

# Response Surface Methods

07.12.2016

# Goals of Today's Lecture

- See how a sequence of experiments can be performed to optimize a response variable.
- Understand the difference between first-order and second-order response surfaces.

# Introductory Example: Antibody Production

- Large amounts of antibodies are obtained in biotechnological processes.
- Mice produce antibodies. Among other factors, pre-treatment by radiation and the injection of an oil increase yield.
- Of course we want to **maximize yield**.
- Due to the complexity of the underlying process we can not simply set-up a (non)linear model.
- We can say that

$$Y_i = h \left( x_i^{(1)}, x_i^{(2)} \right) + E_i,$$

where  $x^{(1)}$  = “radiation dose” and  $x^{(2)}$  = “waiting time between radiation and injection of the oil”.

- However, we do **not** have a parametric form for the function  $h$ . ⚡

## Remarks

- In general, we first have to identify those factors that really have an influence on yield.
- If we have many factors, one of the easiest approaches is to use two levels for each factor (without replicates).
- This is also known as a  $2^k$ -**design**, where  $k$  is the number of factors.
- If this is still too much effort, only a fraction of all possible combinations can be considered, leading to so called **fractional designs**.
- Identification of relevant factors can be done using **analysis of variance (ANOVA)** (regression). We will not discuss this here.

In the following, we assume that we have already identified the relevant factors.

## Back to our example

- The simplest model with an optimum would be a **quadratic function**.
- If we would know  $h$ , we could identify the **optimal setting**  $[x_0^{(1)}, x_0^{(2)}]$  of the process factors.
- $h(x^{(1)}, x^{(2)})$  is also called the **response surface**.
- Typically,  $h$  must be **estimated from data**.

## Hypothetical Example: Reaction Analysis

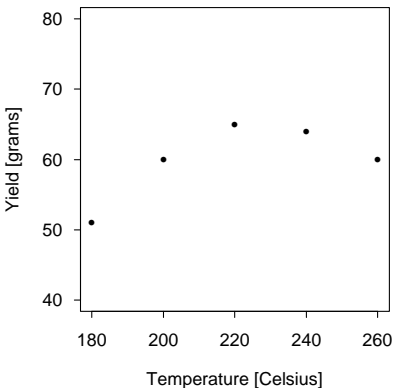
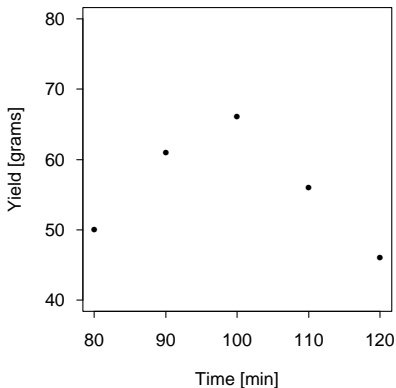
Assume yield [grams] of a process depends on

- Reaction time [min]
- Temperature [ $^{\circ}$ Celsius]

An experiment is performed as follows:

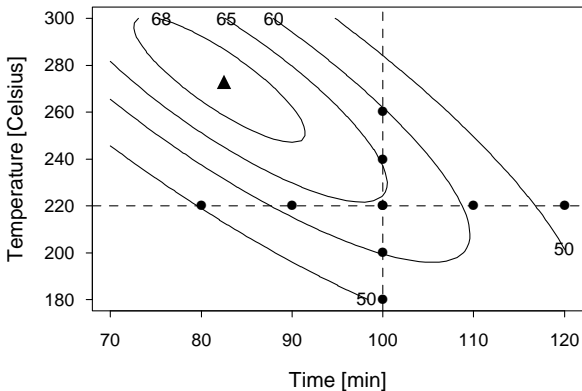
- For a temperature of  $T = 220^{\circ}C$ , different reaction times are used: 80, 90, 100, 110 and 120 minutes.
- The maximum seems to be at 100 minutes (see plot on next slide).
- Hence, we fix reaction time at 100 minutes and vary temperature at 180, 200, 220, 240 and  $260^{\circ}C$ .

- The optimum seems to be at  $220^{\circ}\text{C}$ .
- Can we now conclude that the (global) optimal value is at  $[100\text{min}, 220^{\circ}\text{C}]$ ?



Left: Temperature  $T = 220^{\circ}\text{C}$  / Right: Reaction time  $t = 100$  minutes.

## True response curve



- We miss the global optimum if we simply vary the variables “one-by-one”.
- The reason is that the maximal value of one variable is usually **not** independent of the other one! ⚡



# First-Order Response Surfaces

- As we do **not** know the true response surface, we need to get an idea about it by doing experiments at “**the right design points**”.
- Where do we **start**? Knowledge of the process may tell us that reasonable values are a temperature of  $140^{\circ}\text{C}$  and a reaction time of 60 minutes.
- Moreover, we need to have an idea about a **reasonable increment**, say we want to vary reaction time by 10 minutes and temperature by  $20^{\circ}\text{C}$  (this comes from **your** understanding of the process).

- We start with a so-called **first-order response surface**.
- The idea is to approximate the response surface with a **plane**, i.e.

$$Y_i = \theta_0 + \theta_1 x_i^{(1)} + \theta_2 x_i^{(2)} + E_i.$$

This is a **linear regression** problem!

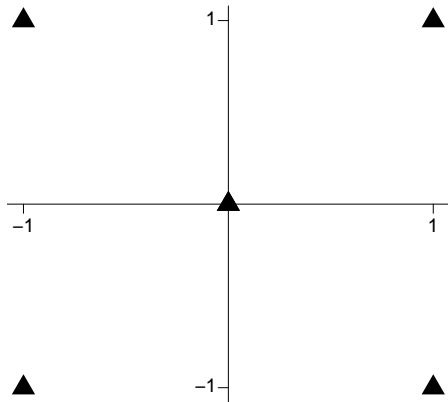
- We need some data to estimate the model parameters.
- By choosing the design points “right”, we get the best possible parameter estimates (with respect to accuracy).

Here, the **first-order design** is as follows:

Run	Variable in original units		Variable in coded units		Yield
	Temperature [ $^{\circ}$ C]	Time [min]	Temperature	Time	Y [grams]
1	120	50	-1	-1	52
2	160	50	+1	-1	62
3	120	70	-1	+1	60
4	160	70	+1	+1	70
5	140	60	0	0	63
6	140	60	0	0	65

First-order design and measurement results for the example “Reaction Analysis”. The single experiments (runs) were performed in **random order**: 5, 4, 2, 6, 1, 7, 3.

A **visualization in coded units** illustrates the special “layout” of the design points.



- The fitted plane, the so called **first-order response surface**, is given by

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x^{(1)} + \hat{\theta}_2 x^{(2)}.$$

- This is only an **approximation** of the real response surface.
- Note: The plane can **not** give us an optimal value. ⚡
- But it enables us to determine the **direction of steepest ascent**. We denote it by

$$\begin{bmatrix} c^{(1)} \\ c^{(2)} \end{bmatrix}.$$

- This will be the “fastest” way to get large values of the response variable. Hence, we should follow that direction.

- This means we should perform additional experiments along

$$\begin{bmatrix} x_0^{(1)} \\ x_0^{(2)} \end{bmatrix} + j \cdot \begin{bmatrix} c^{(1)} \\ c^{(2)} \end{bmatrix},$$

where  $[x_0^{(1)}, x_0^{(2)}]^T$  is the point in the center of our experimental design and  $j = 1, 2, \dots$

## Remarks

- Observations in the center of the design have **no influence** on the parameter estimates of the slopes  $\theta_1$  and  $\theta_2$  (and hence no influence on the gradient).
- This is due to the special arrangement of the other points.

## Why is it useful to have multiple observations in the center?

- They can be used to estimate the measurement error without relying on any assumptions.
- They make it possible to **detect curvature**.
- If there is no (local) curvature, the average of the response values in the center and the average of the response values of the other design points should be “more or less equal”.
- It is possible to perform a statistical test to check for curvature.

- If there is (significant) curvature, we don't want to use a plane but a quadratic function (see later).
- Otherwise we move along the direction of the gradient.

## Randomization

Randomization is **crucial** for all experiments. Hence, we should really perform the experiments in a random sequence, even though it would be usually more convenient to e.g. do the two experiments in the center right after each other.



## Back to our example

- The first-order response surface is

$$\hat{y} = 62 + 5\tilde{x}^{(1)} + 4\tilde{x}^{(2)} = 3 + 0.25 \cdot x^{(1)} + 0.4 \cdot x^{(2)},$$

where  $[\tilde{x}^{(1)}, \tilde{x}^{(2)}]^T$  are the coded  $x$ -values (taking values  $\pm 1$ ).

- The steepest ascent direction is given by the **gradient**.
- If we calculate the gradient for the coded variables this leads to the direction vector

$$\begin{bmatrix} 5 \\ 4 \end{bmatrix},$$

that corresponds to

$$\begin{bmatrix} 100 \\ 40 \end{bmatrix}$$

in original units.

- We get a different result if we directly take the gradient for the original variables. ⚡
- The reason is because we do a coordinate transformation.
- Working with coded units in general makes sense because the scale of the predictors is chosen such that increasing a variable by one unit should have more or less the same effect on the response for all variables.

- Further experiments are now performed along

$$\begin{bmatrix} 140 \\ 60 \end{bmatrix} + j \cdot \begin{bmatrix} 25 \\ 10 \end{bmatrix}, j = 1, 2, \dots$$

which corresponds to the steepest ascent direction with respect to the coded  $x$ -values.

- This leads to new data

	Temperature [ $^{\circ}$ C]	Time [min]	Y
1	165	70	72
2	190	80	77
3	215	90	79
4	240	100	76
5	265	110	70

Experimental design and measurement results for experiments along the steepest ascent direction for the example “Reaction Analysis”.

- We reach a maximum at a temperature of  $215^{\circ}\text{C}$  and at a reaction time of 90 minutes.
- Now we could do a further first-order design around this new “optimal” values to get a new gradient and then iterate many times.
- In general, the hope is that we are already close to the optimal solution. We therefore use a **quadratic approach** to model a “peak”.

## Second-Order Response Surfaces

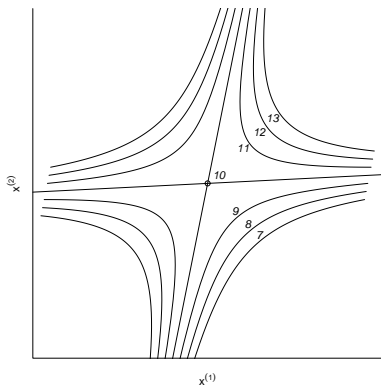
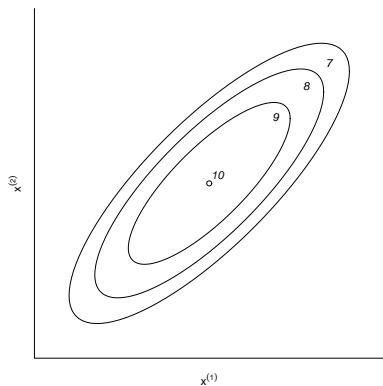
- Once we are close to the optimal solution, the plane will be rather flat.
- We expect that the optimal solution will be somewhere in our experimental set-up (a peak). Hence, we **expect curvature**.
- We can directly model such a peak by using a **second-order polynomial**

$$Y_i = \theta_0 + \theta_1 x_i^{(1)} + \theta_2 x_i^{(2)} + \theta_{11} (x_i^{(1)})^2 + \theta_{22} (x_i^{(2)})^2 + \theta_{12} x_i^{(1)} x_i^{(2)} + E_i.$$

Remember: This is still a **linear model**!

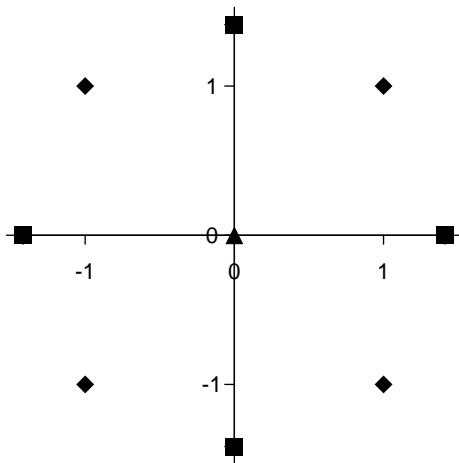
- We now have **more parameters**, hence we need more observations to fit the model.

## Two Examples of Second-Order Response Surfaces



Left: Situation with a maximum / Right: Saddle

- We therefore expand our design by adding **additional points**.



- This is a so called **rotatable central composite design**. All points have the same distance from the center.

- Hence, we can simply add the points on the axis (they have coordinates  $\pm\sqrt{2}$ ) to our first order design.
- Again, by choosing the points that way the model has some nice theoretical properties.
- New data can be seen on the next slide.
- We use this data and fit our second-order polynomial model (with linear regression!), leading to the **second-order response surface**

$$\hat{y} = -278 + 2.0 \cdot x^{(1)} + 3.2 \cdot x^{(2)} + 0.0060 \cdot (x^{(1)})^2 + 0.026 \cdot (x^{(2)})^2 + 0.006 \cdot x^{(1)} x^{(2)}.$$



## New Data

Run	Variables in original units		Variables in coded units		Yield
	Temperature [ $^{\circ}\text{C}$ ]	Time [min]	Temperature	Time	Y [grams]
1	195	80	-1	-1	78
2	235	80	+1	-1	76
3	195	100	-1	+1	72
4	235	100	+1	+1	75
5	187	90	$-\sqrt{2}$	0	74
6	243	90	$+\sqrt{2}$	0	76
7	215	76	0	$-\sqrt{2}$	77
8	215	104	0	$+\sqrt{2}$	72
9	215	90	0	0	80

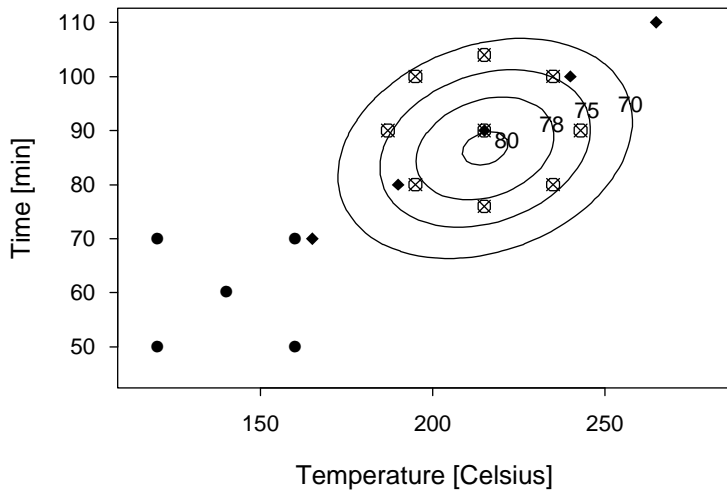
Rotatable central composite design and measurement results for the example “Reaction Analysis”.

- Given the second-order response surface, it's possible to find an **analytical solution** for the optimal value by taking partial derivatives and solving the resulting linear equation system for  $x_0^{(1)}$  and  $x_0^{(2)}$ , leading to

$$x_0^{(1)} = \frac{\hat{\theta}_{12} \hat{\theta}_2 - 2 \hat{\theta}_{22} \hat{\theta}_1}{4 \hat{\theta}_{11} \hat{\theta}_{22} - \hat{\theta}_{12}^2} \quad x_0^{(2)} = \frac{\hat{\theta}_{12} \hat{\theta}_1 - 2 \hat{\theta}_{11} \hat{\theta}_2}{4 \hat{\theta}_{11} \hat{\theta}_{22} - \hat{\theta}_{12}^2}.$$

(see lectures notes for derivation).

## Summary Picture



## Back to our Original Example

- A radioactive dose of 200 rads and a time of 14 days is used as starting point. Around this point we use a first-order design.

Run	Variables in original units		Variables in coded units		Yield
	RadDos [rads]	Time [days]	RadDos	Time	Y
1	100	7	-1	-1	207
2	100	21	-1	+1	257
3	300	7	+1	-1	306
4	300	21	+1	+1	570
5	200	14	0	0	630
6	200	14	0	0	528
7	200	14	0	0	609

First-order design and measurement results for the example “Antibody Production”.

What can we observe? Yield decreases at the borders!

- If we check for significant curvature, we get a 95%-confidence interval for the difference  $(\mu_c - \mu_f)$  of

$$\begin{aligned}\bar{Y}_c - \bar{Y}_f \pm q_{0.975}^{t_{n_c-1}} \cdot \sqrt{s^2(1/n_c + 1/2^k)} &= 589 - 335 \pm 4.30 \cdot 41.1 \\ &= [77, 431].\end{aligned}$$

- As the confidence interval doesn't cover zero there is significant curvature.
- We therefore expand our design to a rotatable central composite design by doing additional measurements.

- New data:

Run	Variables in original units		Variables in coded units		Yield
	RadDos [rads]	Time [days]	RadDos	Time	Y
8	200	4	0	$-\sqrt{2}$	315
9	200	24	0	$+\sqrt{2}$	154
10	59	14	$-\sqrt{2}$	0	100
11	341	14	$+\sqrt{2}$	0	513

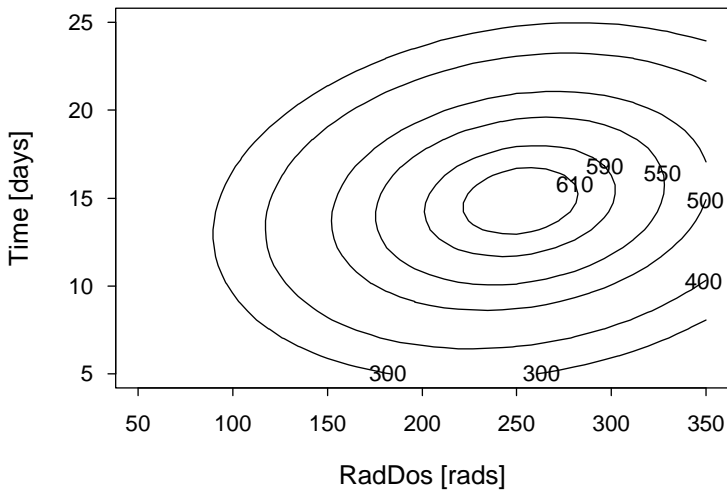
Rotatable central composite design and measurement results for the example “Antibody Production”.

- The estimated response surface is

$$\hat{Y} = -608.4 + 5.237 \cdot \text{RadDos} + 77.0 \cdot \text{Time} \\ - 0.0127 \cdot \text{RadDos}^2 - 3.243 \cdot \text{Time}^2 + 0.0764 \cdot \text{RadDos} \cdot \text{Time}.$$

- We can identify the optimal conditions in the contour plot (see next slide) or analytically:  $\text{RadDos}_{opt} \approx 250$  rads and  $\text{time}_{opt} \approx 15$  days.

## Contour Plot of Second-Order Response Surface



# Summary

- Using a sequence of experiments, we can find the optimal setting of our variables.
- If we are “far away”, we use **first-order designs**, leading to a first-order response surface (a plane).
- As soon as we are close to the optimum, a **rotatable central composite design** can be used to estimate the **second-order response surface**. The optimum on that response surface can be determined analytically.