DoE Principles
e.g.: “Optimization of Injection and Combustion”

Hans – Michael Koegeler

13.05.2013

Introduction to DoE:
Target: **Model based development and optimization:**

- Design of Experiments – What?
- Design of Experiments – Why?
- Design of Experiments – How?
- Comparison of conventional approach to DoE
- Example based theory
- Advantages of DoE
- Preconditions for DoE

Design of Experiments – What?

Design **o**f **E**xperiments

DoE = Statistical Design of Experiments

DoE Methods distribute an optimized low number of Parameter Combinations
in an area of influencing parameters (design space)
in order to get a statistically assured, Empirical Model to predict
the Experimental Result on any position in the design space.

Preknowledge can and shall be used to support the design process.

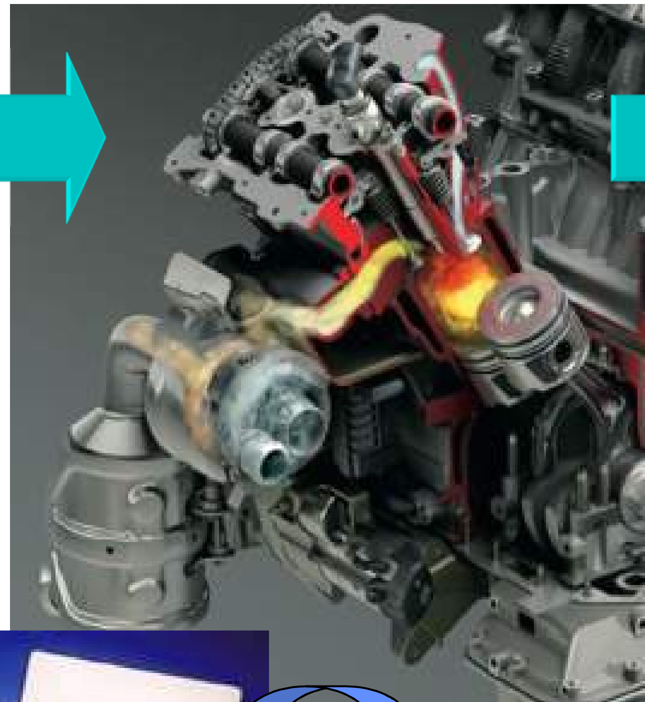
- 1924: DoE started
  based on agricultural Questions:

First systematic experiments to Predict and Optimize the Harvest as Function (dung, soil conditions, watering and others)
by Sir Ronald A. Fisher

Sir Ronald A. Fisher (1890 – 1962)

Design of Experiments – Why?
e.g.: typical task in R+D / Calibration:
Optimization of Injection and combustion:

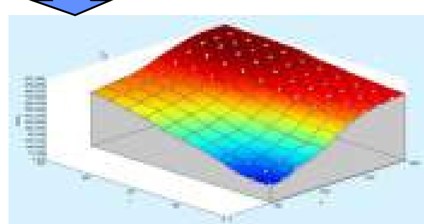
Input Parameters:
High amount of variables



Target:
- **Min fuel consumption**
- **Emission limits**

Diesel:
- nozzle type
- Injection pressure
- Start of Injection
- EGR
- Boost pressure
- several injections
  etc.

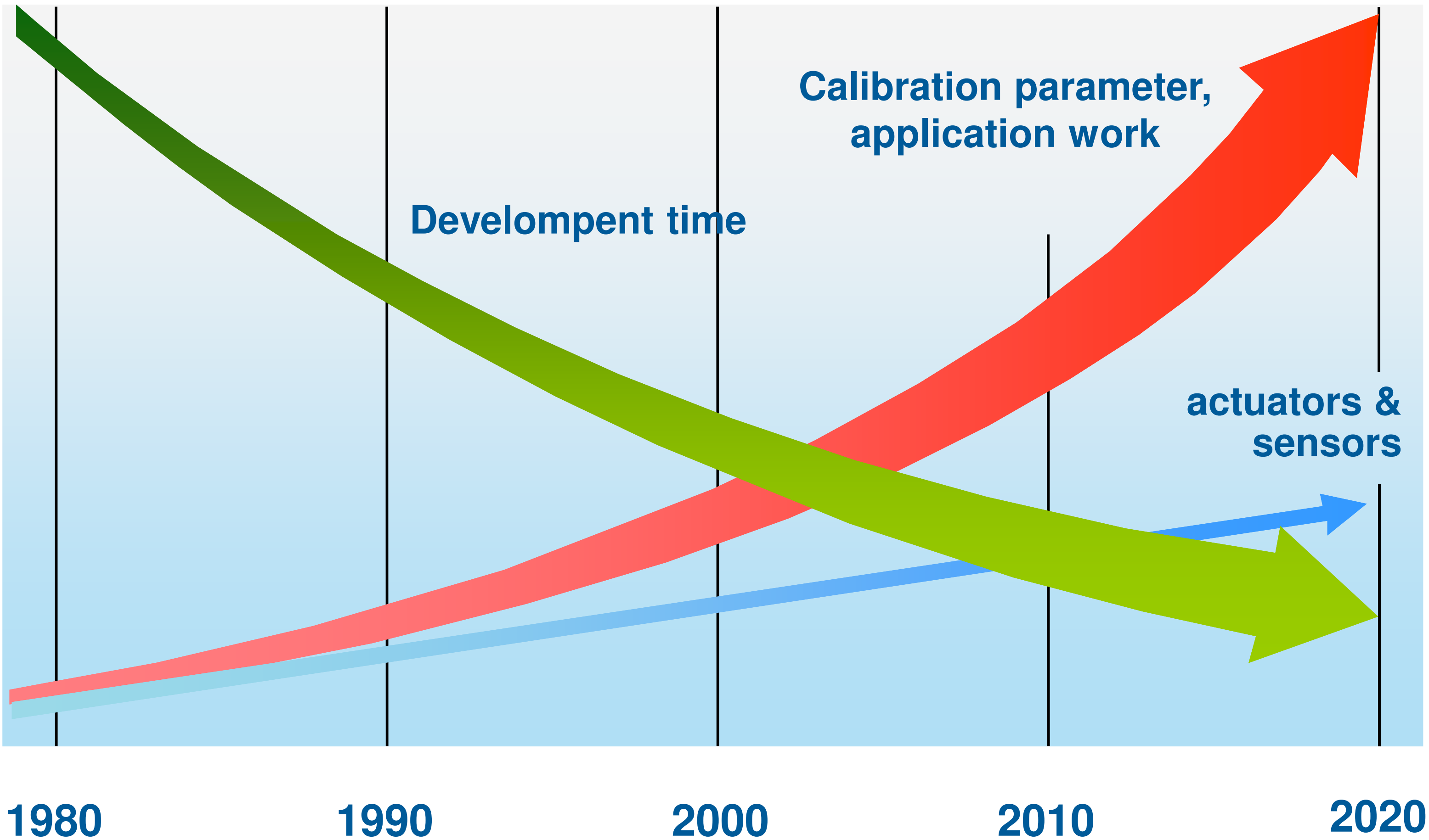
→ Hardware selection and
→ corresponding calibration!



Engine Maps = Application Label

Design of Experiments – Why?

Design of Experiments – Why?
Higher flexibility = dramatic increase of complexity



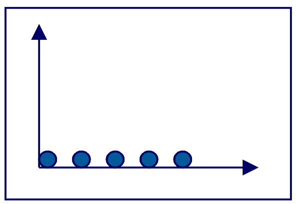
ECU-calibration: new processes and tools are necessary

Design of Experiments – Why?
Full-Factorial optimization

E.g.: 5 measuring points for each direction each

1 parameter

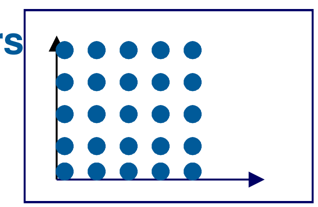
SOI



Number of measurements = 5

2 parameters

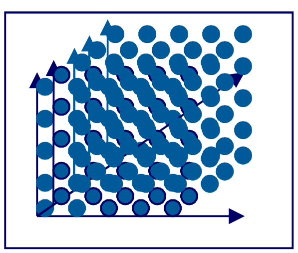
SOI
Rail press.



Number of measurements = 25

3 parameters

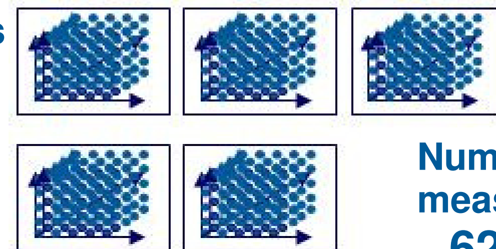
SOI
Rail press.
EGR



Number of measurements = 125

4 parameters

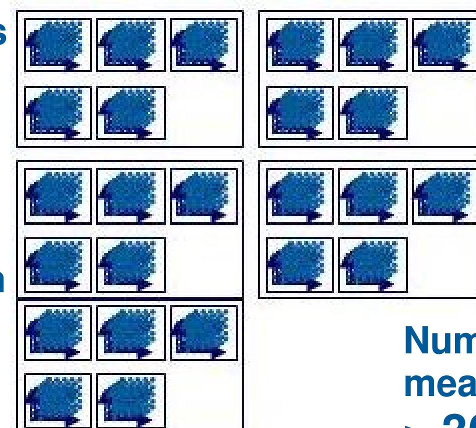
SOI
Rail press.
EGR
Boost press.



Number of measurements = 625

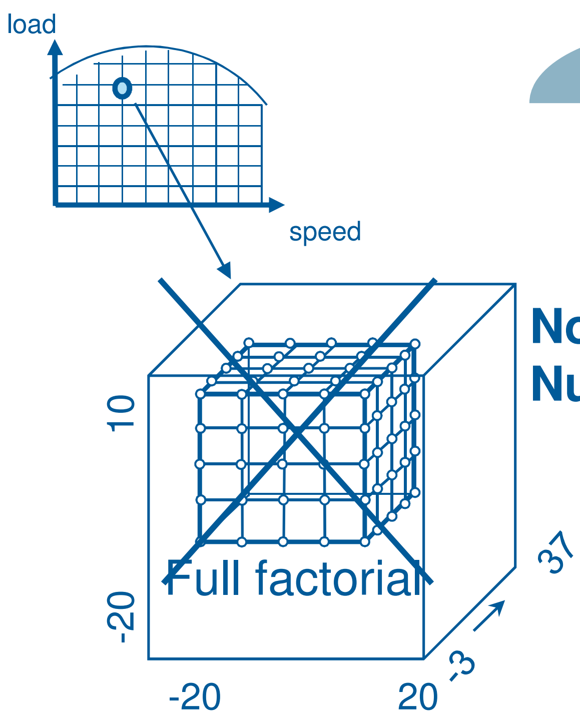
5 parameters

SOI
Rail press.
EGR
Boost press.
Pilot Injection quantity



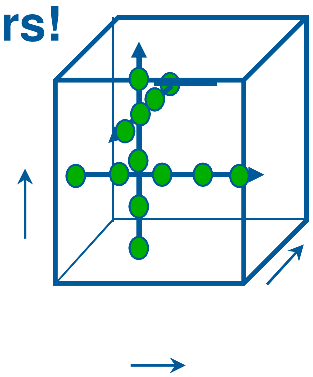
Number of measurements > 3000

Design of Experiments – Why?
Often used process: OFAT – one factor at a time



Not possible:
Number of parameters!

Classic:
Parameter variation
OFAT – one factor at a time



selected experimental space

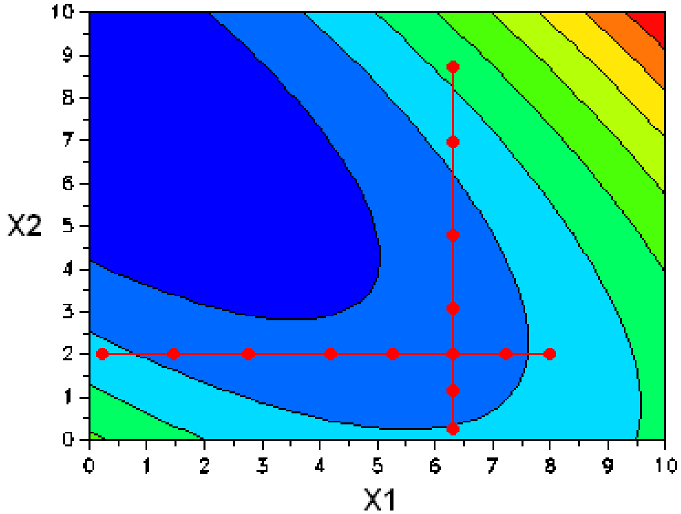
Design of Experiments – Why?
Consequences of „one factor at a time“

Alteration of one variation parameter
while leaving the other one constant,
does not lead automatically to the optimum

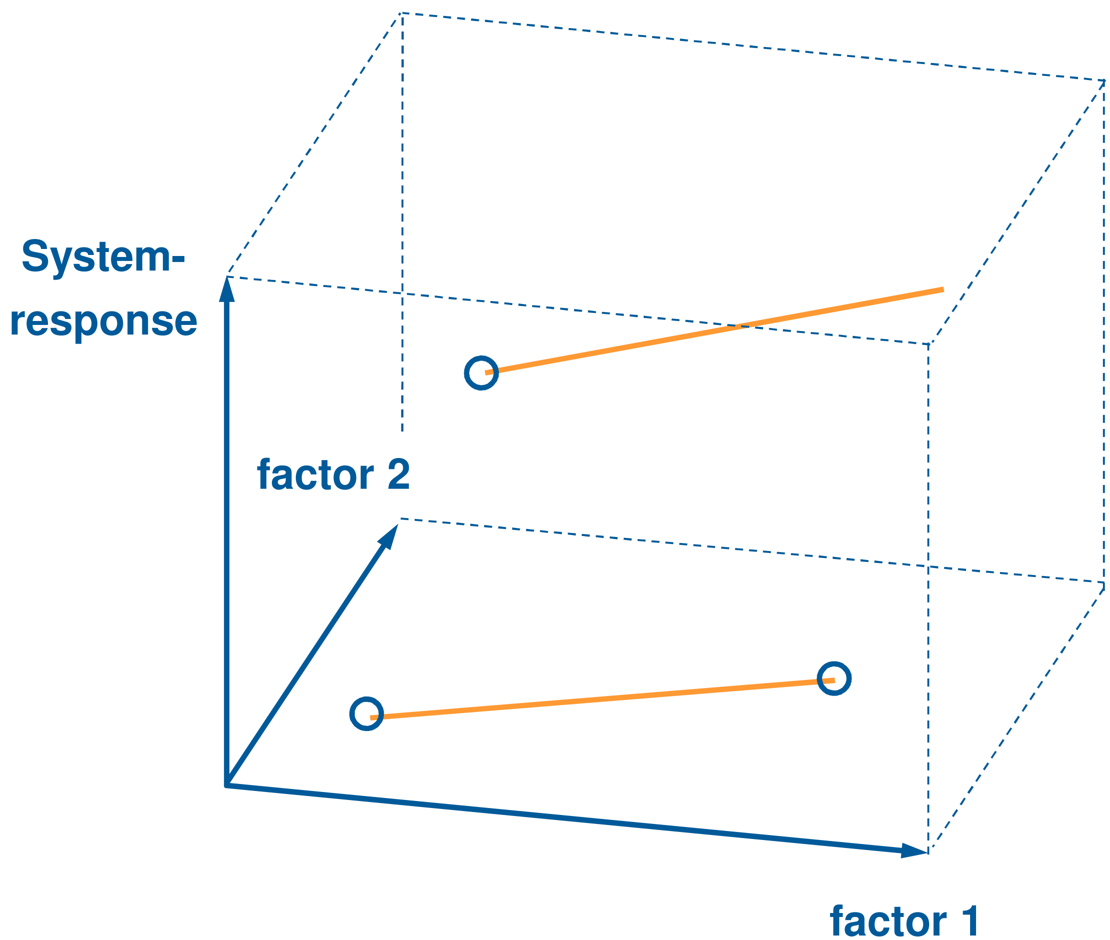
Different start points result in different optima

Many measurements with limited information

No quantification of the interaction of the individual variables



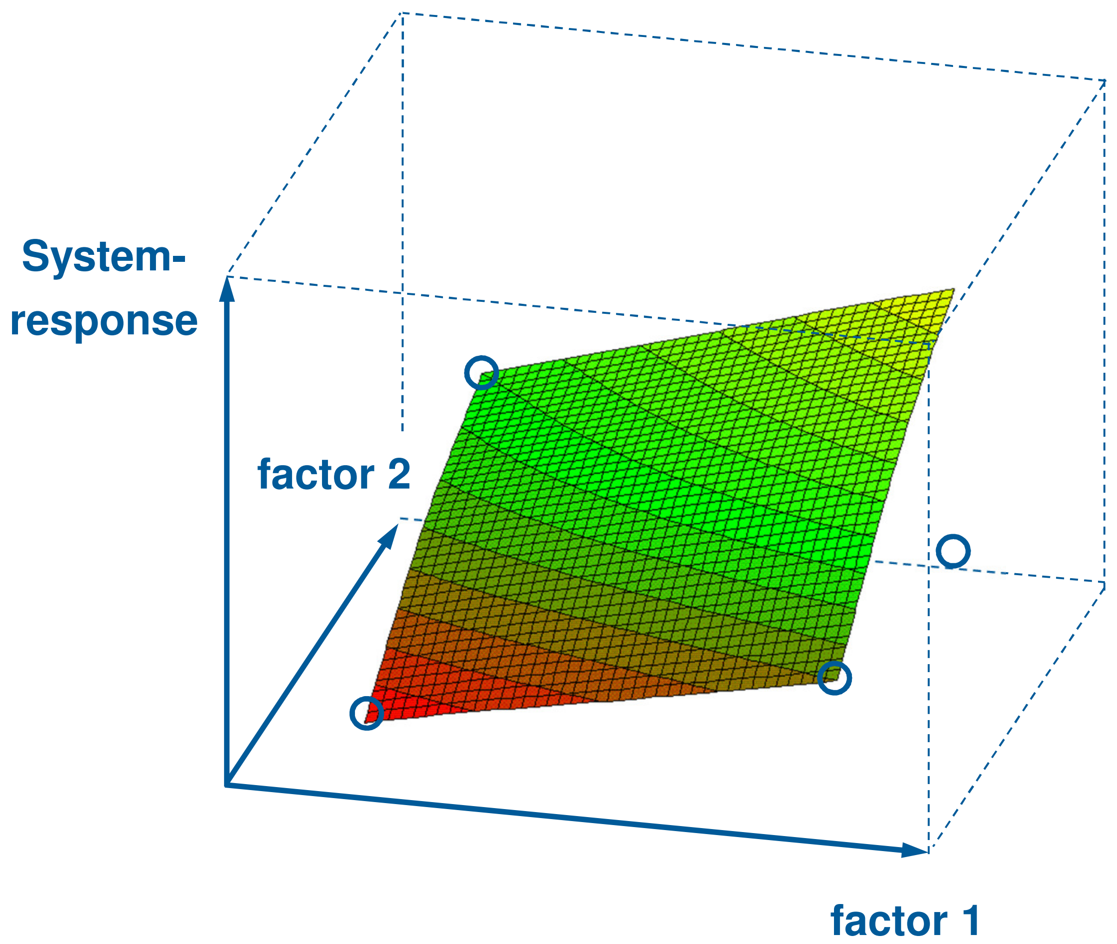
Design of Experiments – Why?
Interaction What is that?



Estimated system behavior when varying factor 2

System behavior when varying factor 1 & factor 2 constant

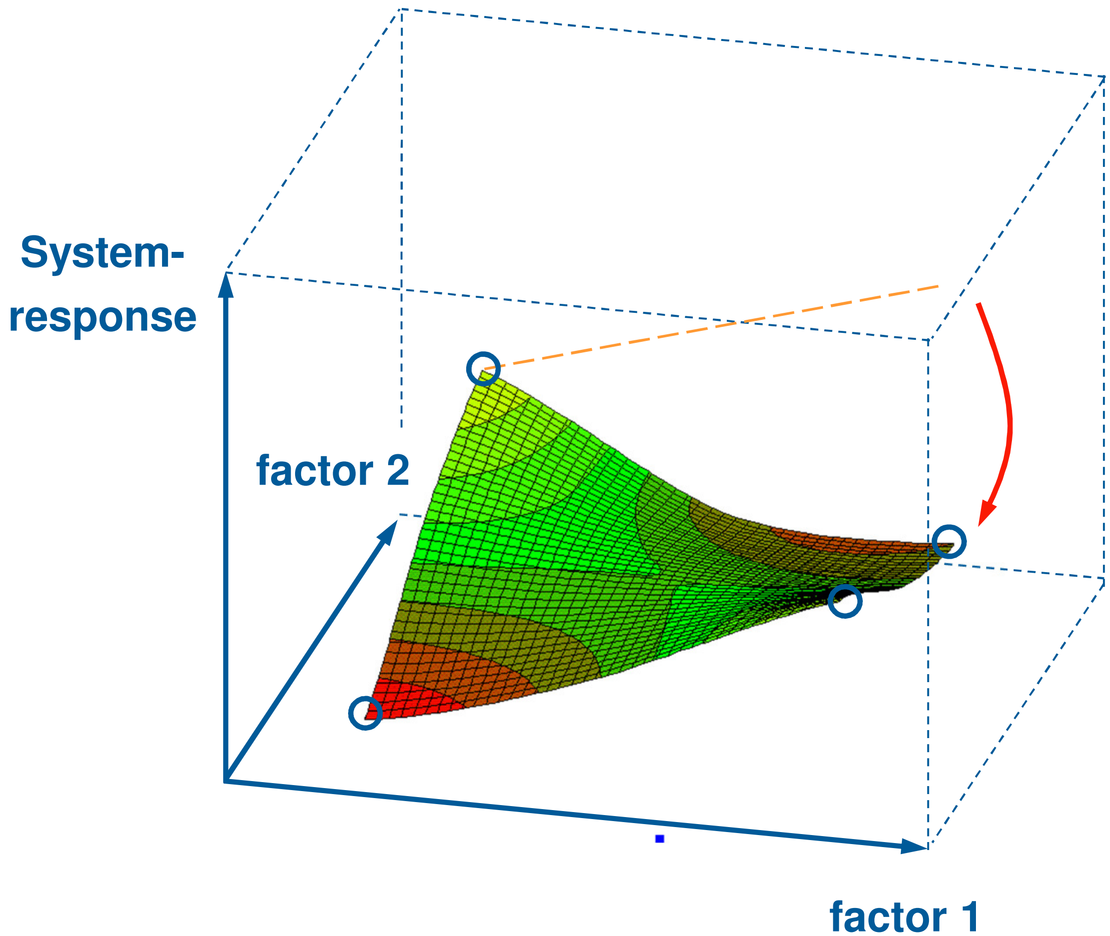
Design of Experiments – Why?
Interaction What is that?



Estimated system behavior when varying factor 2

System behavior when varying factor 2 & factor 2 constant

Design of Experiments – Why?
Interaction What is that?

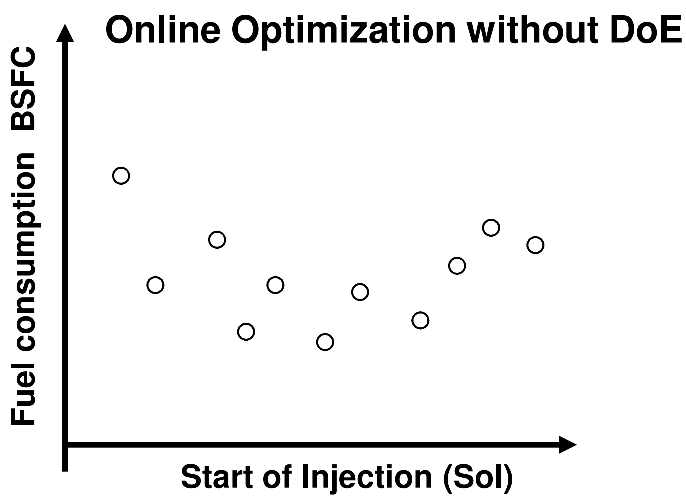


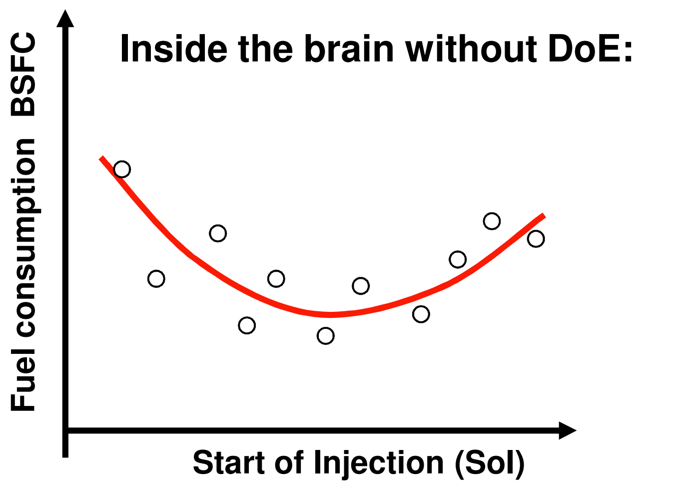
Alternative behavior:

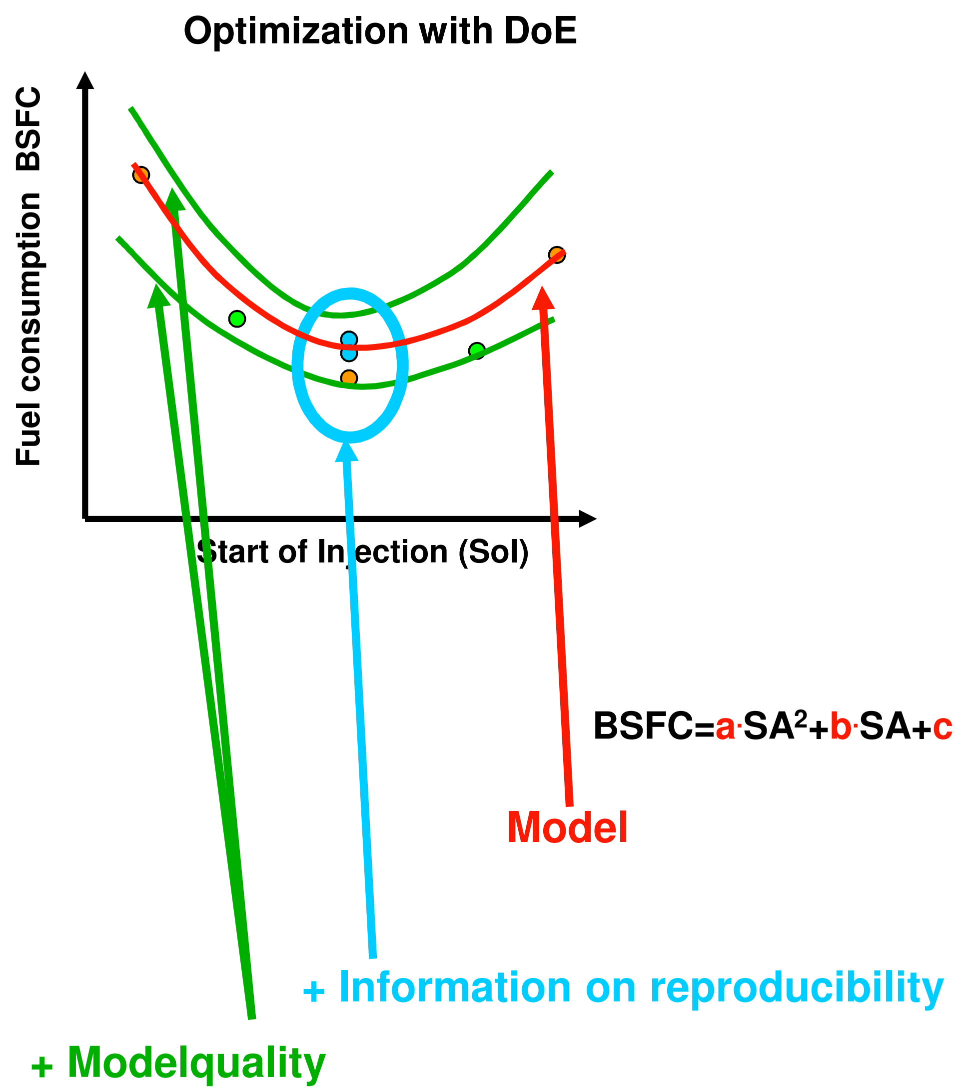
Lateral buckling of the behavior when varying both factors

= Interaction!

Design of Experiments – How?
Model based development and optimization (1 D-example)

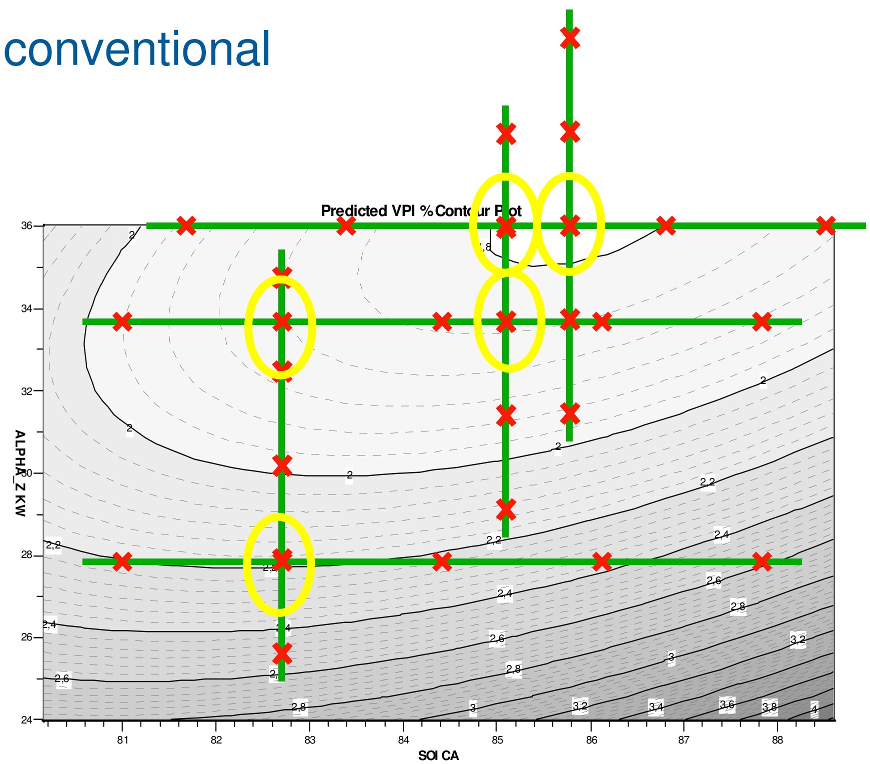




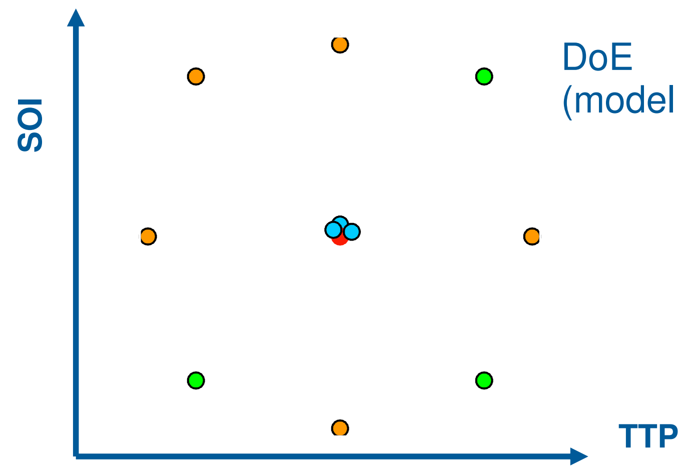


Design of Experiments – How?
Model based development and optimization (2 D-example)

conventional



25 .. 30 measurements



DoE
(model based)

CCD-Design, 2 Parameter: SOI und TTP

BSFC = K2 * TTP^2 + K1 * TTP +
K3 * SOI ^2 + K4 * SOI +
K5 * TTP * SOI
K0

6 measurements for coefficients
3 measurements for model quality
4 measurements for reproducibility testbed – engine

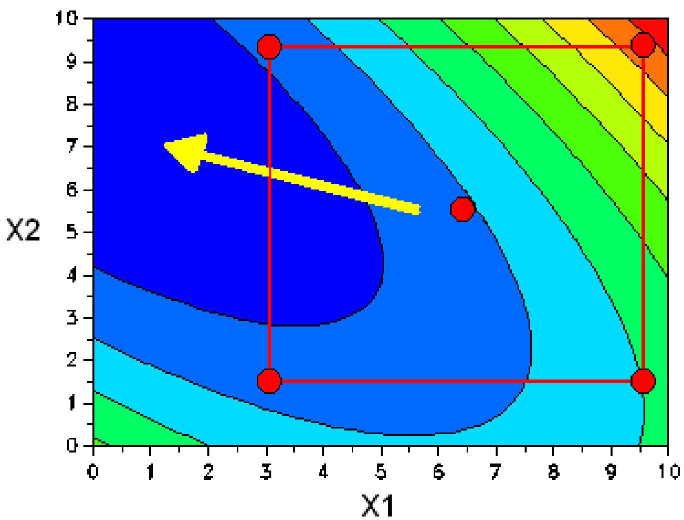
13 measurements

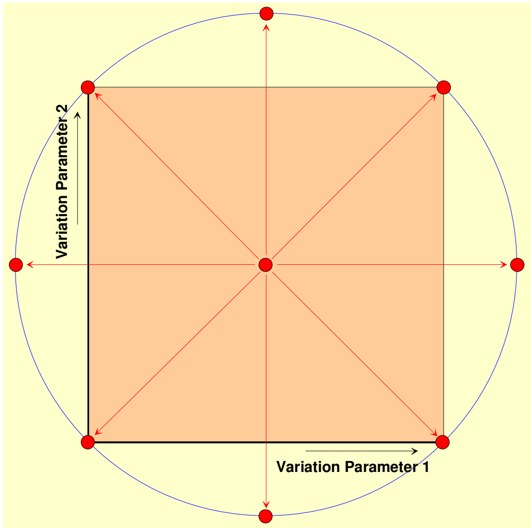
Design of Experiments – How?

Alternative to the conventional process

Create an experimental design, that varies all variables simultaneously and includes the interaction effects too.

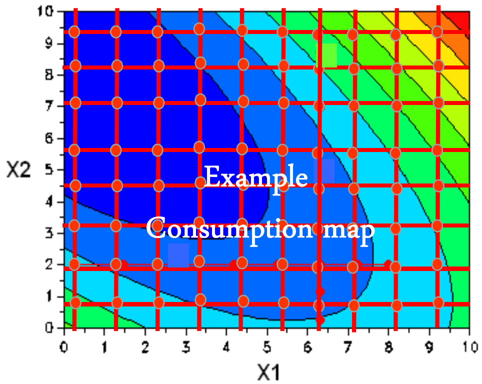
→ Maximum information for a low number of measurements





Comparison of conventional approach to DoE

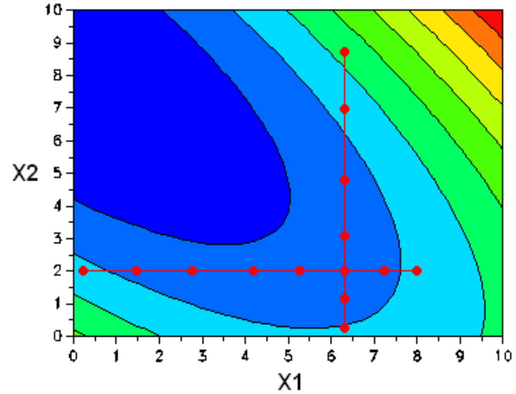
classical: “dragnet“ / „full factorial“ investigation



Advantage:
- complete design space
- model building possible
- almost exact fit
- drift sensitivity ?

Disadvantage:
- With more than 4 parameters, the required number of experiments is not applicable

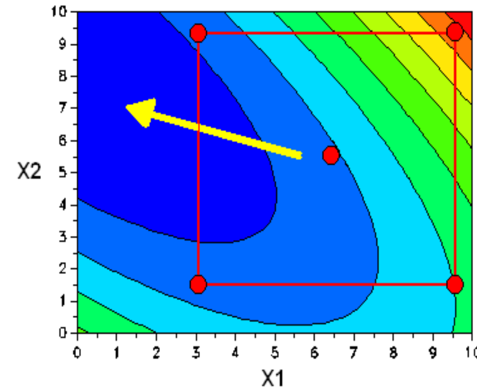
classical: parameter variation (OFAT)



Advantage:
- few measurement points

Disadvantage:
- different start points can lead to different optima
- no modeling possible
- design space incomplete
- drift sensitivity !
- no quantification of interaction

DoE

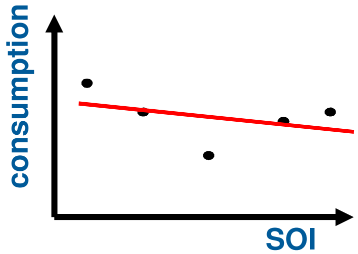


Advantage:
- complete design space
- few measurement points
- exact optimization result
- modeling possible
- no drift sensitivity !
- interactions covered

Disadvantage:
- learn the method once

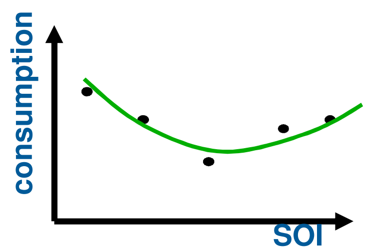
DoE goal: modelling! → Different Designs available

classic DoE´s = f(model equation)



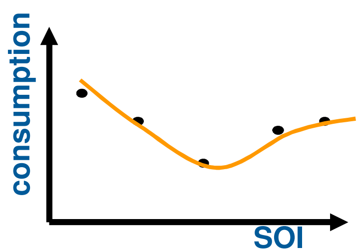
Linear:

y = k*x + d



Quadratic:

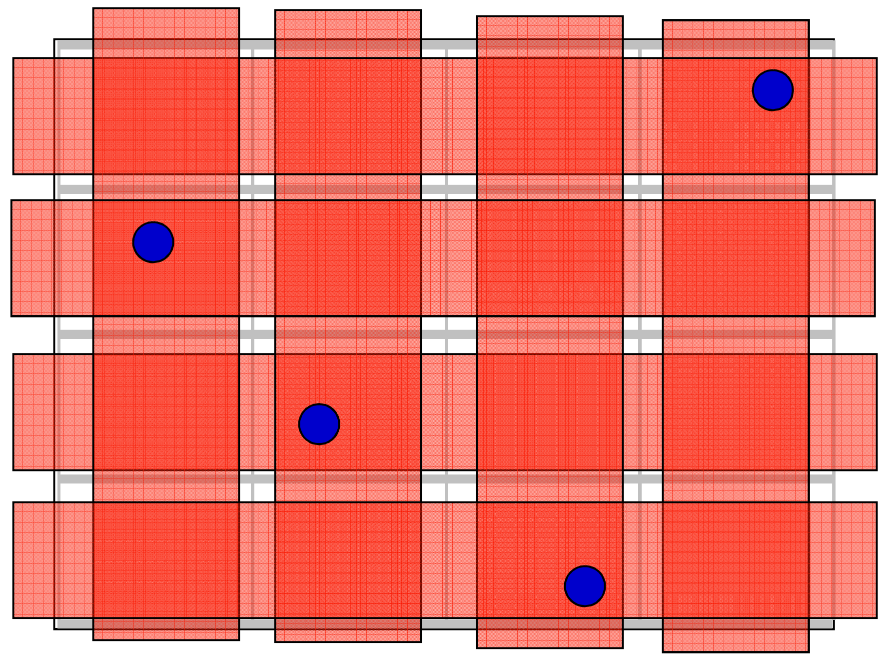
y = a*x^2 + k*x + d



Higher order:

y = b*x^3 + a*x^4 + k*x + d

or: just “space filling”:



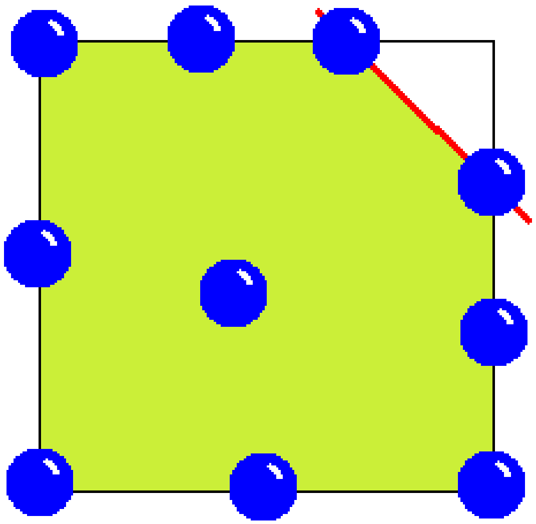
e.g.: Latin Hypercube - Design

Recommended (start-) designs for ICE-Tasks:

a) Preknowledge regarding nonlinearity of the responses for different directions available?

b) Not too strong constraining borderline maps?

→take a D-optimal Design! (most efficient)

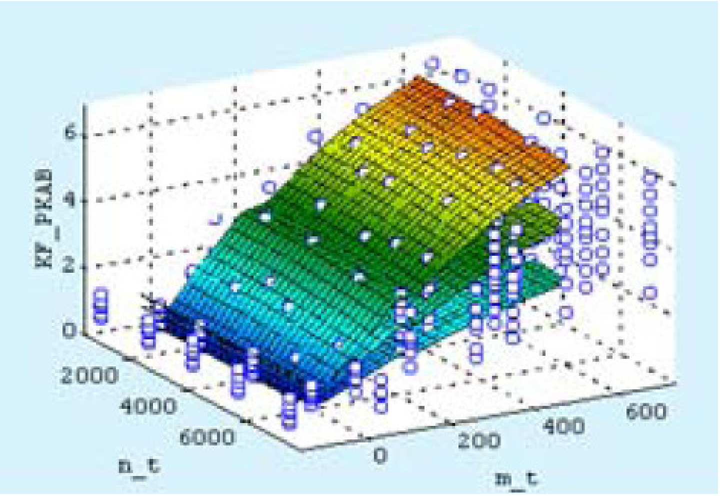


or:

a) ad a): no idea!

b) ad b): narrow borderline maps for global modeling

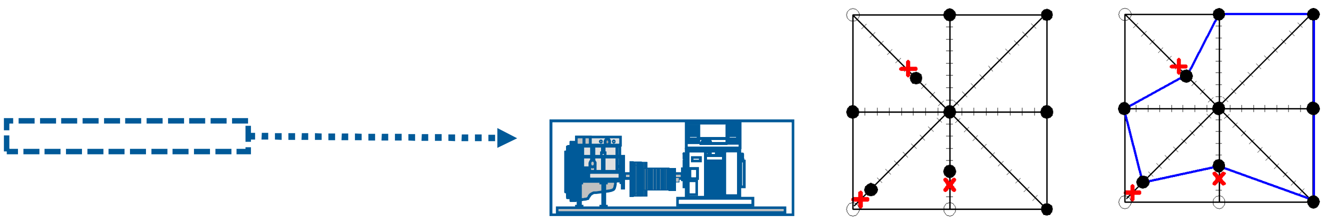
→Take S-optimal Design
(most efficient space filling design)



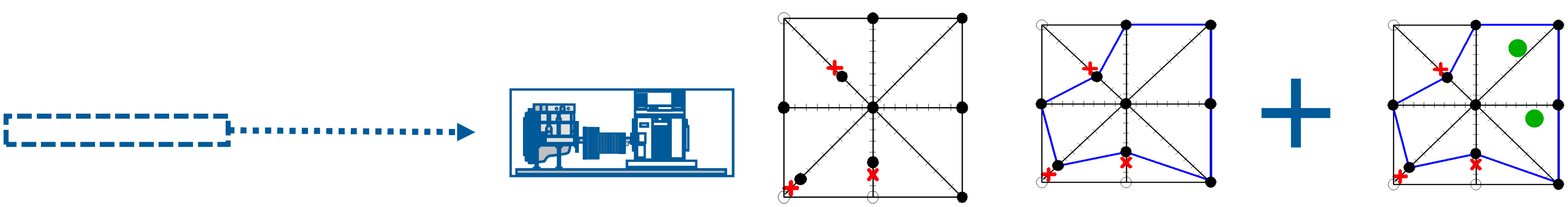
DoE on Internal Combustion Engines?
→ „It can destroy my engine !?!"

- **Variation list**

- **Online DoE screening (adjusts the design-space)**



- **Online D-optimal adaptive**



DoE on Internal Combustion Engines?
→ „It can destroy my engine !?!"

Test Template (iProcedure)

- **Variationlist**

  \+ fast
  \- Limits -> points are lost

- **Online DoE Screening**

  \+ Point number is kept
  \- Design is distorted

- **Adaptive Online DoE**

  1.) + Point number is kept
  \- Design is distorted
  2.) +adaptive Phase (D- / S-optimal)
  +Design is recalculated and adapted to the drivable range

Design (DoE)

- DOE Box Behnken

  \+ Polynomial models of second order
  \- Interaction not fully investigated

- DoE Central Composite

  \+ Polynomial models of second order (fully)
  \- symmetric design space needed

- **D-Optimal**

  \+ free polynomial order (depending on the direction)
  \+ freely shaped design space (by candidate set)
  \+ Inclusions possible
  \+ additional number of points definable
  \- Preknowledge regarding the task benefical

- Latin Hyper Cube

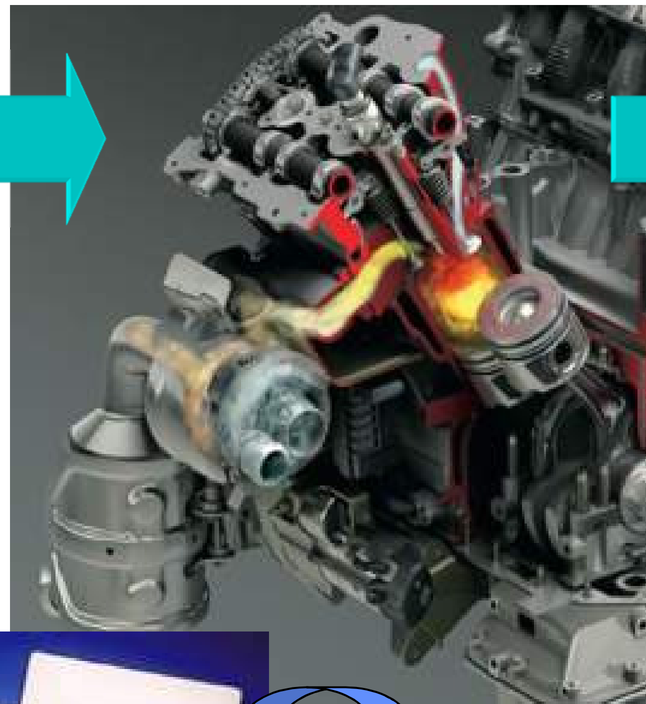
  \+ Just filling the space (defined by number of points)
  \- Outside rarely covered
  \- Symmetric designspace
  \- No direction specific differences possible

- **S-Optimal**

  \+ Fills the space – also in inclined houses
  \+ Asymmetric design space fully supported
  \- Preknowledge regarding the task benefical

Example based theory:
e.g.: typical task in R+D / Calibration:
Optimization of Injection and combustion:

Input Parameters:
High amount of variables

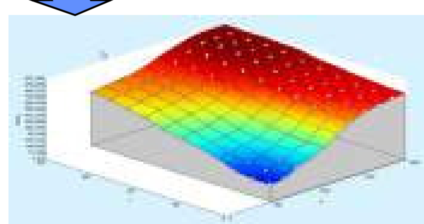


Target:
- **Min fuel consumption**
- **Emission limits**

Diesel:
- nozzle type
- Injection pressure
- Start of Injection
- EGR
- Boost pressure
- several injections

etc.

→ Hardware selection and
→ corresponding calibration!



Engine Maps = Application Label

Model based development and optimization:
“Injector selection with best calibration and trade off view”

Calibration Target Definition

⇩

Experimental Design

⇩

Testing

⇩

Data Analysis and Modeling

⇩

Optimization

⇩

Validation

Target:

To select the best injector (FC = min), that meets both power and emission requirement using trade off optimization

Variations in 4 Modal points:

- Rail Pressure
- SoI (Main Timing Phasing)
- EGR Valve position (Air mass)

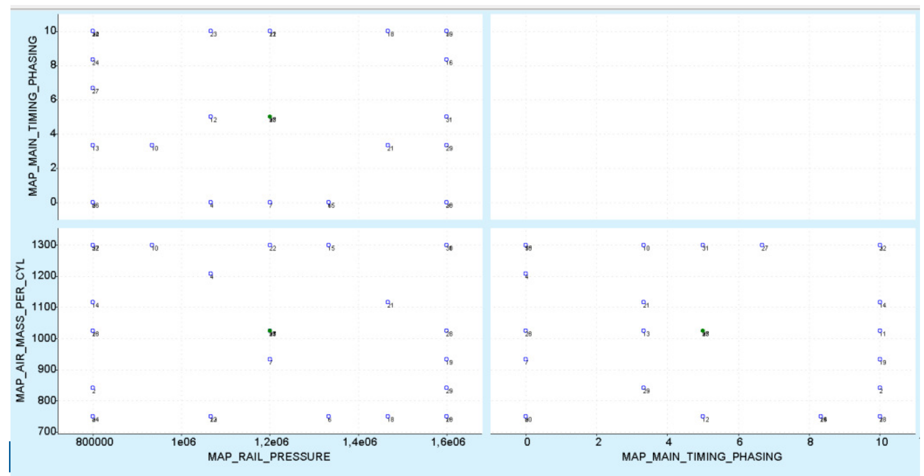
Measurements to be taken in stable conditions:

- **FC** (Fuel Consumption)
- **NOx** + HC Emissions
- **Soot** Emission, **Particle Number**
- **Noise**,
  Maximum cylinder pressure, MFB 50%

Model based development and optimization

- Calibration Target Definition
- ⇩
- **Experimental Design**
- ⇩
- Testing
- ⇩
- Data Analysis and Modeling
- ⇩
- Optimization
- ⇩
- Validation

- 4 Modal points out of 13 Mode-Test
- Variation of
  - Rail Pressure
  - Main Timing Phasing
  - EGR Valve position (Air mass)



- (27 measurements per Operating point)
- On line adaptation keeping the Engine limits for
  - **Maximum Cylinder pressure**
  - **Maximum Turbine inlet temperature**
  - **Maximum turbine speed**

Model based development and optimization

Calibration Target Definition

⇩

Experimental Design

⇩

Testing

⇩

Data Analysis and Modeling

⇩

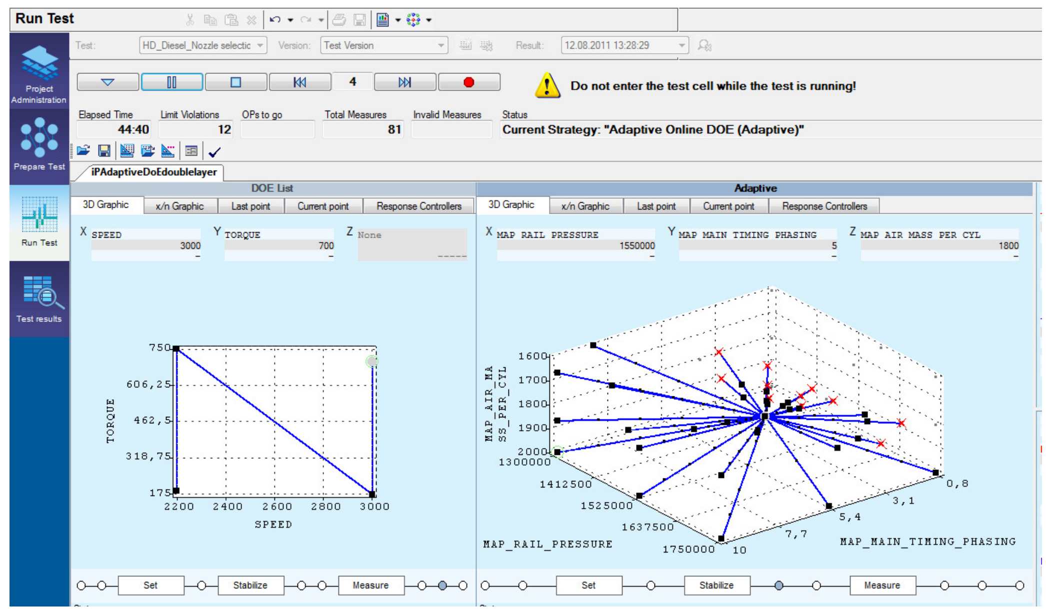
Optimization

⇩

Validation

The Method allows for testing:

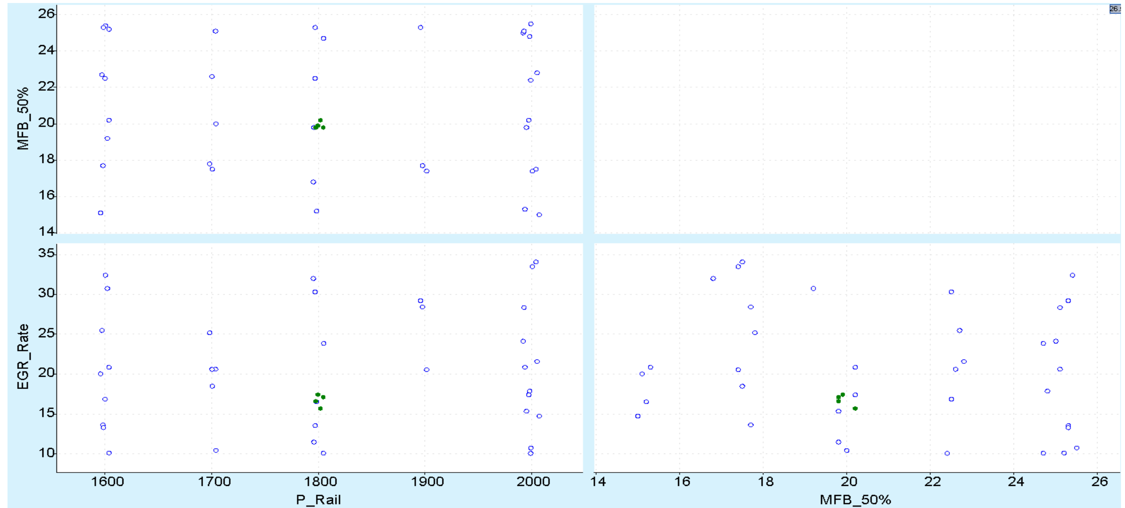
- Fully automated execution of test runs
- Fully automated limit reactions
- Fully automated adaptation of the DoE



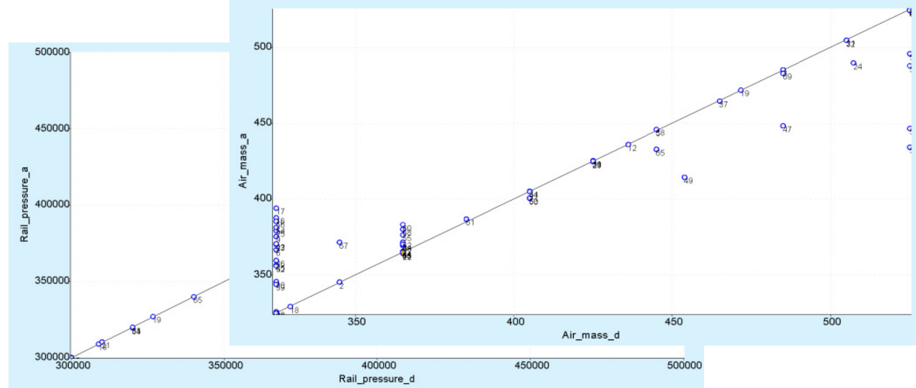
next theory part: → The raw data analysis

Important before starting the modeling of all required channels !

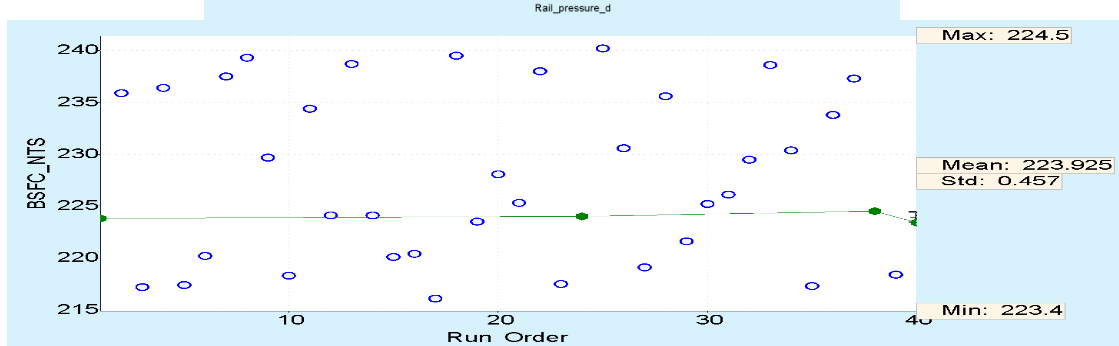
- **Check the DoE Design**
  → Variation vs. run order and
  → Variation vs. Variation



- **Check, if desired settings happened Compare demand values to actual values of the variation parameters**
  → Variation Demand vs. Actual

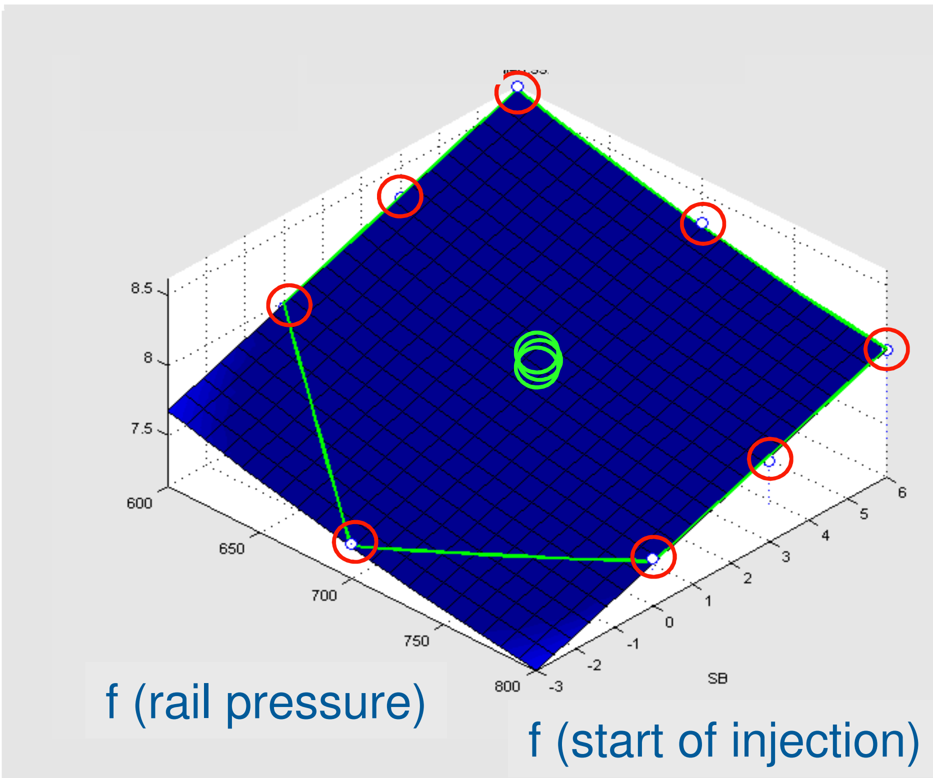


- **Find strong outliers in the measurements**
  → Measured vs. RunOrder
- **Check the Reproducibility of the boundary conditions**
  → Measured vs. RunOrder



Empirical mathematical modells
e.g.: Polynomial Models (= base for many other types)

Fuel consumption [kg/h]



measurement points

repetition points

e.g.: 2nd order model equation:

$z = a_o$ +

Constant

+ a_1*SB **+ b_1***Prail **+**

Linear terms (main directions)

+ a_2*SB^2 **+ b_2***$Prail^2$ **+**

Quadratic terms (main directions)

+ c*SB*Prail

Interaction term 2nd order

Model types

- **Polynomials**

  Model order: arbitrary

- **Free Poly Model (FPM)**

  Model order: arbitrary

  Automatic order reduction

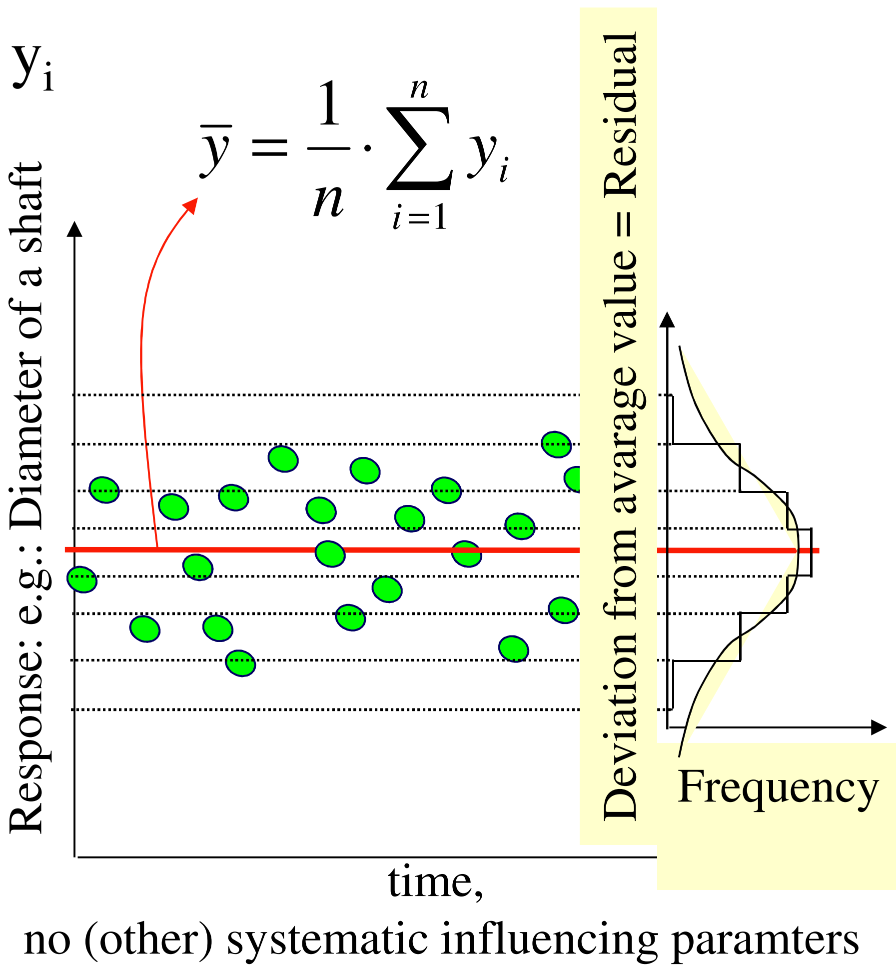
  Deletion of insignificant terms

- **FNN Fast Neural Network** / INN

  combination of several Polynomial models as Neural net work

- **Integration of custom model types**

How to judge model quality?
some statistic basics:



1) **Average value**

2) **Standard Deviation:**
To measure the average deviation between the average value and the individual measurements

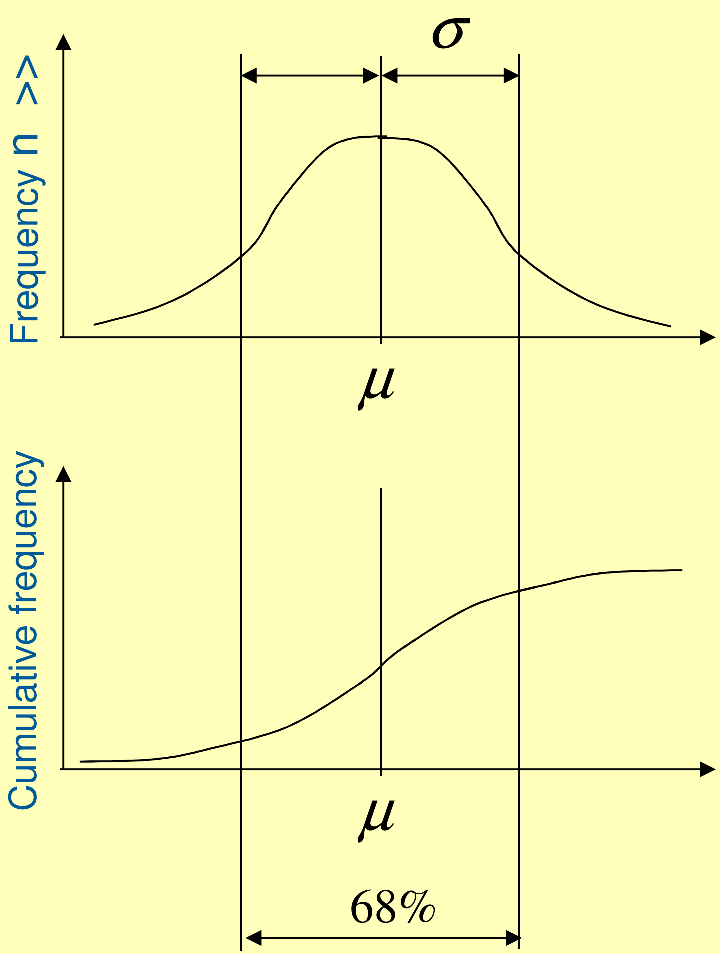
$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

3) Frequency distribution:

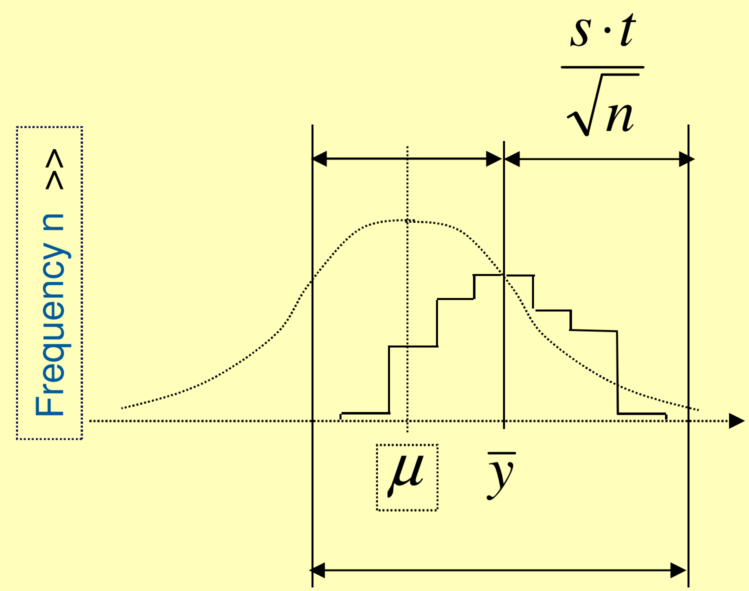
Define an area and count the number of values within the defined area

Truth ← → Sample

Truth: random distribution of individual results around the average value (normal distribution)



Measurable: n samples + s + histogram:



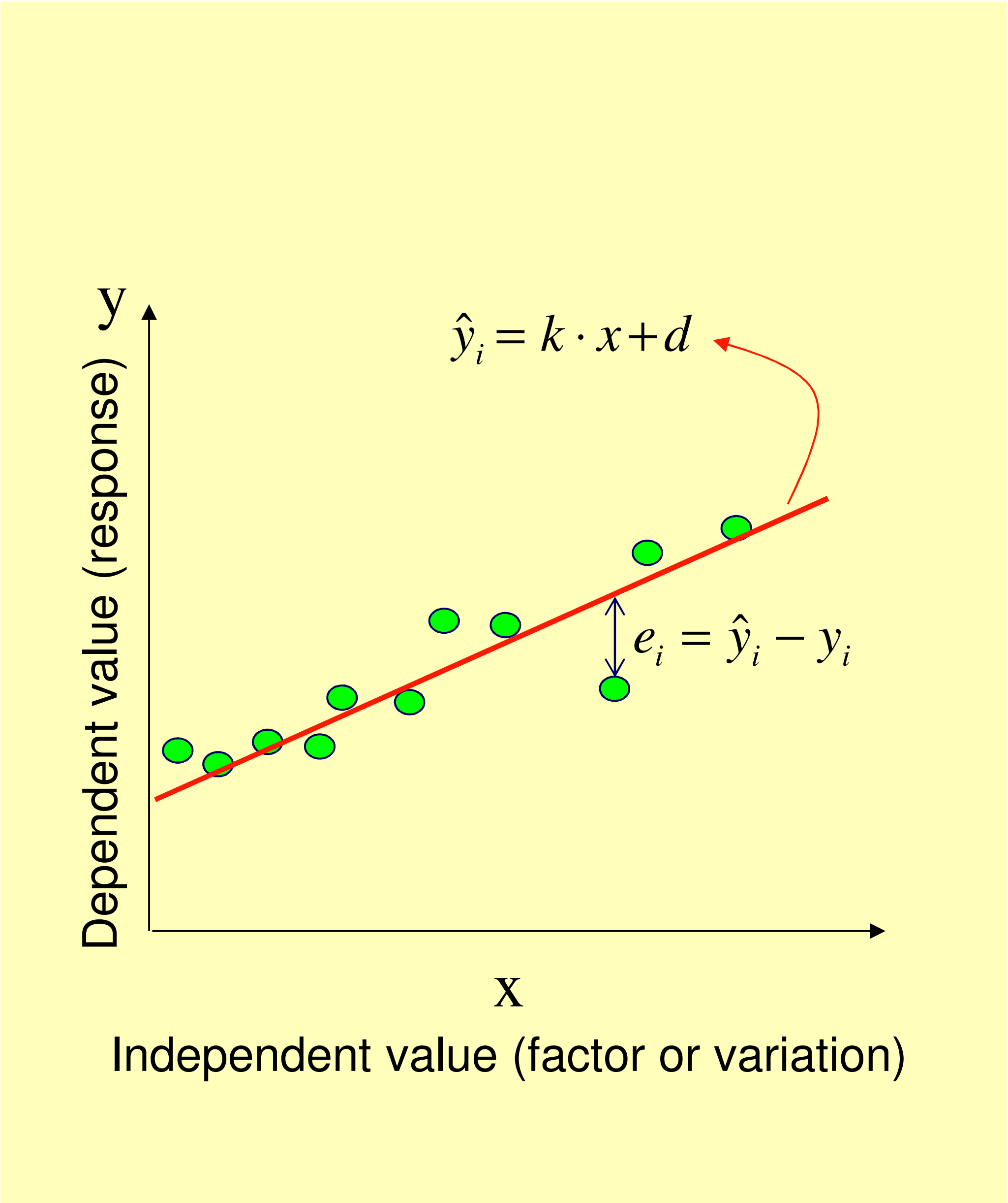
Confidence interval

The true value μ has, for example, a 95% probability of falling within

$$\bar{y} \pm \frac{s \cdot t_{(95\%,n)}}{\sqrt{n}}$$

s = Standard deviation; t = Student factor (table value) / t-distribution: normal equation derived from distribution. Can be interpreted as a signal-noise-ratio

In case of (linear) Regression:



Methode of least square fit:

Adapt the model coefficients (k, d) such, that:

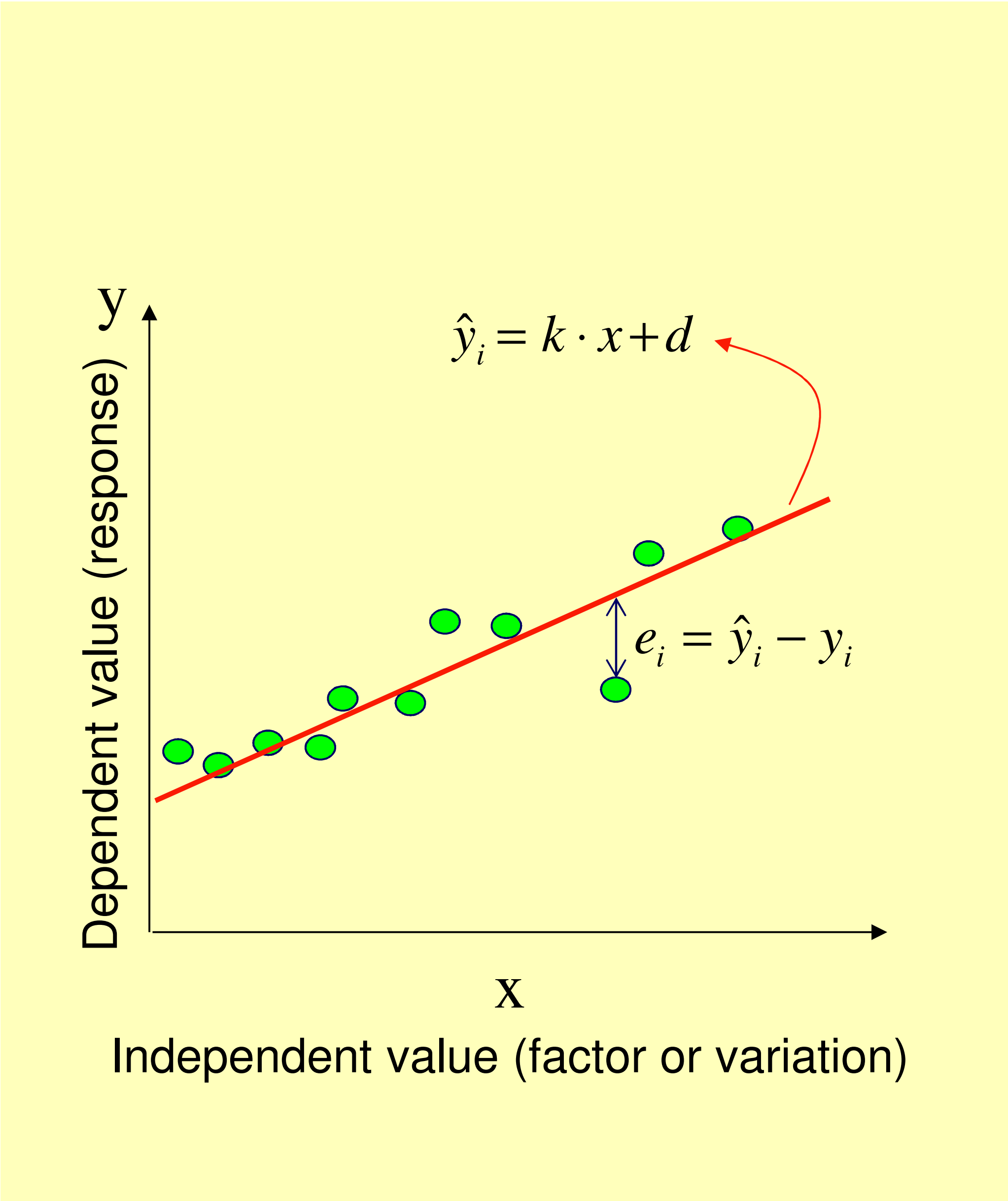
$$\sum_{i=1}^{n} e_i^{\,2} \Rightarrow Min$$

Residua: $e_i = \hat{y}_i - y_i$

$\hat{y}_i$ i*th* model value as a function of x (independent variable)

y_i i*th* measurement value as a function of the response at x_i

In case of (linear) Regression:
→ residua and confidence intervall are still usefull!



1) Regression coefficient

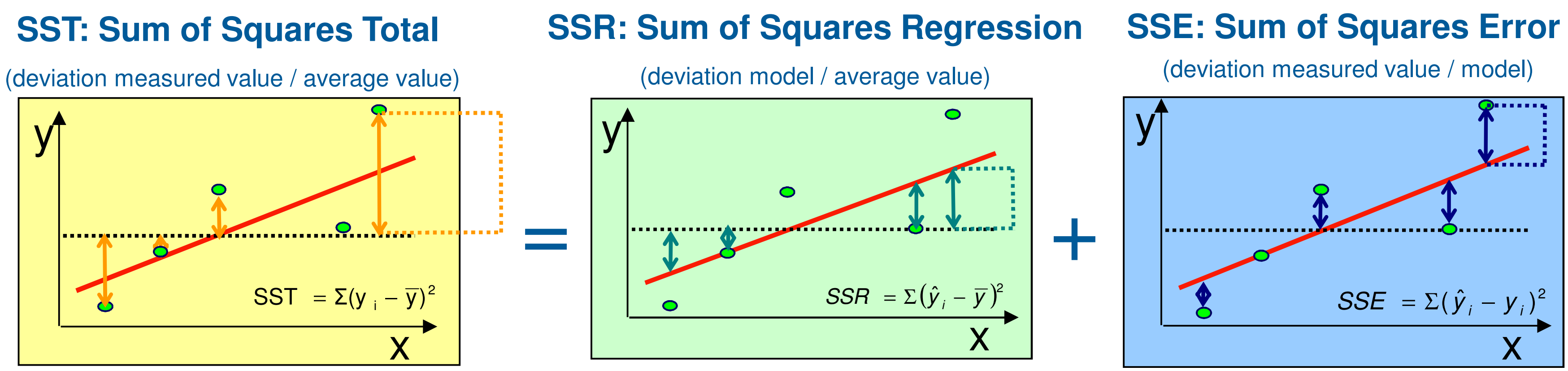
2) Standard deviation

3) Statistic analysis of the variance (ANOVA)

4) Confidence interval of the models

5) Are the base conditions for the results above fulfilled (Normal distribution of the residua)?

Splitting of the average value deviations



SST = SSR + SSE

y_i — i^{th} measured value

$\hat{y}_i$ — i^{th} modeled value

$\bar{y}$ — Average of all measured values

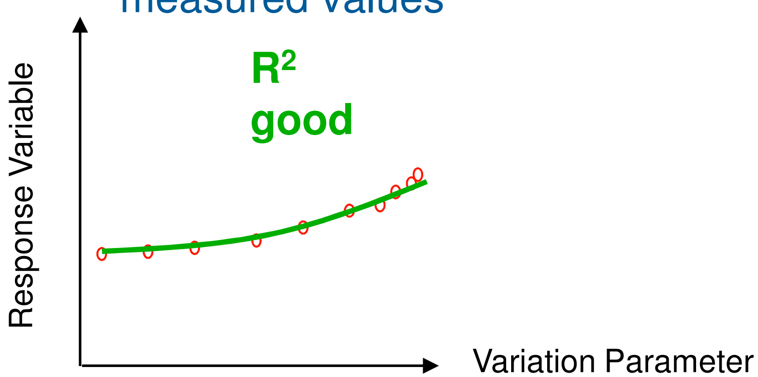
Regression Coefficient (Coefficient of Determination)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

y_i : i^{th} measured value

$\hat{y}_i$: i^{th} model value

$\bar{y}$: average of all measured values



SSR: Deviation of the **modeled** values from the total average value

SST: Deviation of the **measured** values from the total average value

Must be between 0 and +1

It shows how much the model explains the deviation from the average value.

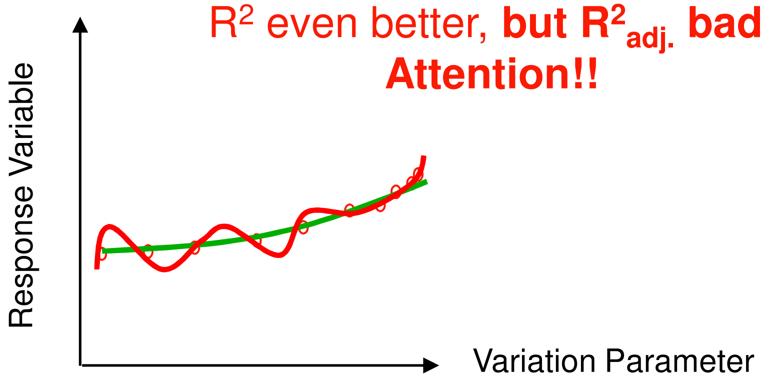
It shows how exactly the model matches to the measurement values.

Adjusted Coefficient of Determination

$$R^2 adj = 1 - \frac{SSE/(n-k)}{SST/(n-1)}$$

R^2 = Regression coefficient (coefficient of determination)
n = number of values
k = number of independent model-coefficients

R^2 even better, **but $R^2_{adj.}$ bad!**
Attention!!



"R^2 adjusted":

Ranges between $-\infty$ and $< R^2$

R^2_{adj} takes into account the model's degrees of freedom (n - k)

R^2_{adj} can decrease with increasing model order, due to reducing degrees of freedom, in cases where R^2 would indicate a more faithful model fit.

Predicted Coefficient of Determination

$$R^2 pred = 1 - \frac{PRESS}{SST}$$

SST: Deviation of the **measured** values from the total average value

PRESS: Sum Squares of the deviations of the **measured** values from the **modeled** values, where the respective measurement is **not** used for the model calculation (otherwise it is the same as SSE).

PRESS: Predictive **Re**sidual **S**um of **S**quares

"R^2 predicted":

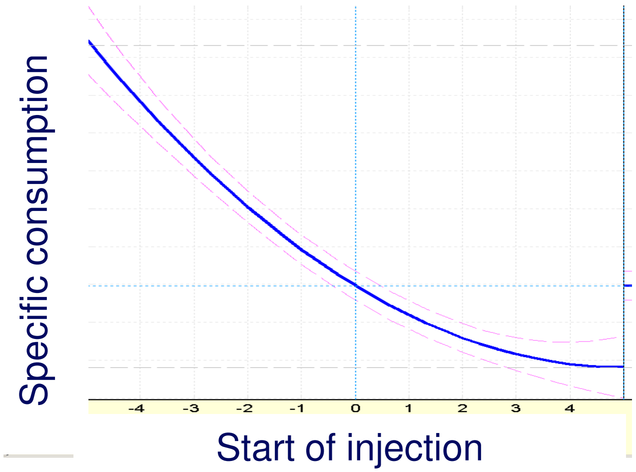
Ranges between - ∞ and < R^2

R^2predicted describes the model's predictive power.

We treat the j^{th} measurement as unavailable for modeling, however this measurement is used for calculating the j^{th} residual.

PRESS is the sum of the squares of residuals calculated in this way.

Confidence Interval



$$Conf = \hat{y}_i \pm t \cdot \sqrt{D_i}$$

$\hat{y}_i$: model value in place i

t : Studentfactor

D_i : local quantile

Shows the boundaries of the range within which the “true model” is valid with a confidence of, for example, 95% .

Shows whether the model value change, as a function of some variation parameter, is significant or not.

In CAMEO, the confidence interval can be set to 90%, 95% or 99%.

Prediction Interval

$$Pred = \hat{y}_i \pm t \cdot \sqrt{(D_i + s^2)}$$

$\hat{y}_i$: model value in place i

t : Studentfactor

D_i : local quantile

s : standard deviation

Shows the boundaries of the range within which results are expected to lie, with a probability of, for example, 95%, if the experiment is repeated.

Shows whether a verification measurement can be expected within this range of the model.

Does a regression explain more than an average value?
→ ANOVA - Analysis of Variance „F-Test“

AVL

	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-Value (F)	Significance (p)
Regression	665	6	110.0	559.8	0.00001
Error	12	64	0.19		
Total	677	70			

SSR

SSE

SST

Number of measurements n

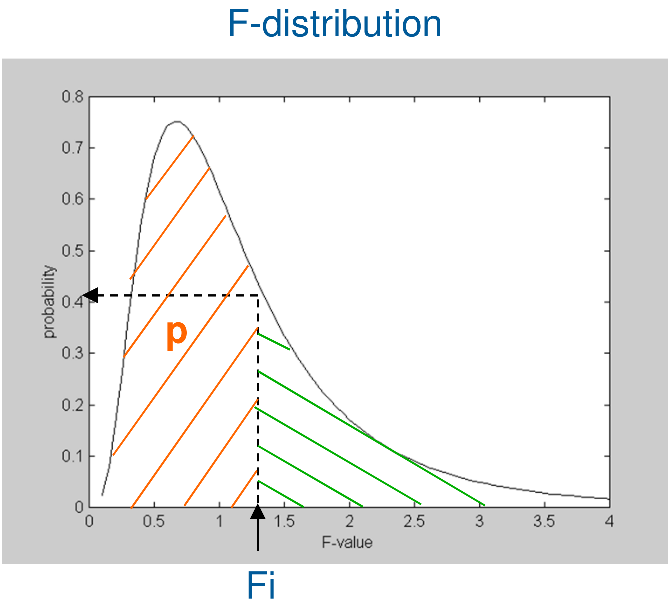
Number of Model-koeffizients k

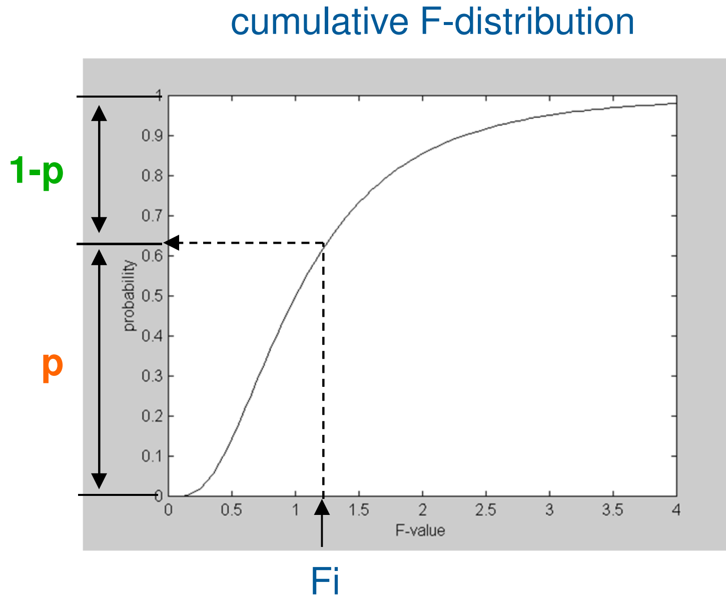
$$F = \frac{MS_{regression}}{MS_{error}}$$

Mean Square: $MS_{regression} = \frac{SSR}{df_{regression}}$ or $MS_{error} = \frac{SSE}{df_{error}}$

Analysis of variance – F-destribution (Fisher-Snedecor- distribution)

1) Define a Hypothesis to be checked (e.g.: "one variance is significantly bigger than the other")
2) Determine the Probability for an F-value in case the Hypothesis is true (or wrong)





p .... Probability of that F_i-value in case the Hypothesis is **true**

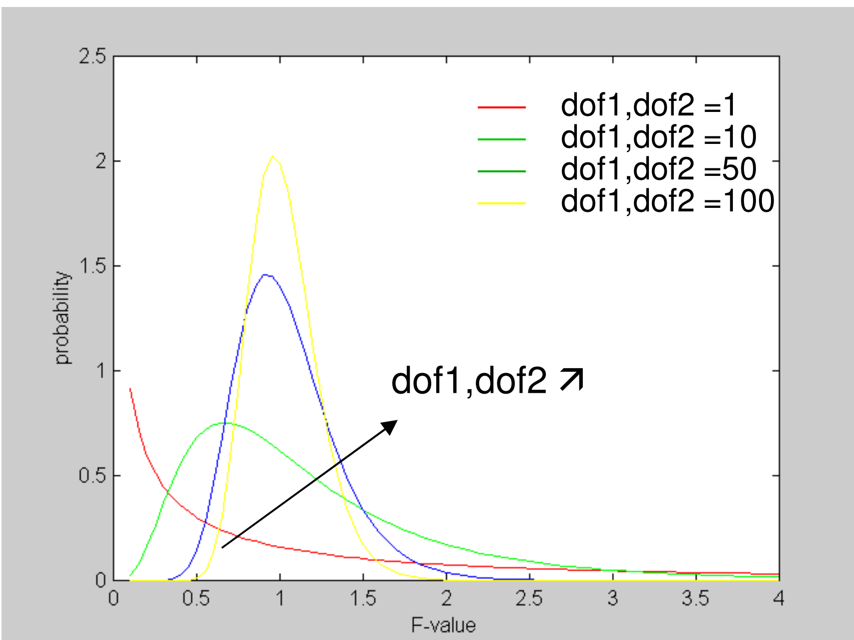
1-p .... Probability of that F_i-value in case the Hypothesis is **wrong**

Varianzanalysis and F-distributions

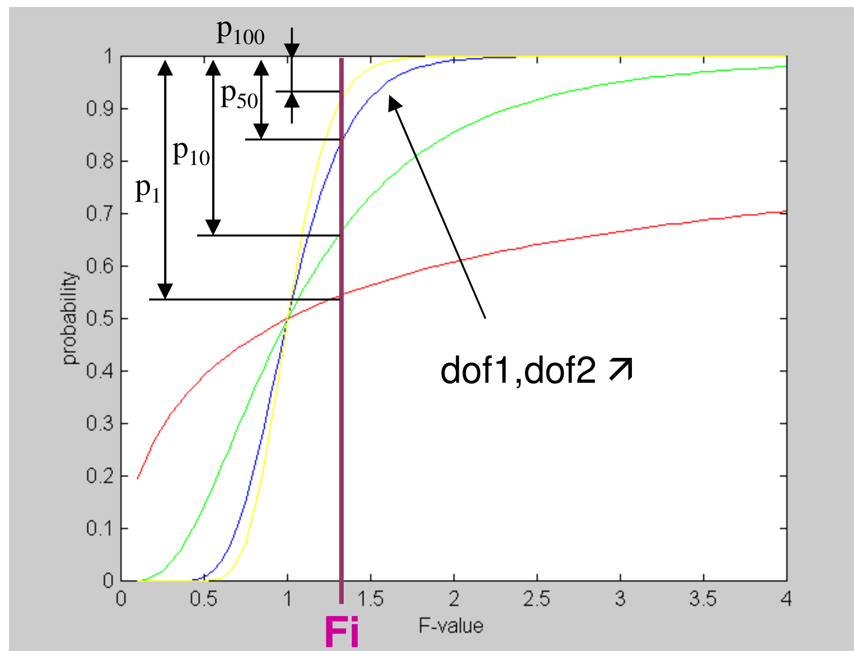
$$p = f(Fi, dof1, dof2)$$

The shape of the F-distribution strongly depends on the degrees of freedom of the compared variances:

F-distribution



cumulative F-distribution



So the „Significance“ shows in case of the F-Test the remaining probabillity, that a regression could not give more information than a poor meanvalue ($MS_{Regression}$ not bigger than MS_{Error})

	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	F-Value (F)	Significance (p)
Regression	665	6	110.0	559.8	0.00001
Error	12	64	0.19		
Total	677	70			

SSR

SSE

SST

Anzahl der Messwerte n

Anzahl der Modell-koeffizienten k

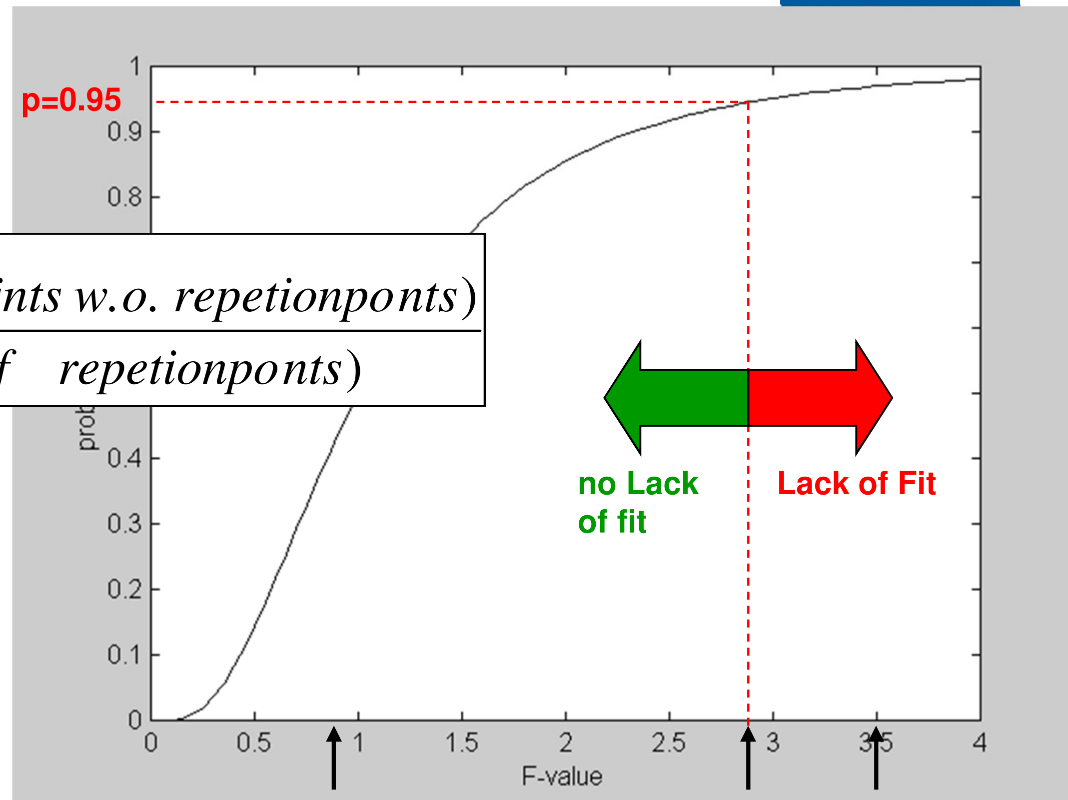
$$F = \frac{MS_{regression}}{MS_{error}}$$

Mean Square: $MS_{regression} = \frac{SSR}{df_{regression}}$ or $MS_{error} = \frac{SSE}{df_{error}}$

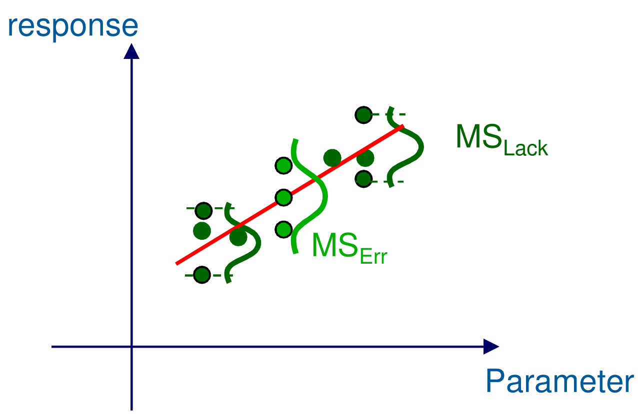
2^{nd} Signifikance Test: Lack of Fit?

Check the Hypothesis:
Varianz of other pionts arround the model >>
>>Varianz of the repetition failure ??

$$F2 = \frac{MS_{Lack}(dof1 = number\ of\ \ pojnts\ w.o.\ repetionponts)}{MS_{Err}(dof2 = number\ of\ \ repetionponts)}$$



F-Distribution: $p = f(Fi, dof1, dof2)$

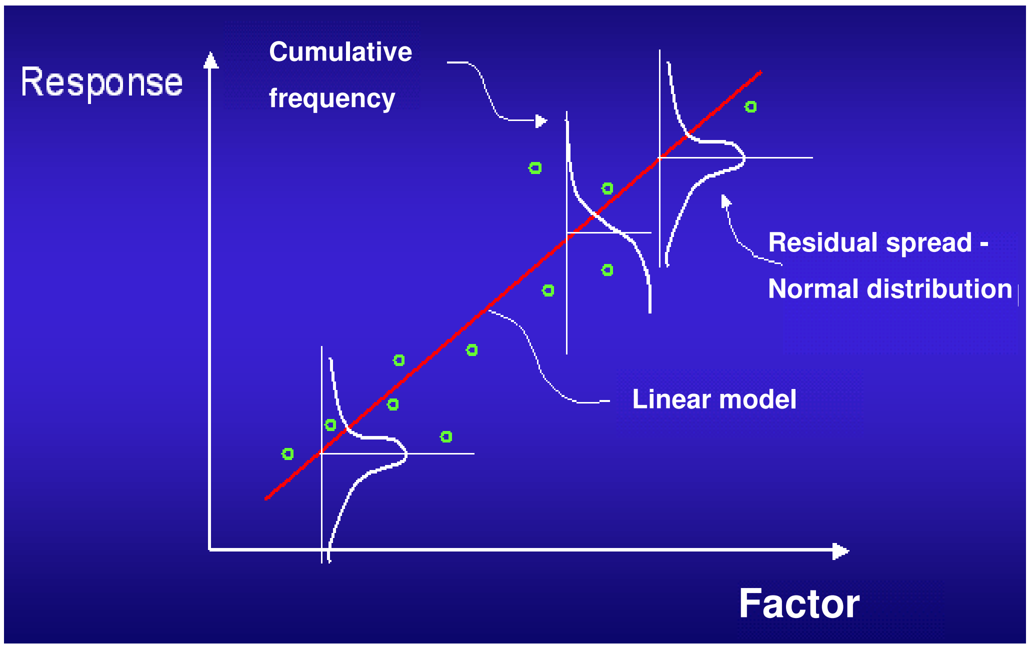


F2=3.5 … Hypothesis is true
$\rightarrow$ Varianz of points arround the model is significantly bigger than in the repetition point → “Lack of Fit”

F_{95} > F2 > F_{50} …Hypothese is probably not true
$\rightarrow$ Varianzes are in the same range → no “Lack of Fit” (optimal Model)

F2=0.8 … Danger of “overfit”

Precondition of normal distribution of the residuals fulfilled?



Small residuals occur more often than large ones, i.e. normal distribution of residuals

→ Model fits to the average behavior!

→ No trend (random errors are independent of each other and no function of time)

Constant standard deviation of the residuals (independent of x)

Summary Statistic base concepts

Parameter	Range	Meaning	Excellent	Good	Average
R^2	0 to 1	Quality of model fits to measurements, ratio of modeled to total deviation from the average value	≤≥ 0.95	≥ 0.8	≥ 0.5
R^2adj	- ∞ to 1	Adjusted to the number of degrees of freedom – the more coefficients (higher model order) and fewer measurements, the lower the statistic.	≥ 0.95	≥ 0.8	≥ 0.5
R^2pred	- ∞ to 1	Model predictive quality for new measured values (be careful with few measurements)	≥ 0.9	≥ 0.7	≥ 0.4
Confidence interval		Range in which the true model value lies with a 95% probability	Intersection Plot: relative to measured range x/y Plot: Color code – 100% equates to measurement range		
Prediction range		Range in which a single new measurement is expected to lie, with a probability of 95%	Intersection Plot: relative to Measured range x/y Plot: Color code – 100% equates to measurement range		

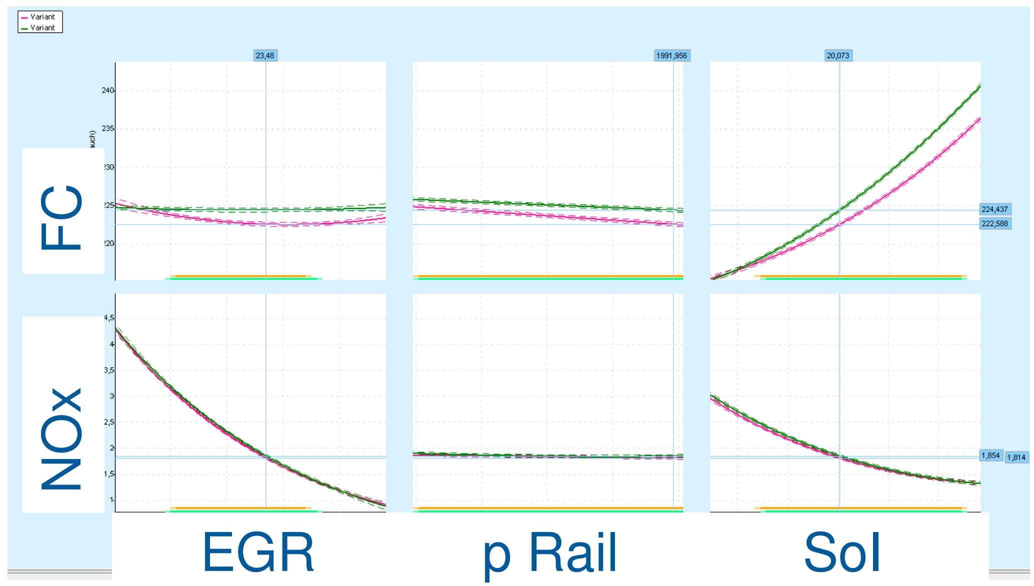
Summary Statistic base concepts

F-test	$MS_{Regression}$ significantly bigger than MS_{Error}?	
Lack of Fit	MS_{Lack} (other measurement points) > ? ?> MS_{Error} (repetition points) ?	Visible in Measured vs. Predicted - Graphic
Leverage	Are there single measurement points with a high influence on the model?	

Model based development and optimization

Calibration Target Definition
⇩
Experimental Design
⇩
Testing
⇩
Data Analysis and Modeling
⇩
Optimization
⇩
Validation

- Modelling of all relevant target channels with Polynomial Models or Neural Networks
- Intersection Graphics to (manually) optimize and understand interactions
- compare variants (injector behaviours)



Model based development and optimization

Calibration Target Definition

⇩

Experimental Design

⇩

Testing

⇩

Data Analysis and Modeling

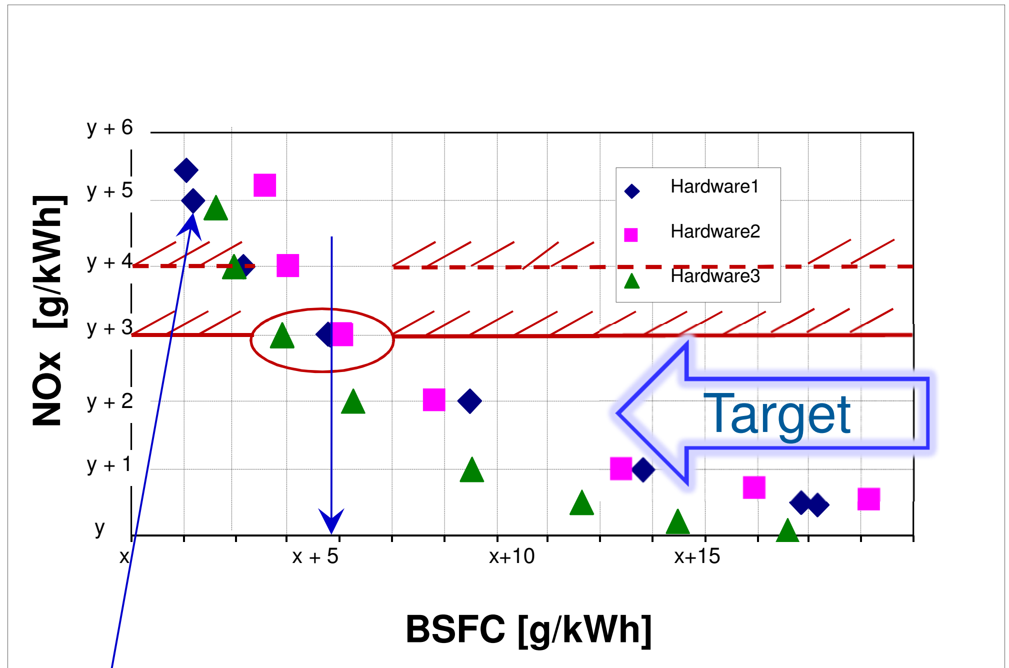
⇩

Optimization

⇩

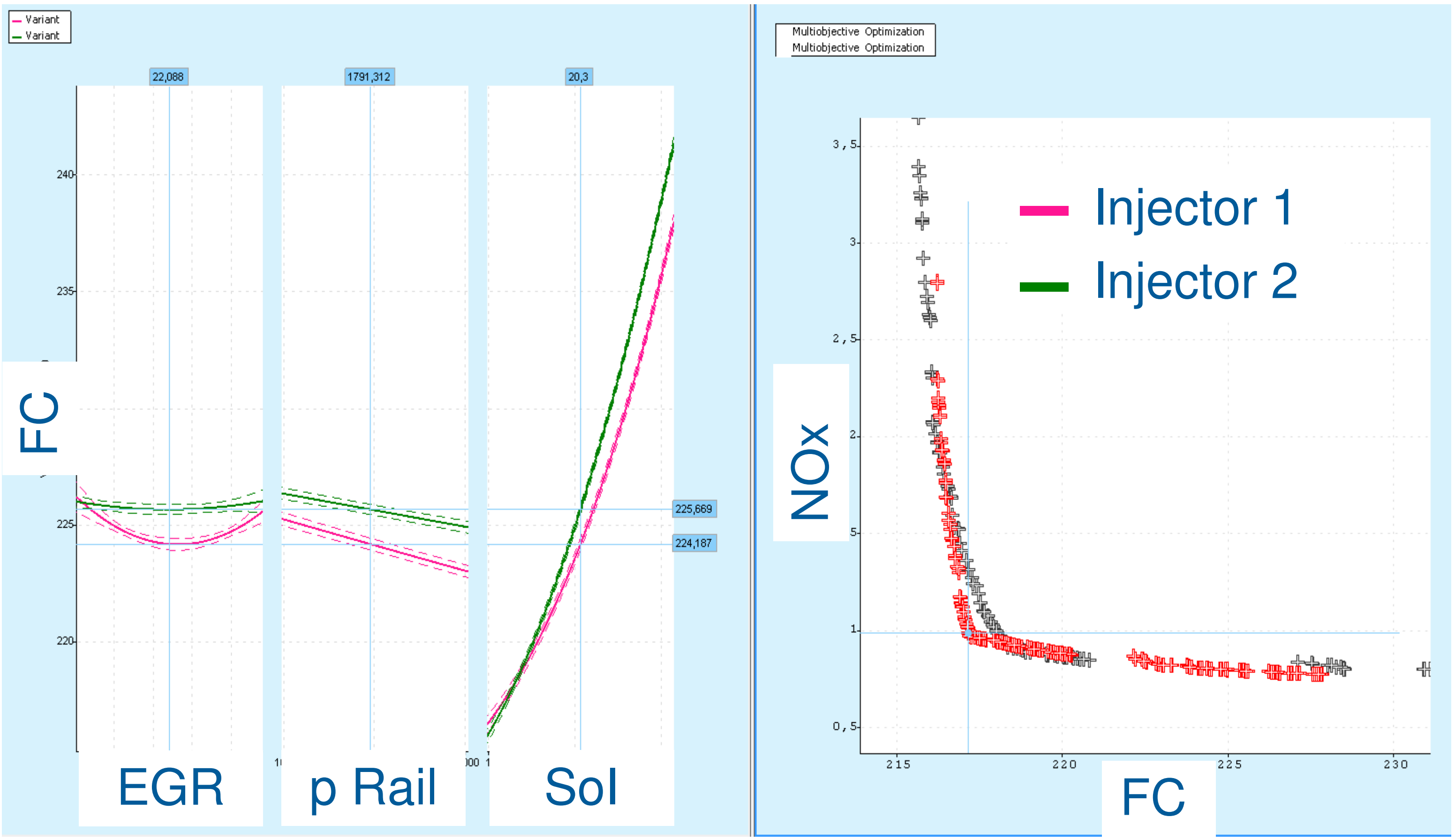
Validation

- How much FC to spend in order to reach the Nox-Traget?
- Which injector gives best FC weighted over the 4 modal points – with his individual best calibration?



Pareto Front: showing the Trade Off behavior

Model based development and optimization: compare 2 Pareto fronts of two hardware variants:



Model based development and optimization

Calibration Target Definition

⇩

Experimental Design

⇩

Testing

⇩

Data Analysis and Modeling

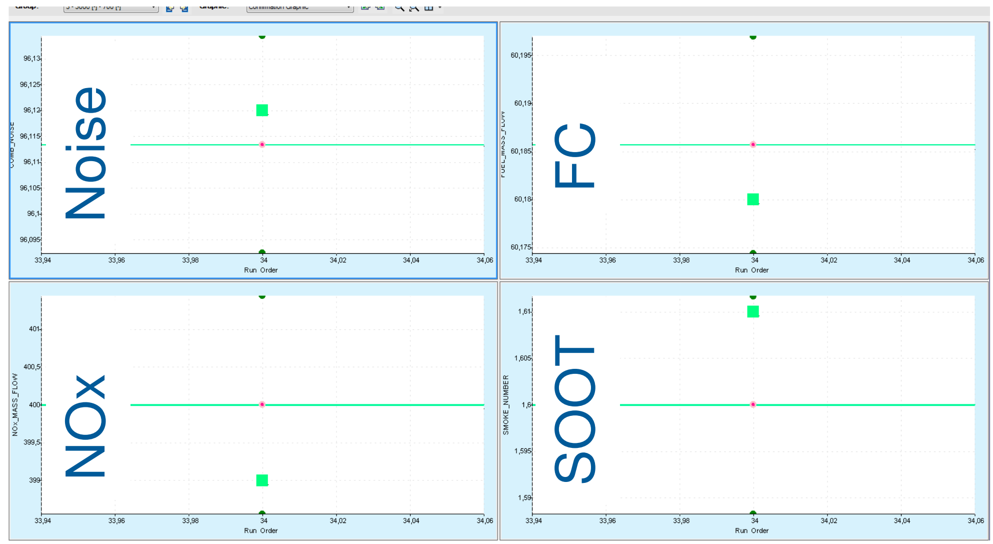
⇩

Optimization

⇩

Validation

- Where the models accurate in the area of the selected optimum?
- Check after the verification test run: measured value should be within the Prediction interval of the models



- All relevant verification measurements with in the prediction intervalls of the models → the whole process performed in trustable way

5 reasons for DoE

→ Strongly reduced number of measurements (maximum information, minimal effort)

→ Noise identification: discrimination between noise and a real effect in the response factor; observable with DoE (confidence level, measurement system stability) – difficult with conventional approach

→ Tested range will be modeled completely (good predictability over the complete range)

→ Results are reproducible and documented

→ Better insight into the variable-interactions (rapid improvement in expertise)

Boundary conditions for DoE

→ Requires testbed systems of higher quality and stability

→ Fundamental knowledge of relationship between parameters and target function; otherwise higher effort

→ Higher automation level is helpful