "Real" Map

- Is this set, or group, of runs suitable to find the maximum?
- What happens with more factors than two? How many runs?
- What happens when a different starting point is chosen?

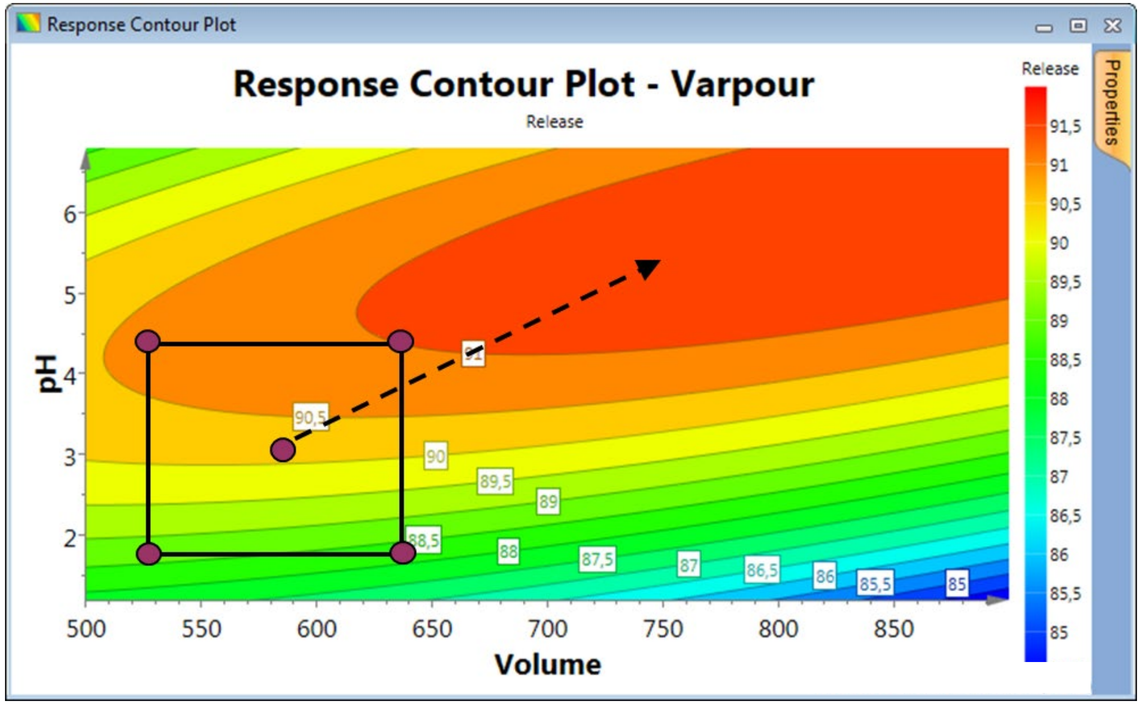





SARTORIUS

# DOE Approach – How to Build the Map

- DOE suggests the correct (often fewer than COST) number of runs needed
- DOE results in a model, a direction to follow
- Many factors can be used

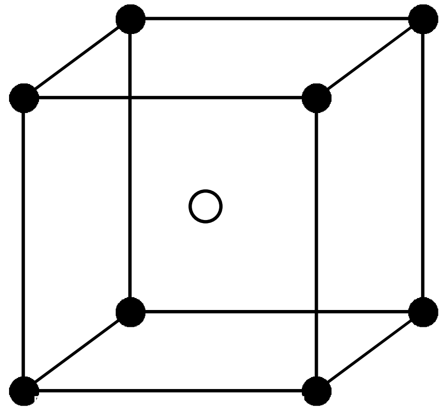


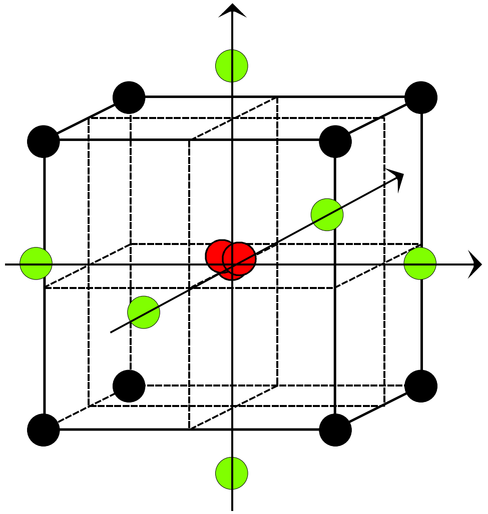
# A Better Approach - DOE

- If not cost, what do we do instead?

- The solution is to construct a carefully prepared set of representative experiments, in which all relevant factors are varied simultaneously

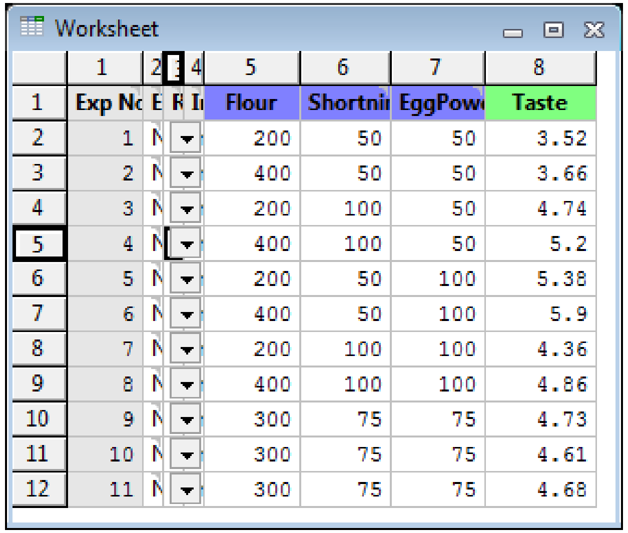
- DOE is about creating an entity of experiments that work together to explore the interesting region





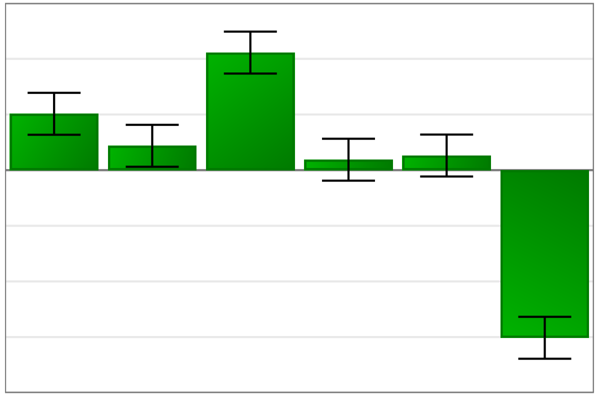
# The Design Encodes a Model to Interpret

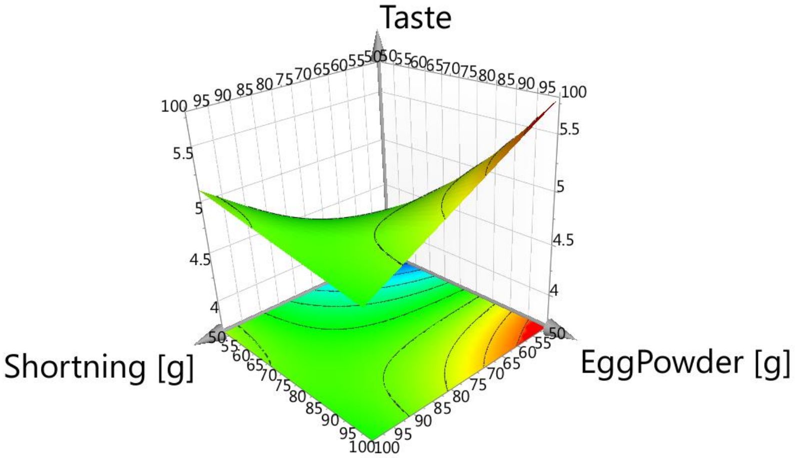
- $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$



Worksheet

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | Exp Nc | E | R | I | Flour | Shortnii | EggPow | Taste |
| 2 | 1 | N | | | 200 | 50 | 50 | 3.52 |
| 3 | 2 | N | | | 400 | 50 | 50 | 3.66 |
| 4 | 3 | N | | | 200 | 100 | 50 | 4.74 |
| 5 | 4 | N | | | 400 | 100 | 50 | 5.2 |
| 6 | 5 | N | | | 200 | 50 | 100 | 5.38 |
| 7 | 6 | N | | | 400 | 50 | 100 | 5.9 |
| 8 | 7 | N | | | 200 | 100 | 100 | 4.36 |
| 9 | 8 | N | | | 400 | 100 | 100 | 4.86 |
| 10 | 9 | N | | | 300 | 75 | 75 | 4.73 |
| 11 | 10 | N | | | 300 | 75 | 75 | 4.61 |
| 12 | 11 | N | | | 300 | 75 | 75 | 4.68 |








SARTORIUS

# Benefits of DOE

- Organized approach which connects experiments in a rational manner
- The influence of and interactions between all factors are estimated
- More precise information is acquired in fewer experiments
- Results are evaluated in the light of variability
- Support for decision-making: Map of the system (response contour plot)


# Mission Popcorn

## Making DOE Understandable to Kids

SARTORIUS

# Making DOE Understandable to Kids

- How do you explain DOE to your kids?
- Mission Popcorn; carried out during summer break of 2006
- Root cause (at Legoland, Billund) was well tasting cotton candy but distasteful popcorn (burnt, unpleasant odor)



SARTORIUS

# Selection of Objective

- Practical objective: To explain to kids what DOE means using an everyday problem (i.e., how to get good popcorn from the microwave) as illustration.

- Experimental objective: Optimization (RSM in software)

# Specification of Responses

- The dataset contains two responses, Kernels and Taste.
  - Kernels, this is simply the number of unpopped kernels.
  - Taste, each person expressed his liking on a five-level scale (1=bad taste, …., 5=optimal taste). The response value is the sum across three persons (we could not use the average as this was too complicated for the little brother).

Responses

| | Name | Abbr. | Units | Transform | Type | Min | Target | Max |
|---|---|---|---|---|---|---|---|---|
| 1 | **Kernels** | Ker | | None | Regular | | 0 | 60 |
| 2 | **Taste** | Tas | | None | Regular | 9 | 15 | |
| | | *Double-click here* | *to add* | *a new response* | | | | |

# Definition of Factors

- The dataset contains two factors, Time and Power, both adjustable on a continuous scale:
  - Time (seconds), low level 170 seconds, high level 210 seconds.
  - Power (watt), low level 600 watts, high level 800 watts.

Factors

| | Name | Abbr. | Units | Type | Settings | Transform | Precision |
|---|---|---|---|---|---|---|---|
| 1 | **Time** | Ti | seconds | Quantitative | 170 to 210 | None | 1 |
| 2 | **Power** | Po | W | Quantitative | 600 to 800 | None | 5 |
| | | *Double-click here* | *to add* | *a new factor* | | | |

# Generation of Experimental Design

- The design used was a CCF optimization design, by default encoding 8+3 experiments in MODDE 12. One centerpoint was dropped since we bought a ten-pack of microwave popcorn.

Worksheet

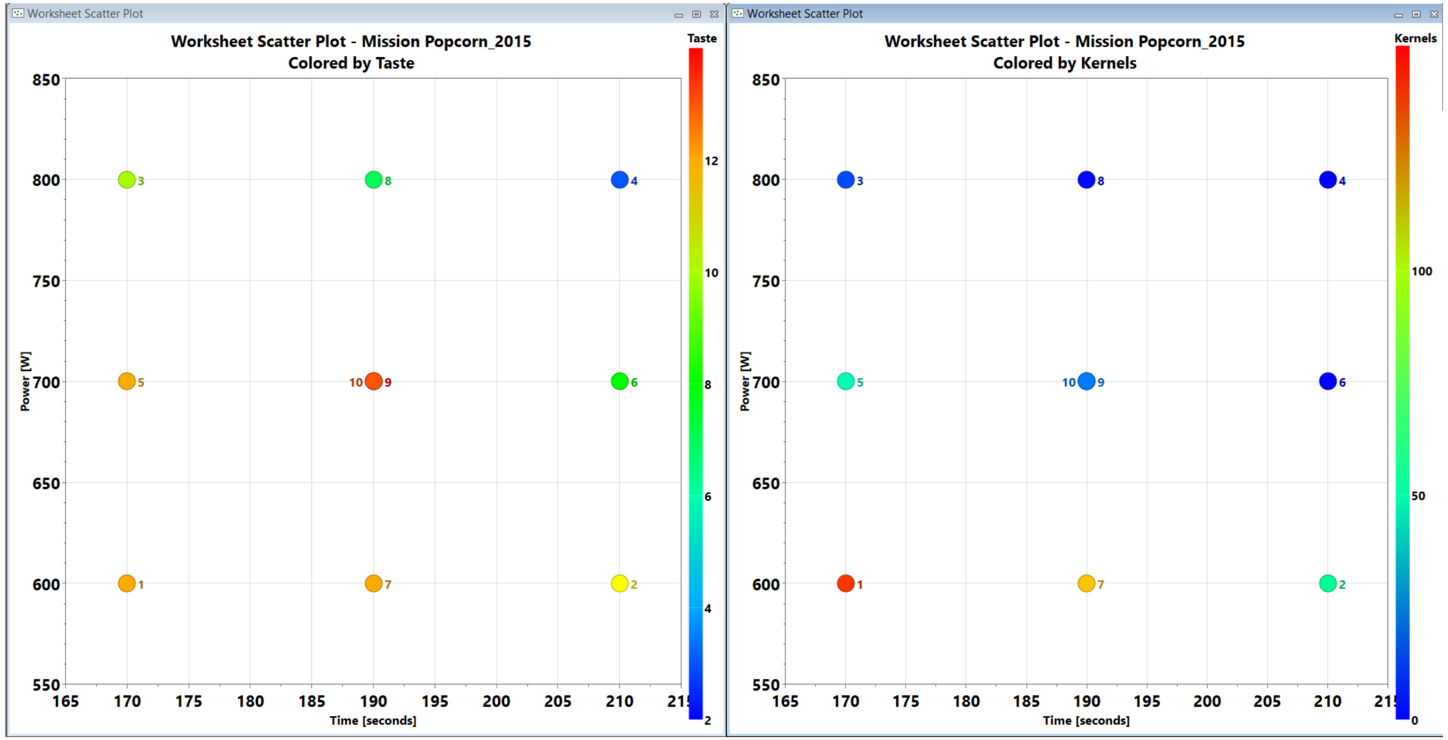
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | Exp No | Exp Name | Run Order | Incl/Excl | Time | Power | Kernels | Taste |
| 2 | 1 | N1 | 3 | Incl | 170 | 600 | 142 | 12 |
| 3 | 2 | N2 | 10 | Incl | 210 | 600 | 53 | 11 |
| 4 | 3 | N3 | 6 | Incl | 170 | 800 | 11 | 10 |
| 5 | 4 | N4 | 4 | Incl | 210 | 800 | 0 | 3 |
| 6 | 5 | N5 | 8 | Incl | 170 | 700 | 48 | 12 |
| 7 | 6 | N6 | 1 | Incl | 210 | 700 | 0 | 8 |
| 8 | 7 | N7 | 7 | Incl | 190 | 600 | 121 | 12 |
| 9 | 8 | N8 | 9 | Incl | 190 | 800 | 1 | 7 |
| 10 | 9 | N9 | 5 | Incl | 190 | 700 | 22 | 14 |
| 11 | 10 | N10 | 2 | Incl | 190 | 700 | 18 | 13 |



SARTORIUS

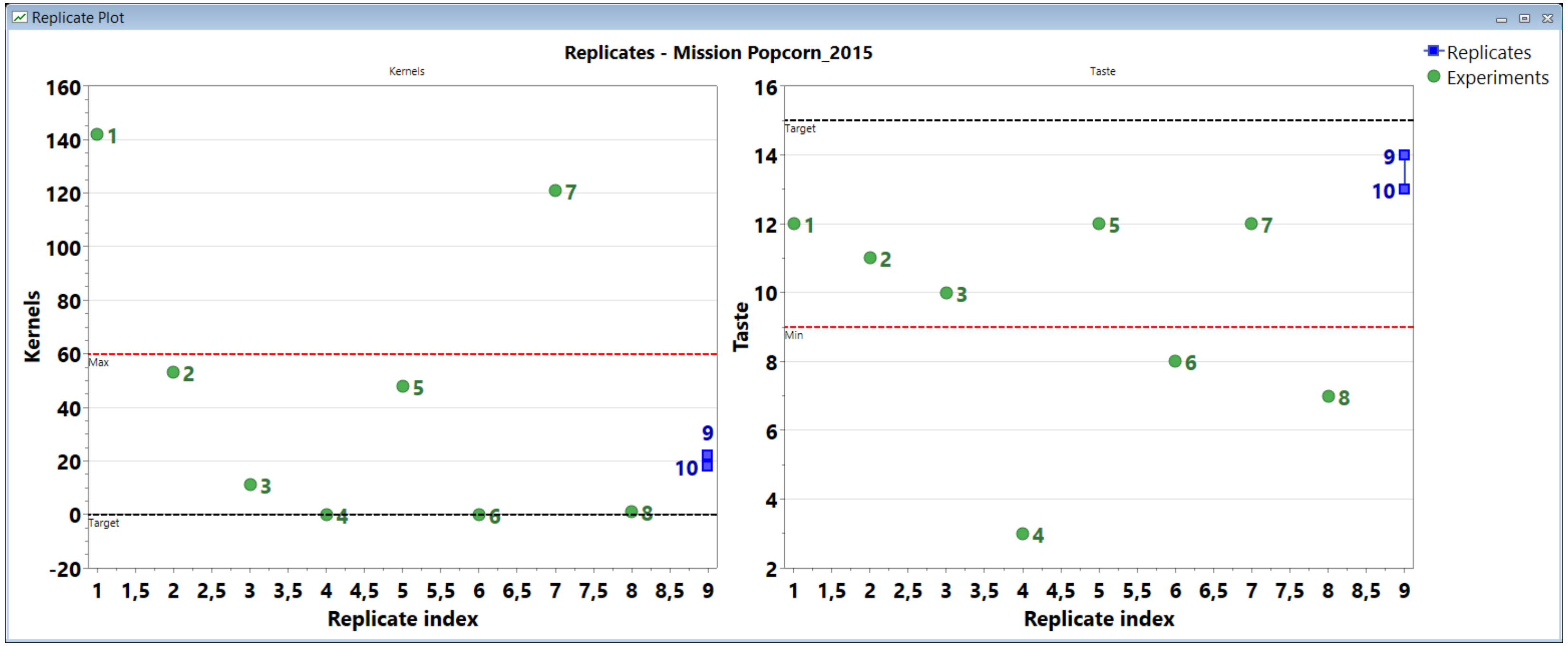
# Visualize Geometry of Design

- Colour coding provides an easy-to-understand overview.
  - Centerpoint promising with simultaneously blue (kernels) and red (taste) color.



# Replicate Plot – Evaluation of Raw Data

- The replicate plots indicate small variability among the replicates.

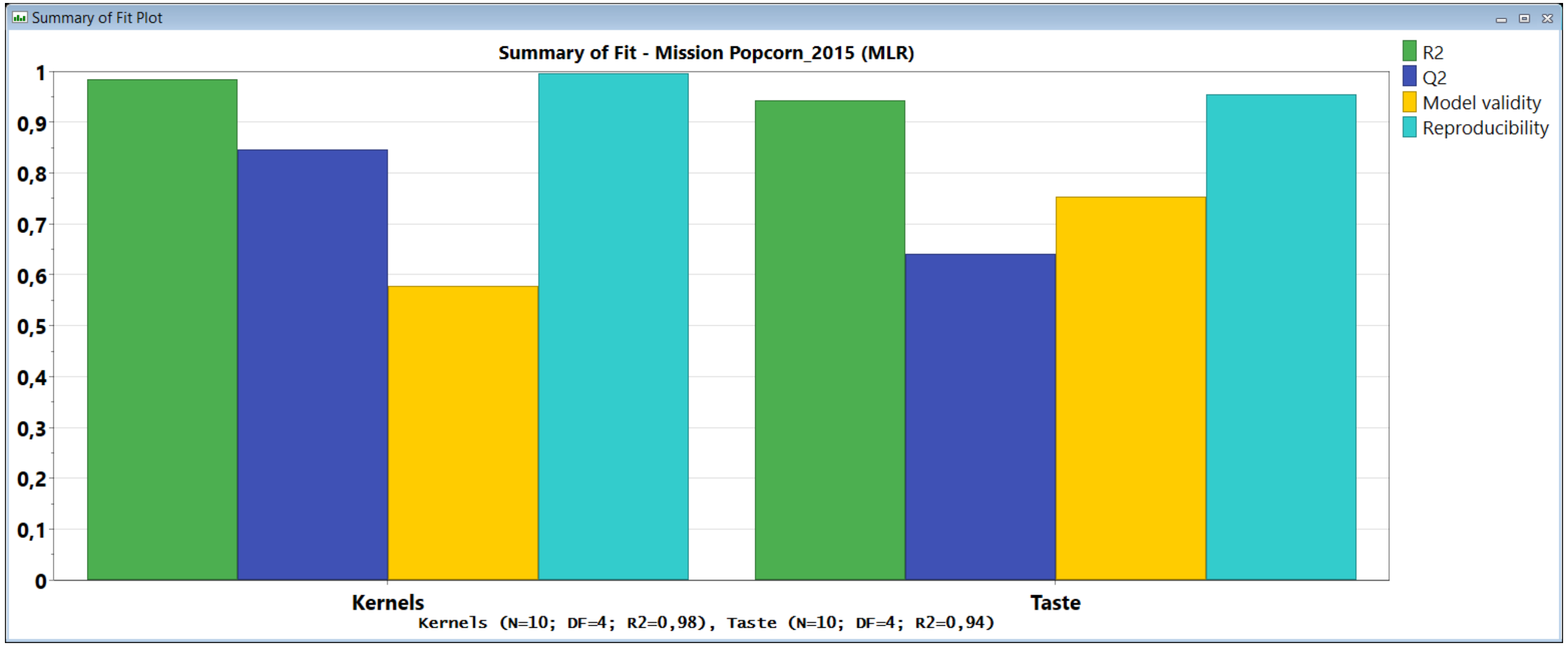





SARTORIUS

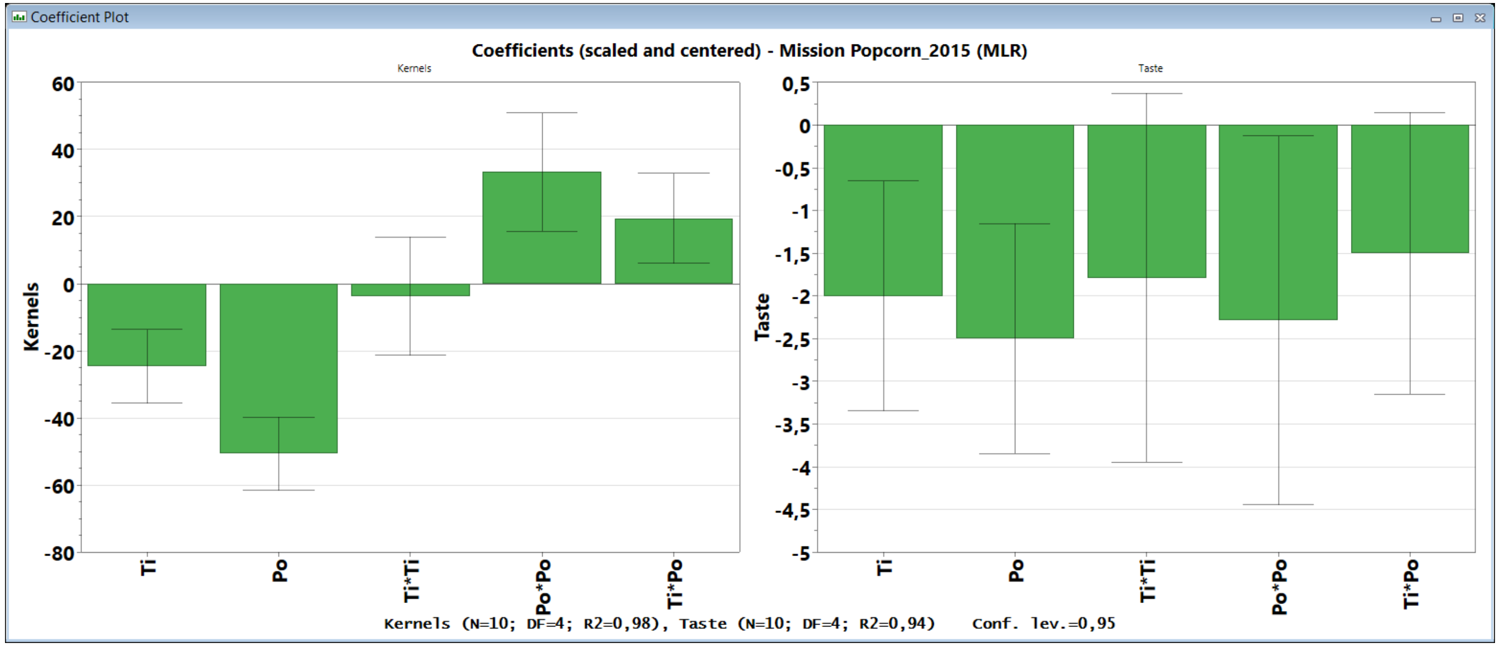
# Summary of Fit Plot – Model Performance

- When fitting the default quadratic model to the data we obtained surprisingly strong models. Model performance protocol displayed below.



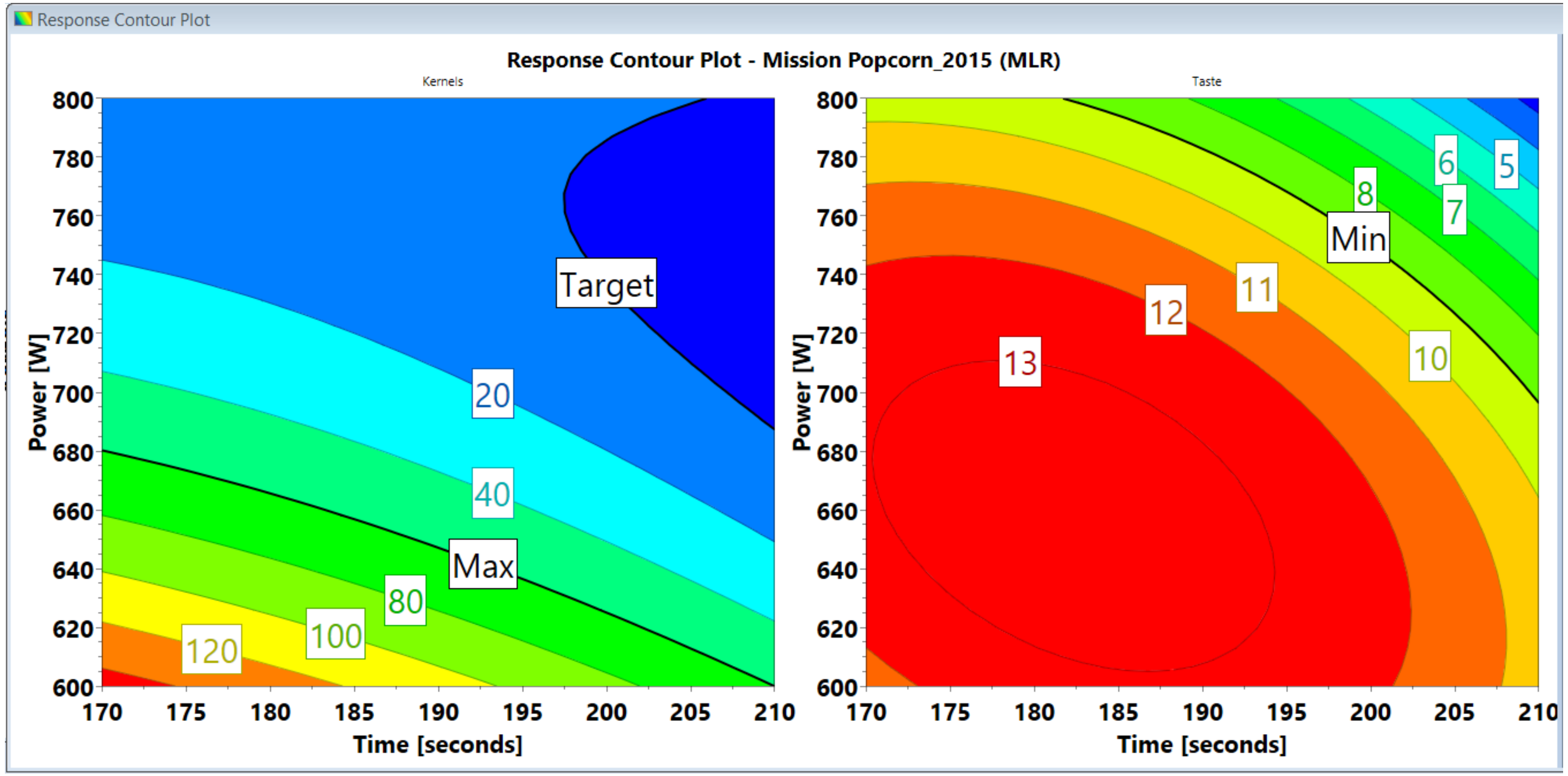
# Regression Coefficients – Model Interpretation

- Coefficients show:
  - To minimize the number of Kernels both factors should be set high.
  - Time and Power seem to have a similar impact on Taste.
  - Adjusting both factors on a lower value corresponds to increasing the Taste.



# Contour Plots – Model Visualization

- Time 182 seconds and Power 657 watts give highest taste. Conflict wrt lowest number of Kernels.



# Response Specifications - Revisited

- To arrive at a 'final' point to use, we sat down and together specified what we wanted.
- We agreed that a Taste of 9 or higher would be fully acceptable. Having 20 kernels per bowl was also deemed OK (hence a total of 60).
- Thus, we set up the following response specifications:

Responses

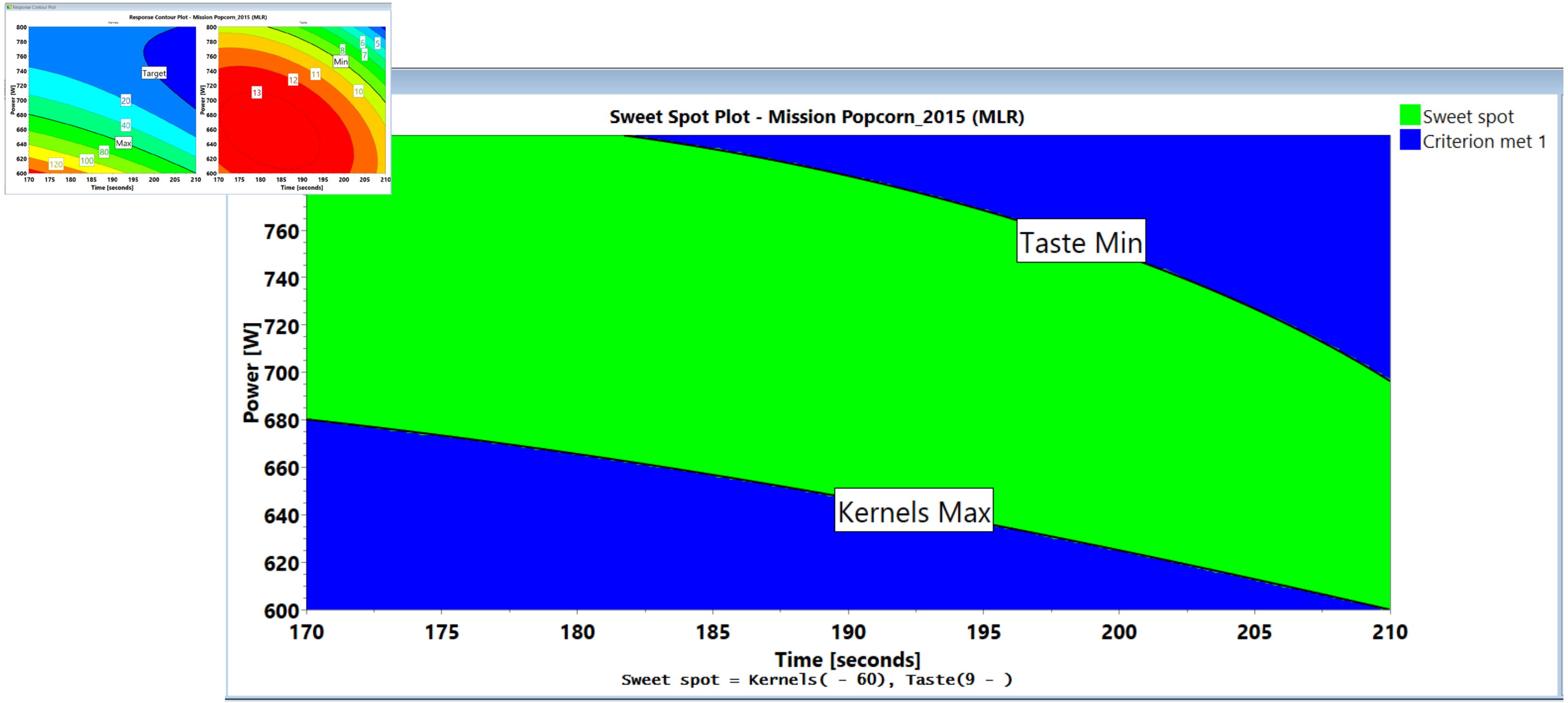
| | Name | Abbr. | Units | Transform | Type | Min | Target | Max |
|---|---|---|---|---|---|---|---|---|
| 1 | **Kernels** | Ker | | None | Regular | | 0 | 60 |
| 2 | **Taste** | Tas | | None | Regular | 9 | 15 | |
| | | *Double-click here* | *to add* | *a new response* | | | | |



SARTORIUS

# Sweet Spot Plot – Overlay of Contour Plots

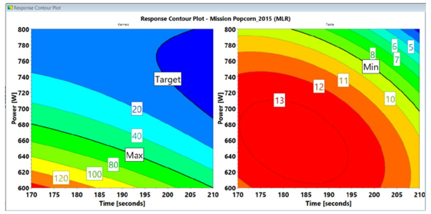
- A region of optimum exists inside the searched space (a k a knowledge space)

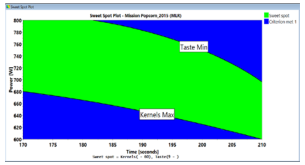


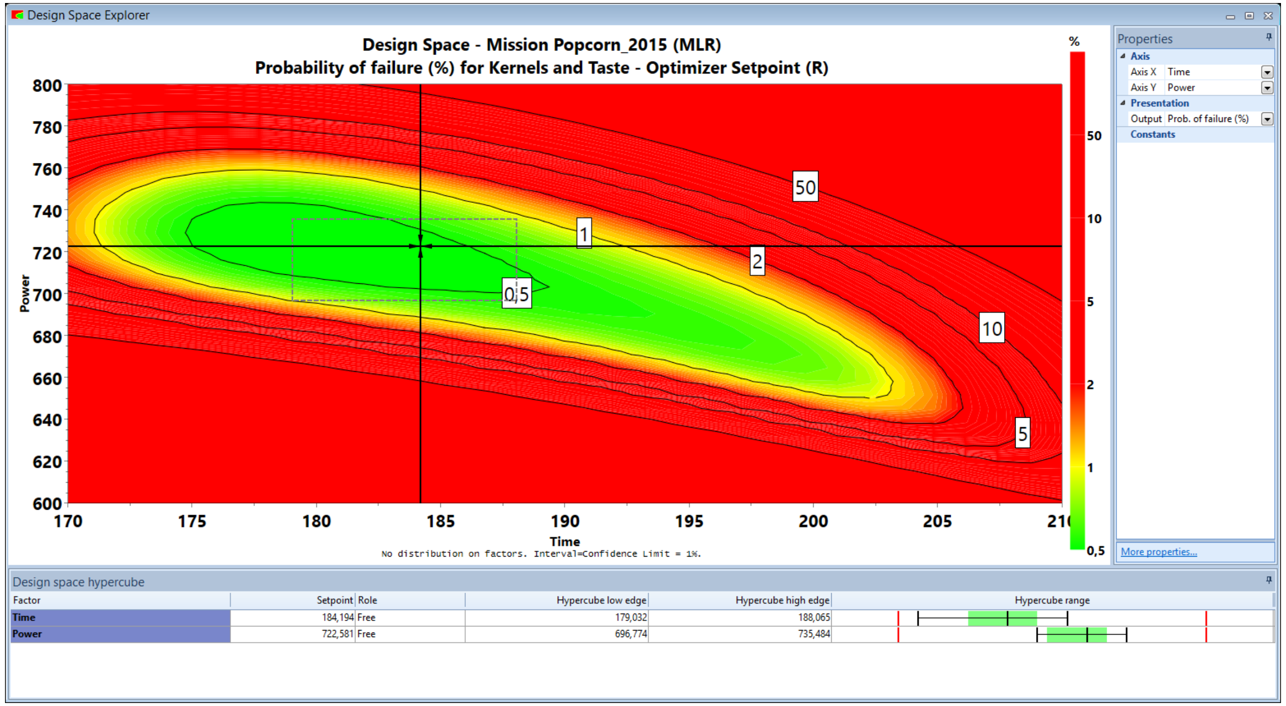



SARTORIUS

# Design Space Plot








- Design space smaller than suggested by Sweet Spot plot



SARTORIUS


# SciLife_1

## Investigating the transfection efficiency by an Ambr®15 experiment

SARTORIUS

# Background to Example Data

- The Protein Expression and Characterization Facility at SciLife Lab in Stockholm is working on the identification of novel therapeutic antibodies.
- Part of the work is to create a library of lead candidates against a specific target.
- To produce this library they want to establish an efficient transfection protocol and find out the optimal settings for transient gene expression with the FectoPRO® transfection reagent.
- The experiment was designed to investigate how the transfection efficiency was influenced by changing amounts of added DNA and the FectoPRO®:DNA reagent.
- SciLife Lab used the Ambr® 15 system.
- Ambr® 15 is set up in sets of 12 bio-reactors and with only two factors in the current investigation, 1 experimental run with 12 bio-reactors gives a good optimization design



SARTORIUS

# Settings for Critical Quality Attributes, Responses in DOE Nomenclature

- The main objective was to find the conditions yielding the highest transfection efficiency of Expi293 cells, which SciLife Lab knows correlates with a higher protein titer

- The responses measured are Transfected cells (%) and Viability (%).
  - Theoretical maximum for both responses is 100%.
  - For Transfected cells 80% was seen as a very good result.
  - Critical Quality Attributes maybe should be called Key Performance Indicator (KPI) instead

Responses

| | Name | Abbr. | Units | Transform | Type | Min | Target | Max |
|---|---|---|---|---|---|---|---|---|
| 1 | **Transfected cells** | Trans | % | None | Regular | 60 | 80 | |
| 2 | **Viability** | Via | % | None | Regular | 80 | 100 | |

# Critical Process Parameters, Factors in DOE Nomenclature

- The factors investigated are DNA amount and FectoPRO®:DNA ratio.
- DNA amount was varied by 2 equally spaced steps from 0.4 to 1.2.
- FectoPRO®:DNA ratio was varied by 2 unequally spaced steps from 0.6 to 1.6.
- Precision is the estimated variation around the given experimental point.
- Precision was retained at is default value , ±2,5% of the factor range.

Factors

| | Name | Abbr. | Units | Type | Settings | Transform | Precision |
|---|---|---|---|---|---|---|---|
| 1 | **DNA amount** | DNA | µg/mL | Multilevel | 0.4, 0.8, 1.2 | None | 0.02 |
| 2 | **FectoPRO:DNA** | Ratio | | Multilevel | 0.6, 0.8, 1.6 | None | 0.025 |

# Worksheet

- Experiment 6 was excluded due to equipment issues.
- Run Order in MODDE® -Q for Ambr® is replaced by placing each vessel in a specific culture station in a randomized order.
  - See CS1-5, CS1-1 …

Worksheet

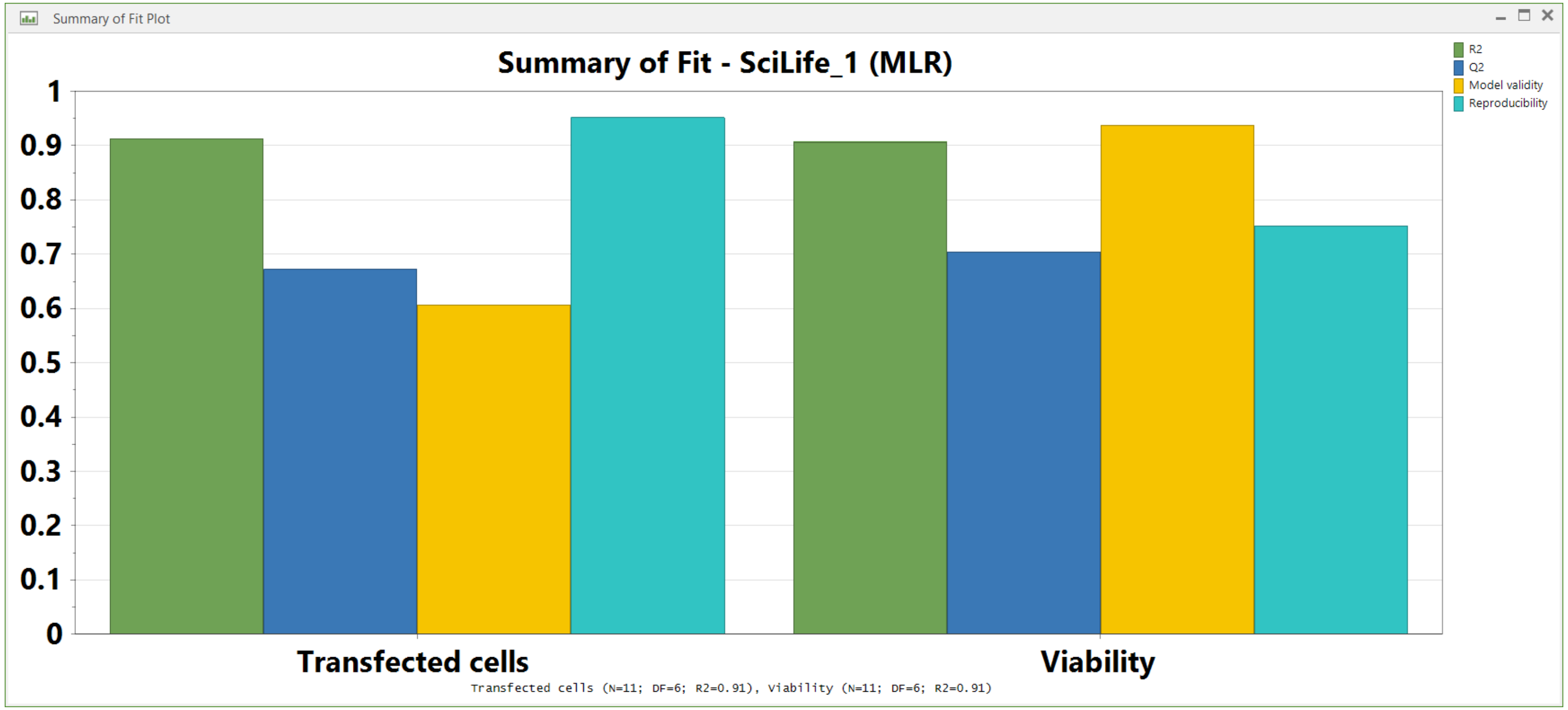
| | Exp No | Exp Name | Run Order | Incl/Excl | DNA amount | FectoPRO:DNA | Transfected cells | Viability |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 1 | CS1-5 | 5 | Incl | 0.4 | 0.6 | 37 | 87 |
| 2 | 2 | CS1-1 | 1 | Incl | 0.4 | 0.6 | 32 | 89 |
| 3 | 3 | CS1-3 | 3 | Incl | 0.4 | 0.8 | 50 | 91 |
| 4 | 4 | CS1-8 | 8 | Incl | 0.4 | 1.6 | 59 | 84 |
| 5 | 5 | CS1-12 | 12 | Incl | 0.8 | 0.6 | 58 | 81 |
| 6 | 6 | CS1-6 | 6 | Excl | 0.8 | 0.8 | 58 | 80 |
| 7 | 7 | CS1-7 | 7 | Incl | 0.8 | 0.8 | 67 | 85 |
| 8 | 8 | CS1-2 | 2 | Incl | 0.8 | 1.6 | 69 | 77 |
| 9 | 9 | CS1-4 | 4 | Incl | 1.2 | 0.6 | 75 | 82 |
| 10 | 10 | CS1-10 | 10 | Incl | 1.2 | 0.8 | 66 | 82 |
| 11 | 11 | CS1-11 | 11 | Incl | 1.2 | 1.6 | 64 | 77 |
| 12 | 12 | CS1-9 | 9 | Incl | 1.2 | 1.6 | 61 | 81 |



SARTORIUS

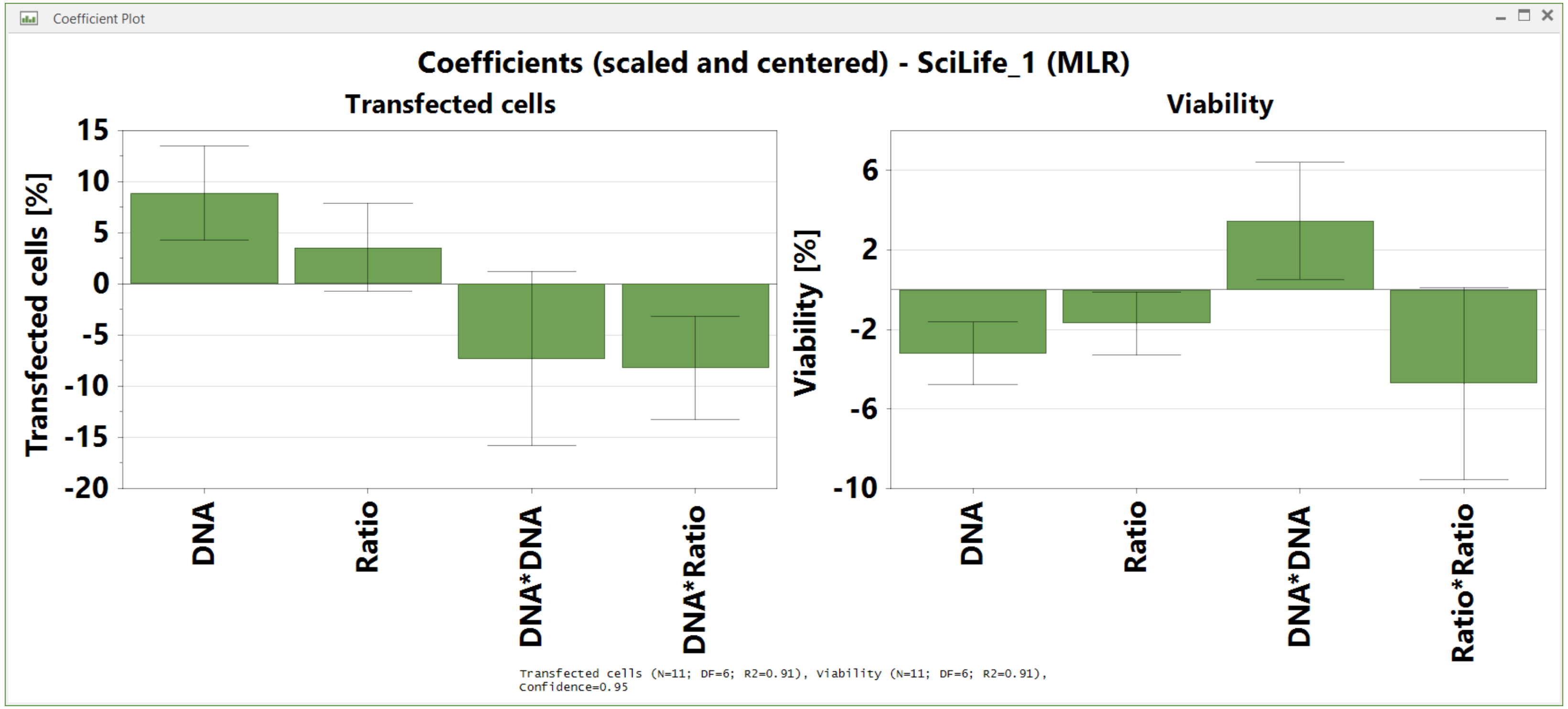
# Very Reliable Results (Good Modeling Statistics)

- A wizard will guide the user through the essential data analysis steps



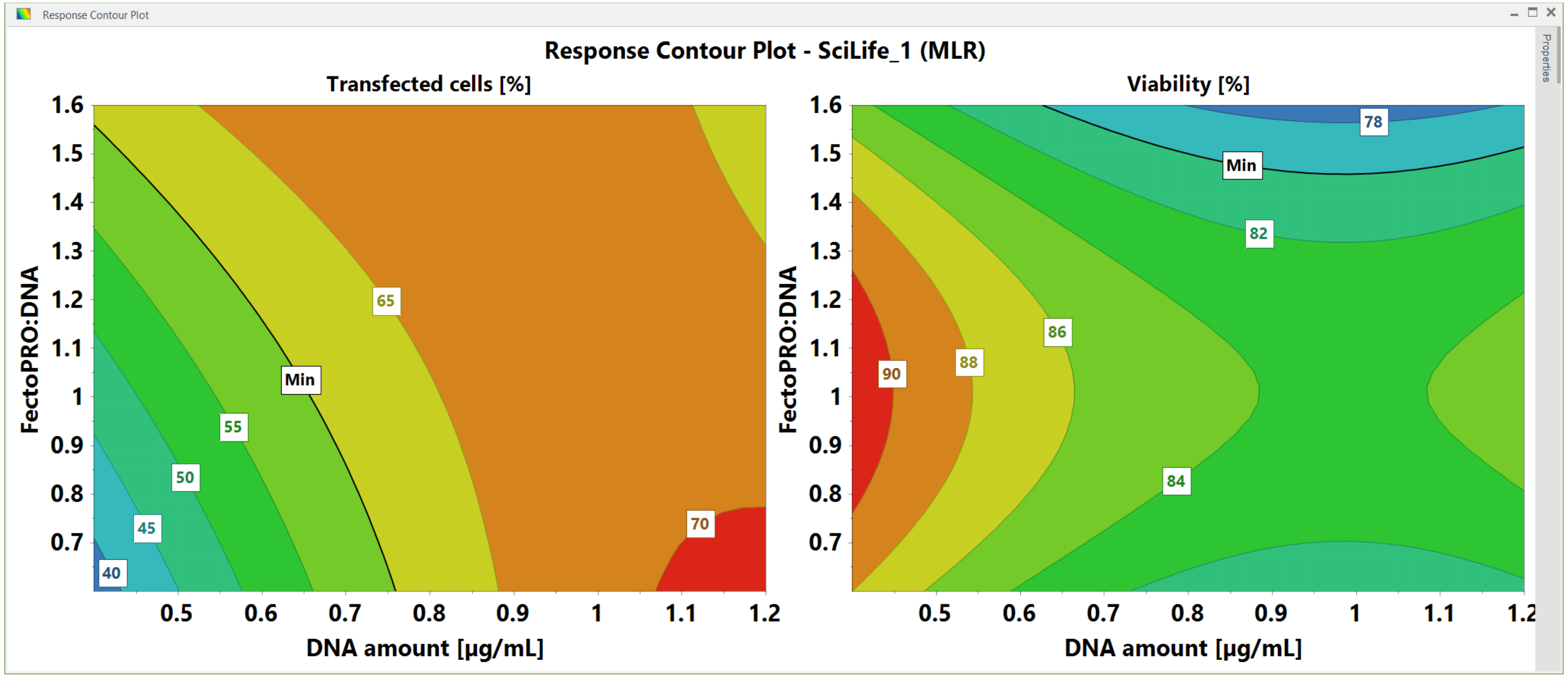
# How Factors Influence Responses

- Regression coefficients show non-linear dependencies



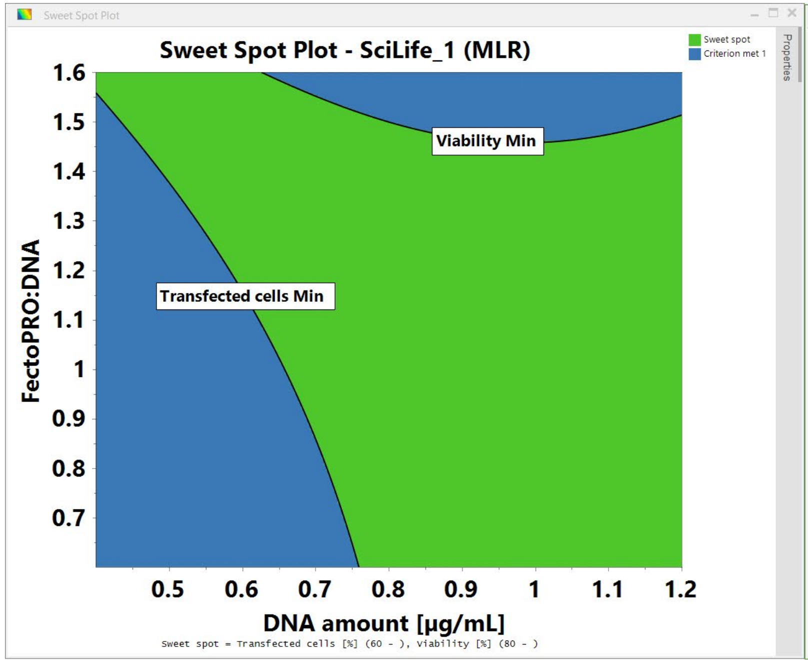
# How Factors Influence Responses

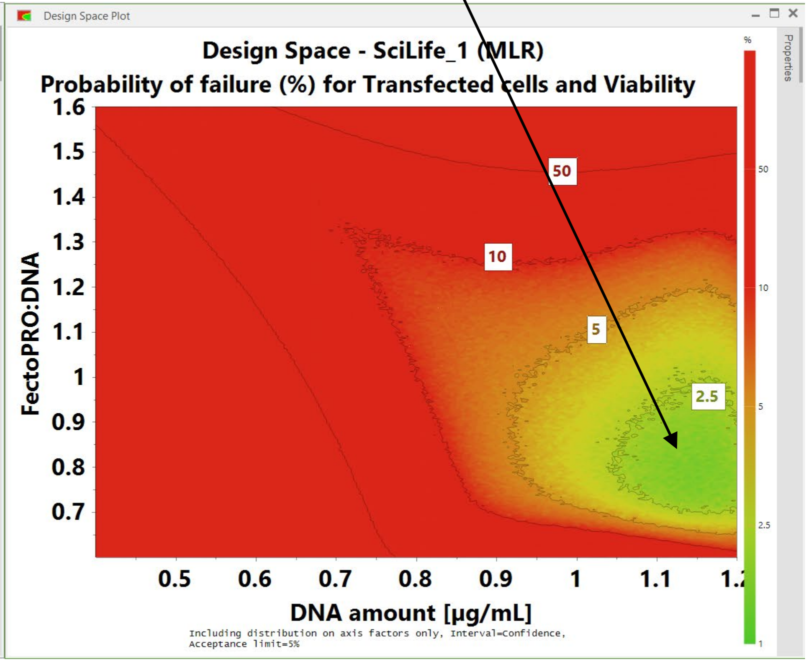
- Response contour plots



# Where Is the Best Operating Condition?

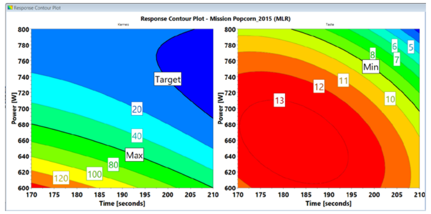
- Which combination of the factors (DNA amount and Ratio) fulfils the specifications on the responses (Transfected cells > 60% and Viability > 80%) ?
- Sweetspot plot show possible region. Design space plot show low risk region.

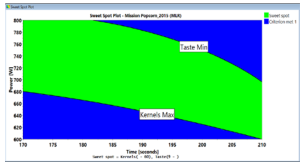


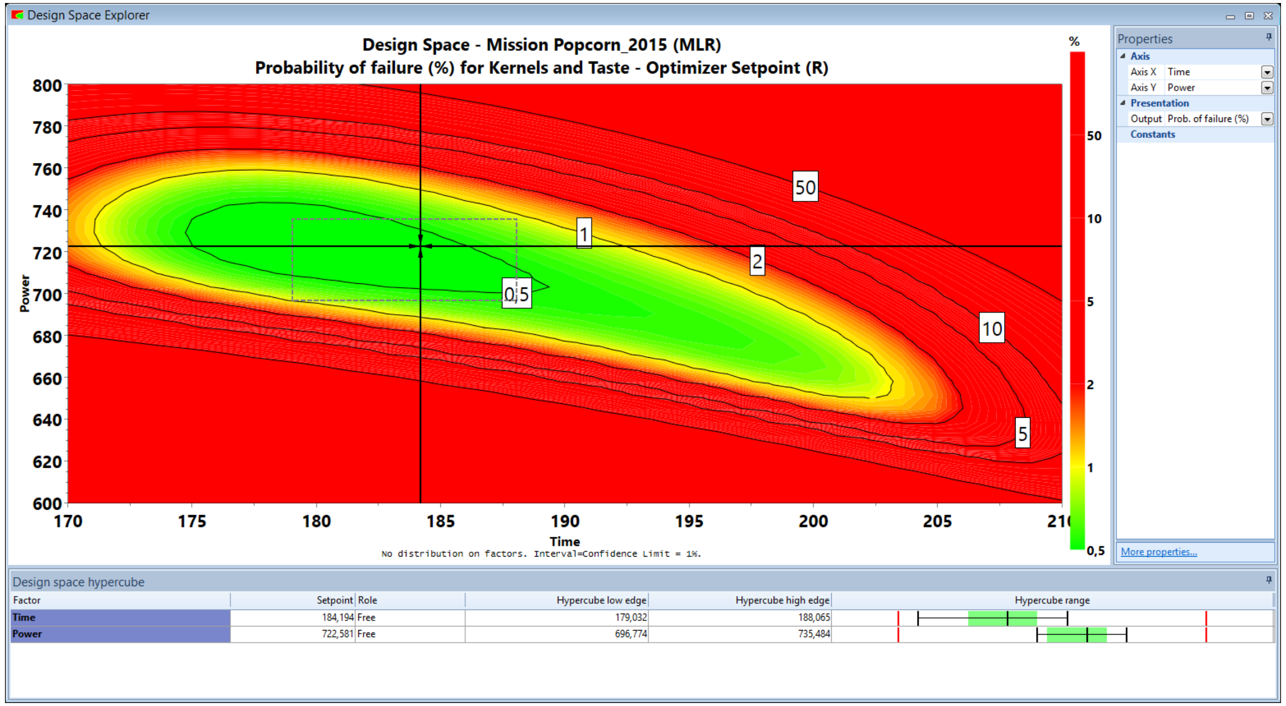



# Demo

# Design Space Plot




- Design space smaller than suggested by Sweet Spot plot



SARTORIUS

# Mission Popcorn: End Result

- Based on our joint efforts we were able to find out a suitable combination of Time (= 190 secs) and Power (= 700 watts).
- We are currently using this combination with great satisfaction. It produces well tasting popcorn without undesirable side effects such as burning and unpleasant odor. One resulting bag is seen to the right.
- The final result (apart from the popcorn) for the two end users (i.e., the two boys) was better understanding for dad's work plus having a lot of fun together with their father.




SARTORIUS

# Conclusions From Second Example

- From the results of the experiment SciLife Lab was able to set up a robust protocol, while minimizing both the plasmid DNA and transfection reagent to lower the experimental costs
- Using DOE helped them to understand the limitations of their transfection system and how to push the system towards the lowest use of raw materials



SARTORIUS

# Summary

- DOE results in a set of experiments.
- All factors are varied, systematically and independently.
- The number and type of factors and regression model specify the prerequisites.
- The DOE defines the optimal number of runs and the best factor combinations for the runs.
- DOE is used for three primary experimental objectives
  - screening: which factors are important and what are their appropriate ranges?
  - optimization: what are the optimal factor settings?
  - robustness testing: how sensitive is a response to small factor changes?
- Advantages with DOE compared to COST:
  - factor interactions are estimable
  - reliable maps of the systems
  - seen effects and noise are separable and estimable
  - probability analysis

# Upcoming Webinars (https://www.sartorius.com/en/company/exhibition-conferences)

UPCOMING EVENTS | PAST EVENTS | CALENDAR | SCHEDULE

Search

STANDARD
**Design of Experiments (DOE) for the Beginner**
TUE, JAN 26, 2021 03:00 PM - 04:00 PM CET

STANDARD
**Multivariate Data Analysis (MVDA) for the Beginner**
THU, JAN 28, 2021 03:00 PM - 04:00 PM CET

STANDARD
**Lean-and-clean DOE using One-click analysis**
TUE, FEB 16, 2021 03:00 PM - 04:00 PM CET

STANDARD
**OPLS® in process modeling**
THU, FEB 18, 2021 03:00 PM - 04:00 PM CET

STANDARD
**Robust optimization made easy**
TUE, MAR 2, 2021 03:00 PM - 04:00 PM CET

STANDARD
**Analyzing batch process data, a step-by-step guide**
THU, MAR 4, 2021 03:00 PM - 04:00 PM CET

STANDARD
**From Design of Experiments to Design Space Estimation**
TUE, MAR 23, 2021 03:00 PM - 04:00 PM CET

STANDARD
**Multiblock Orthogonal Component Analysis (MOCA) – A Novel Tool for Data Integration**
THU, MAR 25, 2021 03:00 PM - 04:00 PM CET