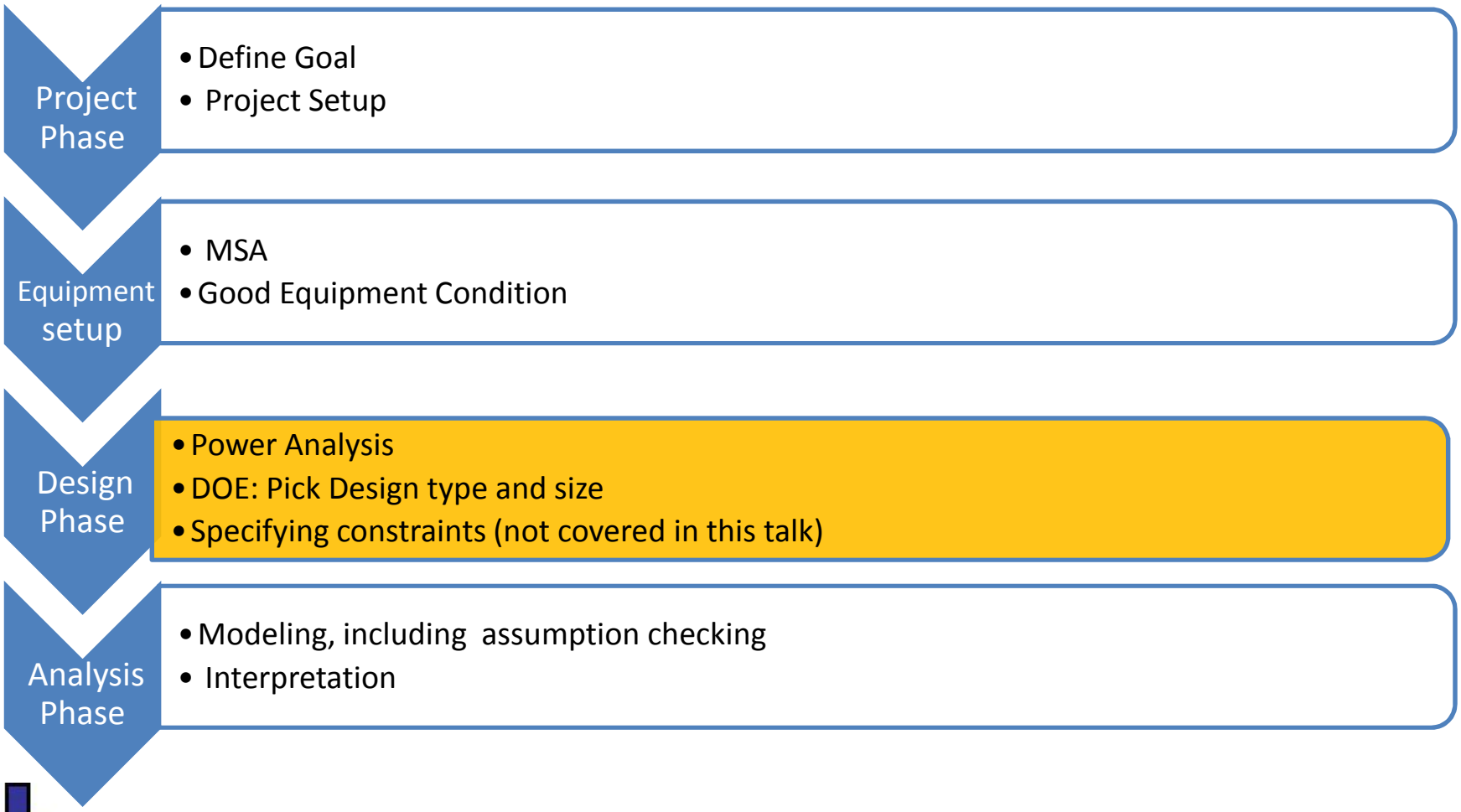
Common Mistakes in Designing Experiments

(process factors only)

StatCon Bertram Schäfer

STATCON

The Design Phase



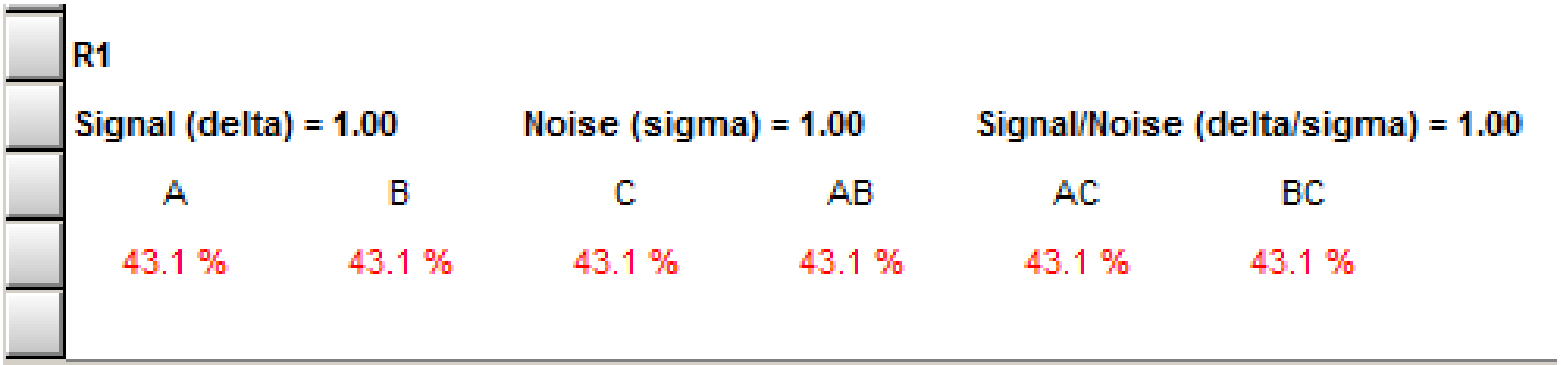
STATCON

Lack of Power Reasons

- Budget Restrictions
- Time Restrictions
- simplified training = minimal design
- no idea of experimental / measurement error
- lack of statistical knowledge

Prospective Power
Uni factorial approach

- Input is minimum effect size
- scenario: α =0.05, β = 0.8, 3 factors, 2 levels each, 2FI model, $\Delta\mu$=1, N(0,1), 2 reps



- -> power=43.1%

STATCON

Prospective Power
Multi factorial approach

- Inputs hypothesized individual averages
- scenario: α =0.05, β = 0.8, 3 factors, 2 levels each, 2FI model, N(0,1), 2 reps

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Response 1 Untitled
-1.00	-1.00	-1.00	0.5
1.00	-1.00	-1.00	-1.5
-1.00	1.00	-1.00	-0.5
1.00	1.00	-1.00	1.5
-1.00	-1.00	1.00	-0.5
1.00	-1.00	1.00	1.5
-1.00	1.00	1.00	0.5
1.00	1.00	1.00	1.5

- Results:
  - All effect sizes $\Delta\mu$=1 -> power >43.1%
  - Strong interaction $\Delta\mu$=2 -> $power_{(\Delta\mu=1)}$ = 43.1% & $power_{(\Delta\mu=2)}$ = 94.4%
  - Weak interaction $\Delta\mu$=.5 -> $power_{(\Delta\mu=1)}$ = 43.1% & $power_{(\Delta\mu=0.5)}$ = 14.4%

STATCON

Lack of Power Effect

- individual Important factors not significant
  - Often not recognized
  - In case of borderline effects
    - Increase significance level to at least 0.1
    - retrospective poweranalysis proven to be useless!
- all effects nonsignificant
  - simple to recognize
  - Post-hoc Poweranalysis useless anyway
  - Design new experiment with sufficient Power

STATCON

Wrong design type

- starting with RSM design in screening phase
  - Often more runs than necessary (assuming small # factors)
  - Utilizes smaller number of factors (assuming higher # factors describe the problem)
- Using two level factorial in RSM optimization phase
  - Curvature effects not covered
  - Optimum either in a corner of the design space or at center point
  - Interaction effects covered properly only when resolution is greater or equal 5

STATCON

Wrong number of factors in screening

(assuming there will be an RSM design run after the screening)

- Too many factors
  - No problem for the design phase. Even in classical designs up to 31 factors, assuming Res III designs. Higher number of factors through supersaturated designs.
  - Many non significant factors drop out in the analysis
- Too few or missing factors
  - The missing factor is assumed to switch from one level to another after some runs
  - the missing factor canceling out a design factor
  - the missing factor emphasizes a design factor

STATCON

Too few or missing factors
Cancelling out – Example

the simulation model+1+A+B+C-2D+AB+AC-AD

- Using a 2^3 full factorial randomized design adding a fourth factor D which changes it‘s level after the 5 run
- Factor D is deleted for the analysis
- The amount of cancelling depends on the randomization and on the position of change in the design (i.e. 5th run)

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Factor 4 D:D	Response 1 Untitled
1.00	1.00	1.00	-1	9
-1.00	-1.00	-1.00	-1.00	1
1.00	-1.00	-1.00	-1.00	1
-1.00	-1.00	1.00	-1.00	1
-1.00	-1.00	-1.00	-1.00	1
1.00	-1.00	-1.00	1.00	-5
-1.00	1.00	1.00	1.00	-1
1.00	-1.00	1.00	1.00	-1
-1.00	1.00	-1.00	1.00	-1
1.00	1.00	-1.00	1.00	-1
-1.00	-1.00	1.00	1.00	-1
1.00	-1.00	1.00	1.00	-1
1.00	1.00	1.00	1.00	3
-1.00	1.00	1.00	1.00	-1
1.00	1.00	-1.00	1.00	-1
-1.00	1.00	-1.00	1.00	-1

STATCON

Too few or missing factors emphasizing – Example

- the simulation model+1+A+B+C+2D+AB+AC+AD
- Using a 2^3 full factorial randomized design adding a fourth factor D which changes it‘s level after the 9 run
- Factor D is deleted for the analysis
- The amount of emphasizing depends on the randomizatior and on the position of change in the design (i.e. 9th run)

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Factor 4 D:D	Response 1 Untitled
-1.00	-1.00	1.00	1.00	1
1.00	1.00	1.00	1.00	9
-1.00	-1.00	1.00	1.00	1
-1.00	1.00	-1.00	1.00	1
1.00	-1.00	1.00	1.00	5
1.00	1.00	-1.00	1.00	5
-1.00	1.00	1.00	1.00	1
1.00	-1.00	1.00	1.00	5
1.00	1.00	1.00	1.00	9
1.00	-1.00	-1.00	-1	-5
1.00	-1.00	-1.00	-1.00	-5
1.00	1.00	-1.00	-1.00	-1
-1.00	-1.00	-1.00	-1.00	-1
-1.00	1.00	1.00	-1.00	-1
-1.00	1.00	-1.00	-1.00	-1
-1.00	-1.00	-1.00	-1.00	-1

STATCON

Wrong number of factors in RSM

- Too many factors
  - many runs
- Too few or missing factors
  - the forgotten factor canceling out or emphasizing a design factor is rare
  - Missing the “true” optimum

STATCON

Improper settings of factor levels

- All factors wide (full design space)
- Single factor narrow
- Single factor shifted (high or low) narrow
- All factors narrow (conservative design space)
- Two factors Shifted (high or low) narrow

The „true“ simulation model

$$25 \cdot Sine(A) + 2 \cdot A - 12 \cdot B + AC + 3 \cdot BC - 3.5 \cdot (C-4)^2$$

- Reasonable function, not too bumpy, but still showing two maxima, interaction and squared behavior (highly subjective)
- A,B,C all ranging between 1 and 9
- Used to simulate response values without additional random noise

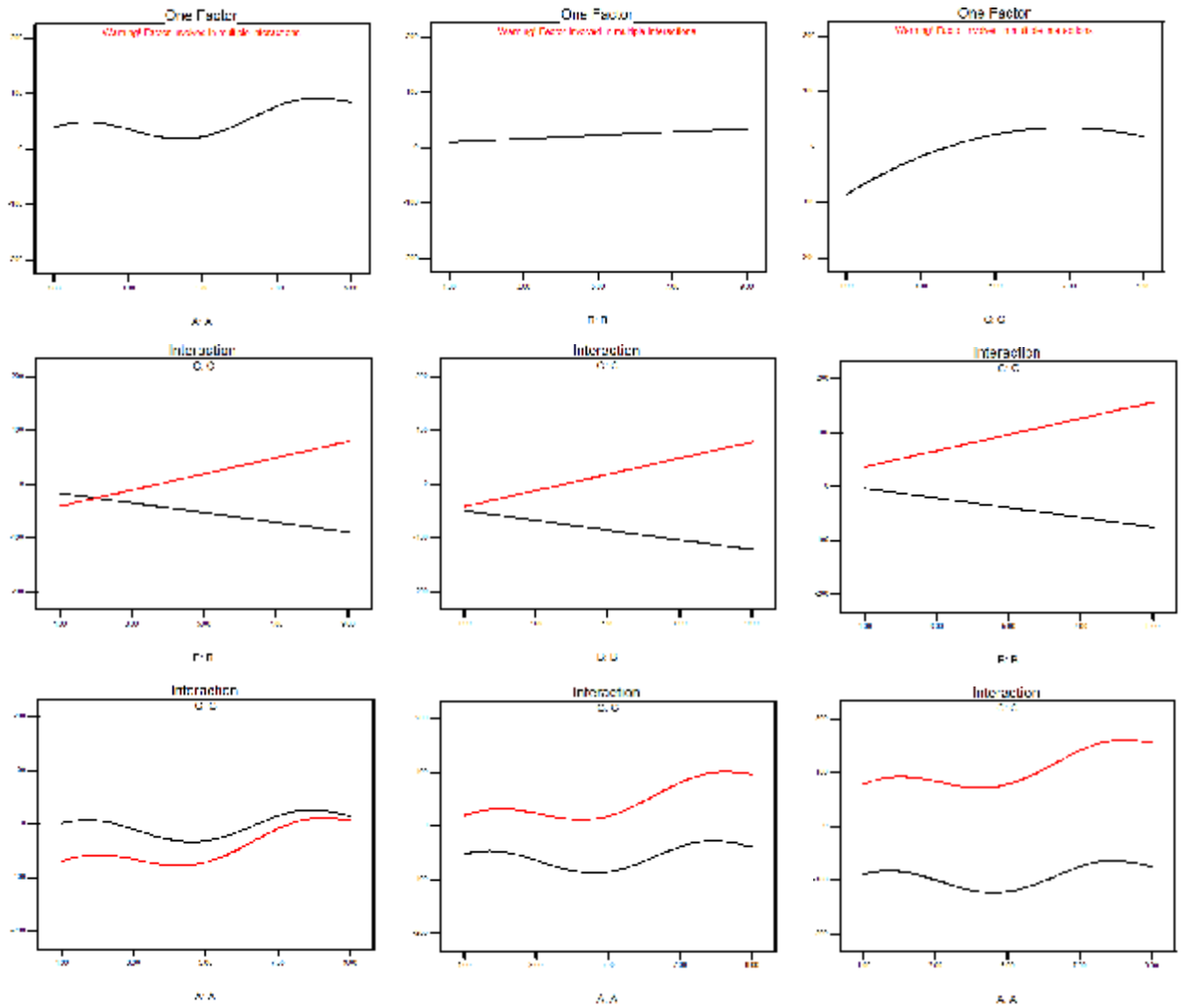
STATCON

The simulation model

Main effects

AC

BC



1 5 9

STATCON

All factors wide „Screening“

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Response 1 y
9.00	9.00	1.00	-75.197
9.00	1.00	1.00	-3.19704
9.00	9.00	9.00	156.803
5.00	5.00	5.00	22.5269
5.00	5.00	5.00	22.5269
1.00	9.00	1.00	-88.4632
9.00	1.00	9.00	36.803
1.00	1.00	9.00	-40.4632
1.00	1.00	1.00	-16.4632
1.00	9.00	9.00	79.5368

STATCON

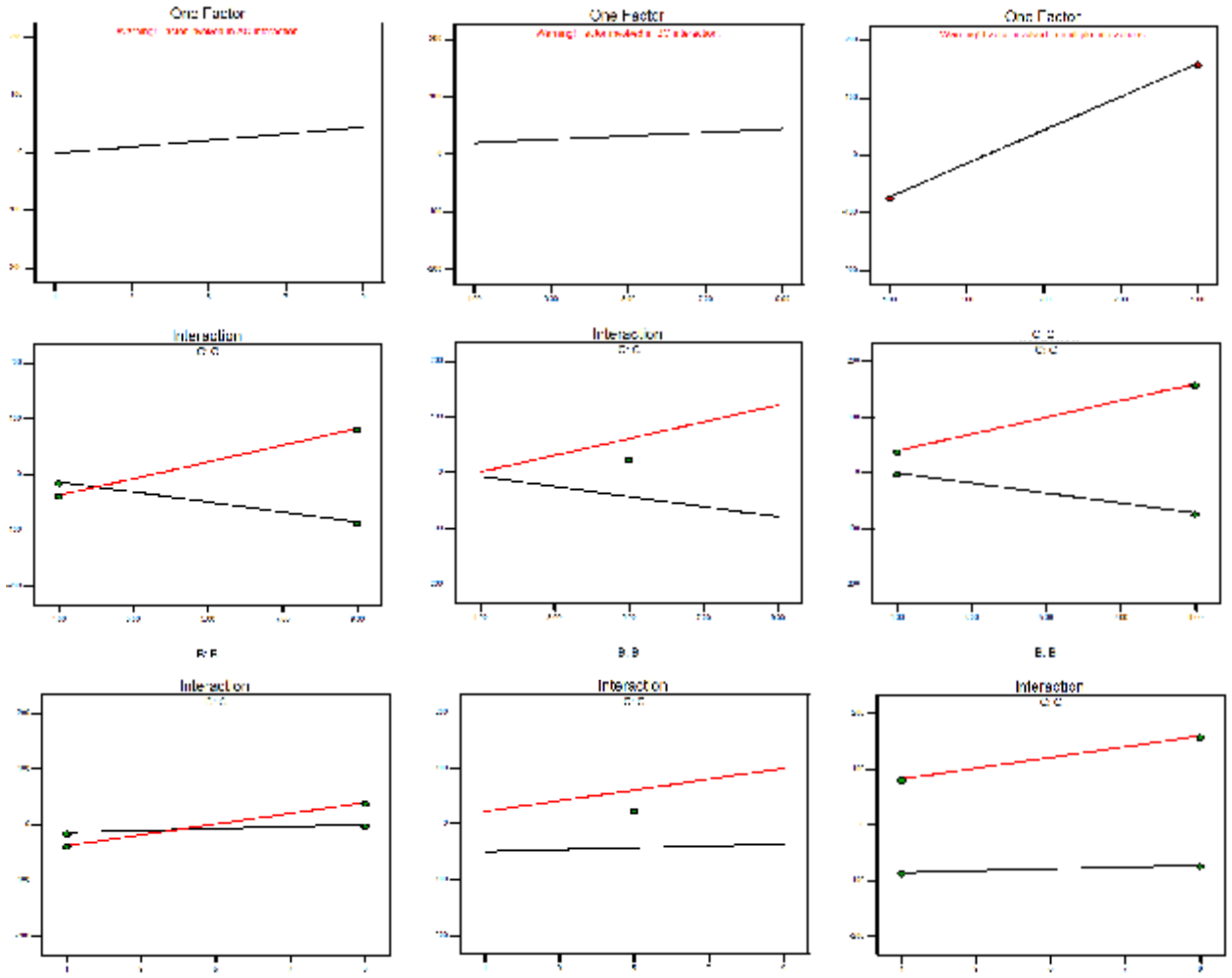
Singel factor narrow

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Response 1 y
4.00	9.00	9.00	72.5799
4.00	1.00	1.00	-47.4201
5.00	5.00	5.00	22.5269
4.00	1.00	9.00	-47.4201
4.00	9.00	1.00	-119.42
6.00	9.00	9.00	106.515
6.00	9.00	1.00	-101.485
5.00	5.00	5.00	22.5269
6.00	1.00	1.00	-29.4854
6.00	1.00	9.00	-13.4854

Factor A ranging between 4 and 6

STATCON

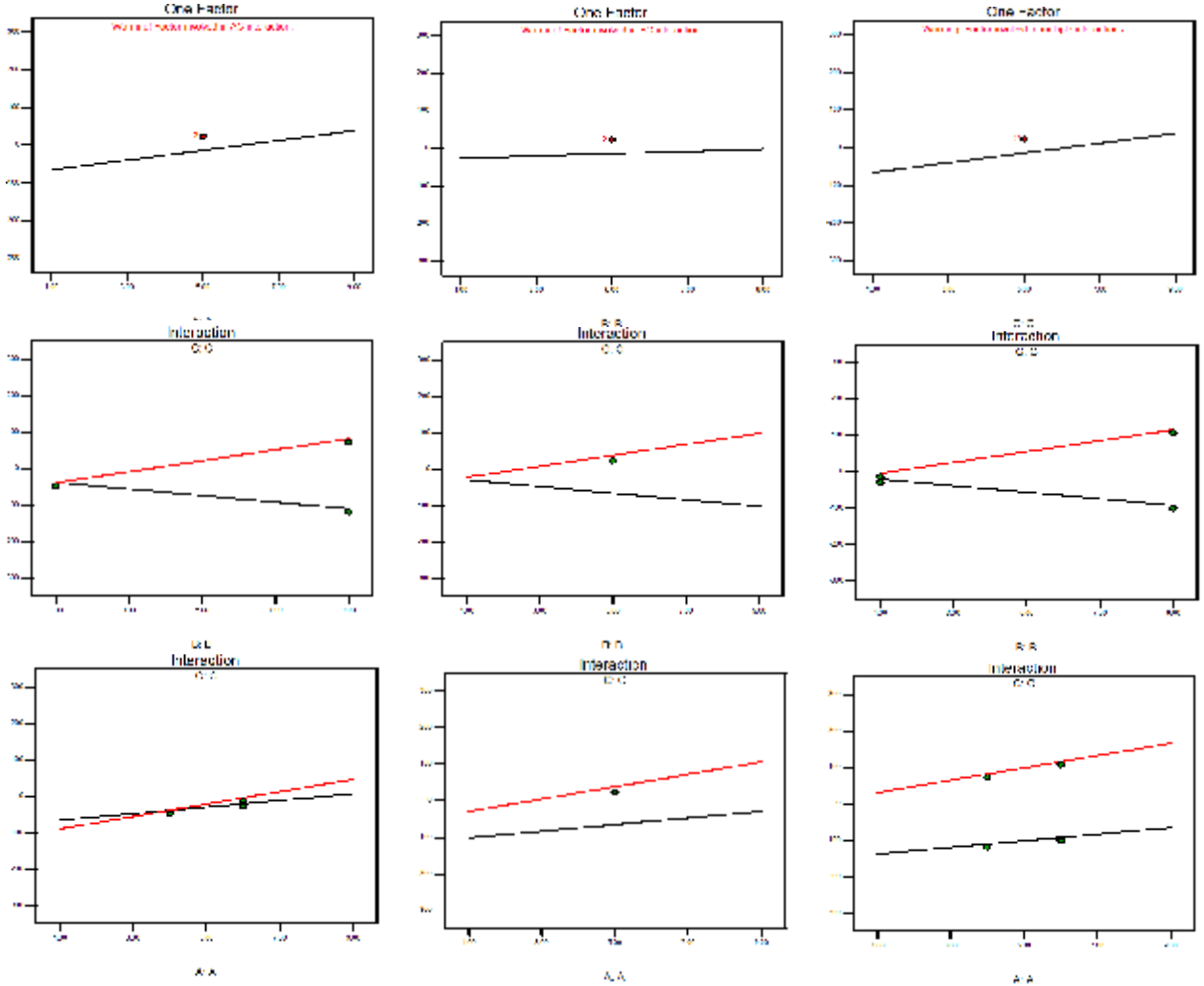
All factors wide „Screening“



1 5 9

STATCON

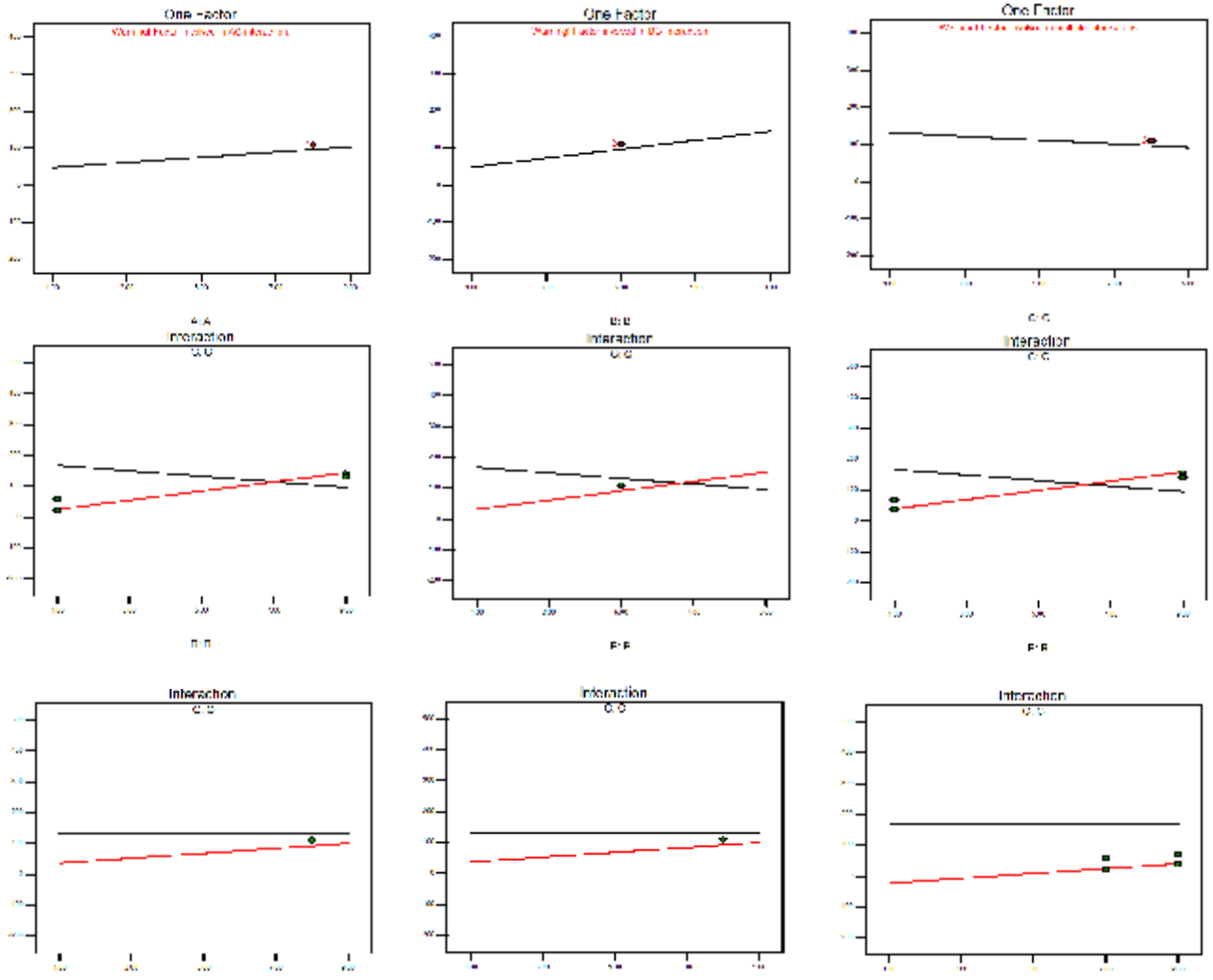
Singel factor narrow „Screening“



1 5 9

STATCON

Two factors narrow „Screening“



STATCON

Improper Randomization

- All factors sorted in standard order
- One factor sorted - others random

- These scenarios reflect situations were the user does not switch to a proper split plot design.
- the simulation model+1+A+B+C-D+AB+AC-AD

STATCON

Improper randomization fully sorted data

- Using a 2^3 full factorial randomized design adding a fourth factor D which changes it‘s level gradually
- Factor D is deleted for the analysis

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Factor 4 D:D	Response 1 Untitled
-1.00	-1.00	-1.00	-1.00	0
-1.00	-1.00	-1.00	-1.00	0
1.00	-1.00	-1.00	-0.75	-0.5
1.00	-1.00	-1.00	-0.75	-0.5
-1.00	1.00	-1.00	-0.50	0
-1.00	1.00	-1.00	-0.50	0
1.00	1.00	-1.00	-0.25	2.5
1.00	1.00	-1.00	-0.25	2.5
-1.00	-1.00	1.00	0.25	0
-1.00	-1.00	1.00	0.25	0
1.00	-1.00	1.00	0.50	1
1.00	-1.00	1.00	0.50	1
-1.00	1.00	1.00	0.75	0
-1.00	1.00	1.00	0.75	0
1.00	1.00	1.00	1.00	4
1.00	1.00	1.00	1.00	4

STATCON

Improper randomization one factor sorted

- Using a 2^3 full factorial randomized design adding a fourth factor D which changes it‘s level gradually
- Factor D is deleted for the analysis

Factor 1 A:A	Factor 2 B:B	Factor 3 C:C	Factor 4 D:D	Response 1 Untitled
1	1.00	-1.00	-1.00	4
1.00	1.00	-1.00	-1.00	4
-1.00	1.00	-1.00	-0.75	0
-1.00	-1.00	-1.00	-0.75	0
1.00	-1.00	-1.00	-0.50	-1
-1.00	1.00	-1.00	-0.50	0
1.00	-1.00	-1.00	-0.25	-1.5
-1.00	-1.00	-1.00	-0.25	0
1.00	-1.00	1.00	0.25	1.5
-1.00	-1.00	1.00	0.25	0
-1.00	1.00	1.00	0.50	0
1.00	1.00	1.00	0.50	5
1.00	-1.00	1.00	0.75	0.5
1.00	1.00	1.00	0.75	4.5
-1.00	-1.00	1.00	1.00	0
-1.00	1.00	1.00	1.00	0

STATCON

Conclusion 1/2

- Screening:
  - One factor narrow rarely changes the interpretation
  - Two or more factors narrow i.e. an overly conservative design space can cause misleading conclusions (assuming a non bumpy function)
  - Make sure to avoid symmetric factor levels around minima or maxima in assumed quadratic effects
- RSM
  - Narrow design space is less of a problem, sometimes even wanted
  - If the true model shows several steep ups and downs, inside the design space (which is very rare) don't use screening designs

Conclusions 2/2

- Power analysis based on hypothesized averages gives a better picture than based on minimum effect size
  - Minimum effect size is rarely estimated correct in advance
  - Sizing for precision is recommended for RSM
- In screening focus should be given to discover the full (wide) design space
- avoid symmetric effects, consultants need to ask for symmetry
- Randomize as much as possible or use proper split plot design

STATCON