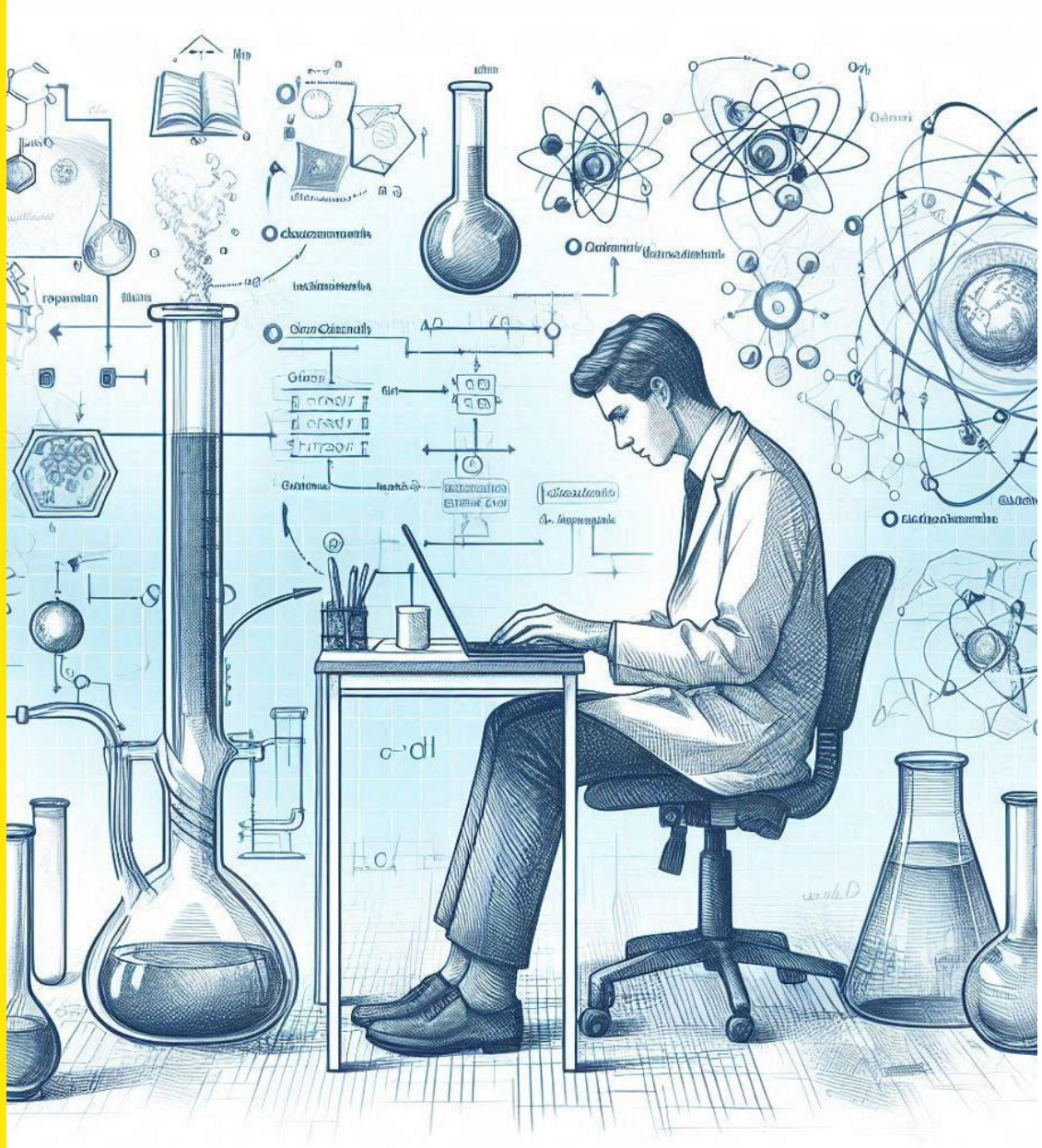
University of Applied Sciences and Arts Northwestern Switzerland
School of Life Sciences

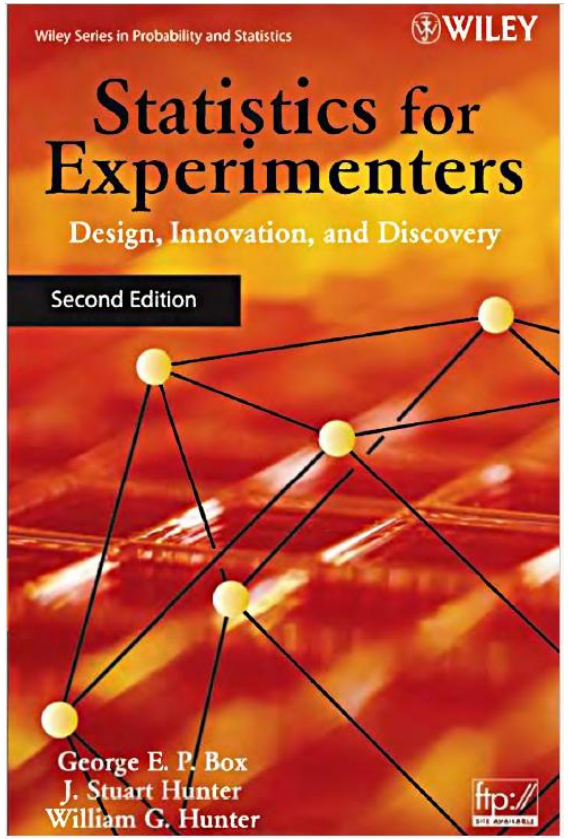
”Classic” DoE vs. Bayesian Optimization / Active learning: A risk-based approach

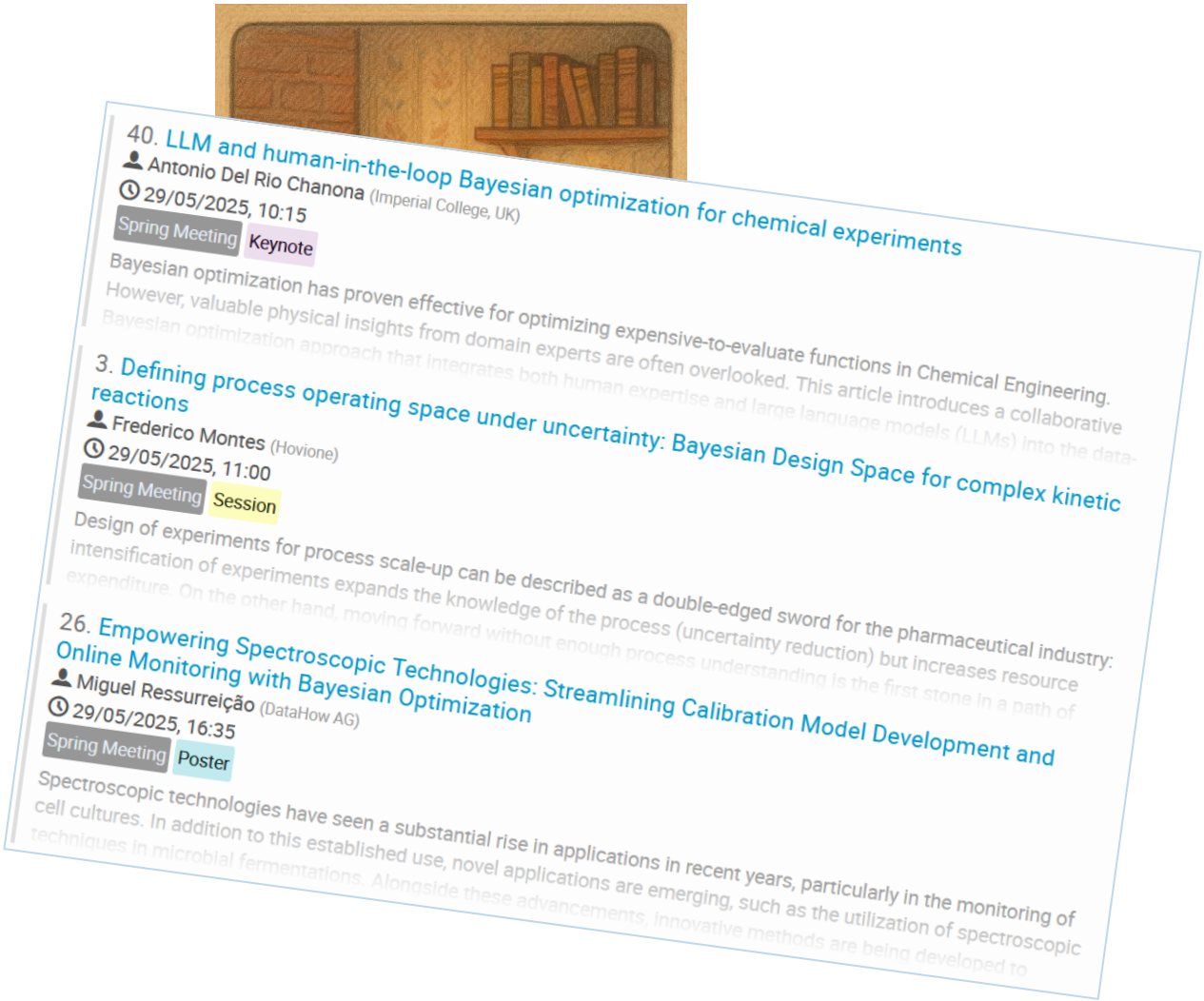


Stefanie Feiler

30.5.2025

Modern times





Questions

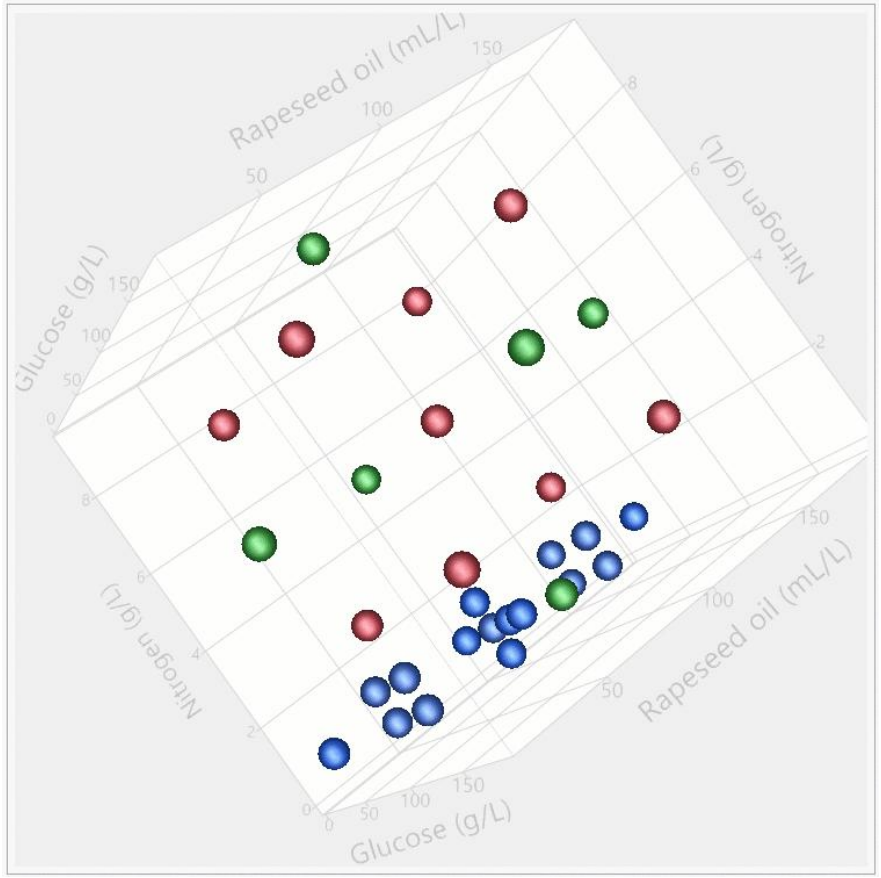
Is classic DoE obsolete?

If not, when to use which method?

Why does it feel so old-fashioned?

What is your opinion?

JMP DoE example



Bayesian optimisation (BO)

Presentation of Phil Kay (JMP)

Bayesian Optimization

Why? What pains could this relieve for experimenters?

- We don't like lots of runs upfront, especially "wasted" runs in "bad" factor regions
- We don't know how complex the system might be

- We need a solution, not a model, and fast!

Bayesian Optimization

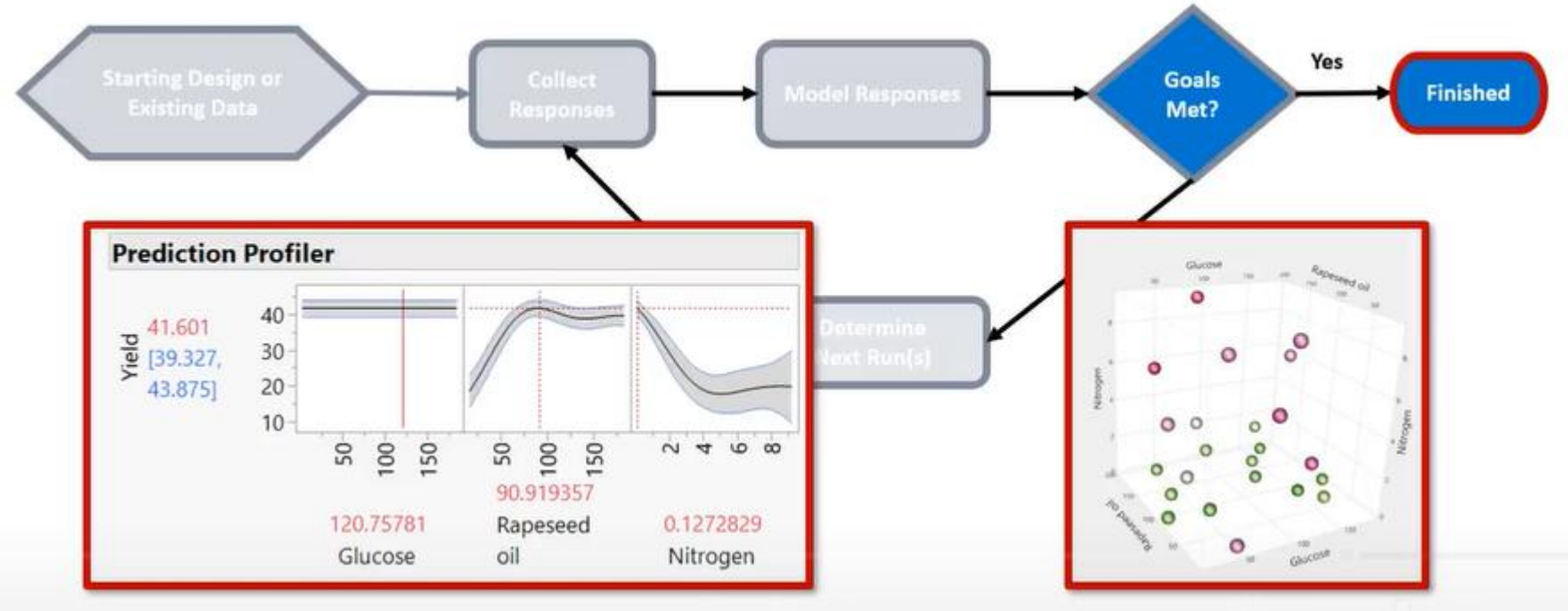
What is it?

- Goal-oriented, model-agnostic sequential experiments
- "Active Learning", "Adaptive Learning"
- An "AI recommender" for product and process innovation

- A revolution in how we experiment for innovation?

BO example

"… more than 20 runs in total"



Comparison: Sequential DoE

3 steps

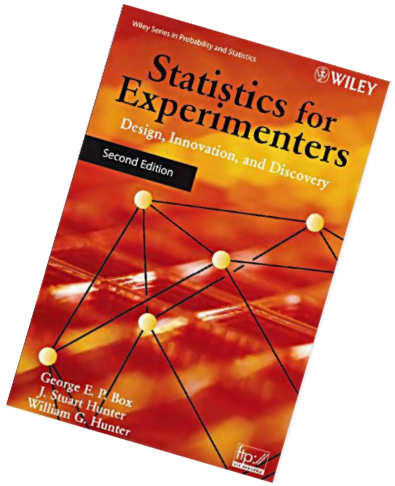
- Full factorial design, 2 replicates: 16 runs
- 14 additional runs
- 32 detailed ones

➔ 62 runs

We need a solution, not a model, and fast!

	Glucose (g/L)	Rapeseed oil (mL/L)	Nitrogen (g/L)	Sophorolipid (g/L)
	184 / 3.07	184 / 15.9	9.2 / 0.13	39.6 / 11.1
1	50	50	7.5	14.3
2	150	150	2.5	27.1
3	50	150	7.5	18.2
4	50	50	2.5	17.7
5	150	50	7.5	14.7
6	50	150	2.5	25.9
7	50	50	2.5	16.6
8	150	50	2.5	12.6
9	50	150	7.5	11.6
10	150	50	2.5	13.2
11	50	50	7.5	12.3
12	150	50	7.5	16.7
13	50	150	2.5	23.0
14	150	150	7.5	12.2
15	150	150	7.5	17.5
16	150	150	2.5	29.4

DoE philosophy



Box / Hunter / Hunter:

- In an ongoing investigation, a rough rule is that only a portion (say 25%) of the experimental effort and budget should be invested in the first design.
- Perfection is not possible: it‘s always an approximation.

⇔ Observation (nowadays):

DoE is propagandised as “be on the safe side”

➔ duplicate runs

➔ tolerance intervals at lab scale

➔ power considerations

Lots of runs!

“Classic DoE”?

- That a conclusion reached in one environment (say the laboratory) will apply in a different environment (say the full-scale process) is based not on statistical reasoning but what Deming called „a leap of faith.“ Good statistics and subject matter knowledge can narrow the chasm but not eliminate it.
- Among the factors to be considered there will usually be the vital few and the trivial many. (J. M. Juran)

The need to work closely with subject matter specialists.

The value of a sequential approach to problem solving and in particular the sequential assembly of experimental designs.

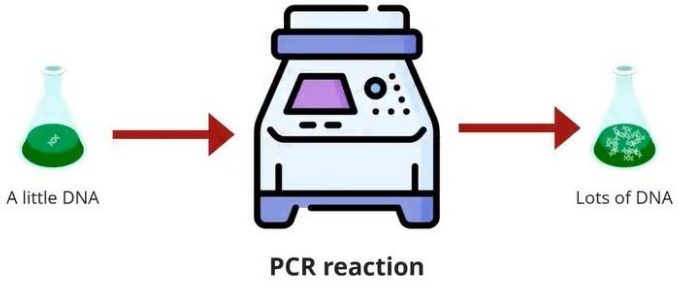
2nd example: PCR (https://virtual-pcr.ico2s.org/pcr)

BO (LinkedIn post of Morten Borman Nielsen):

12 factors

➔ Aim > 500 ng/µL

We want to improve the PCR reaction:
A method used to chemically multiply DNA



We have a solution after 25 runs, but maybe it is possible to reach an even better result?

And after *35 runs* we have a very good solution:

Can your present strategy handle this problem?

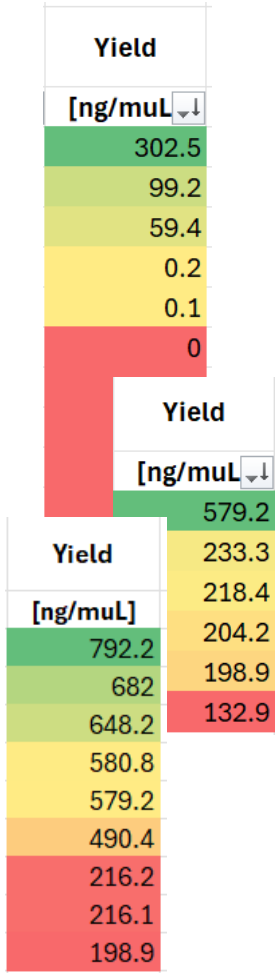
If not, you might want to add Bayesian Optimization to your scientific toolkit ;)

My strategy (“granny DoE”)

- Screening: Plackett-Burman (12 runs)
- Risky: D-optimal design for 2 factors
  (8 runs, max. 194 ng/µL ✗)
- 2nd stage:
  D-optimal design for 4 factors (6 runs)
- Optimisation (RSM):
  Full factorial for 2 factors (9 runs)
- Confirmation

➔ 36 including 8 unnecessary;

➔ already 2nd run of 2nd stage > 500 ng/µL

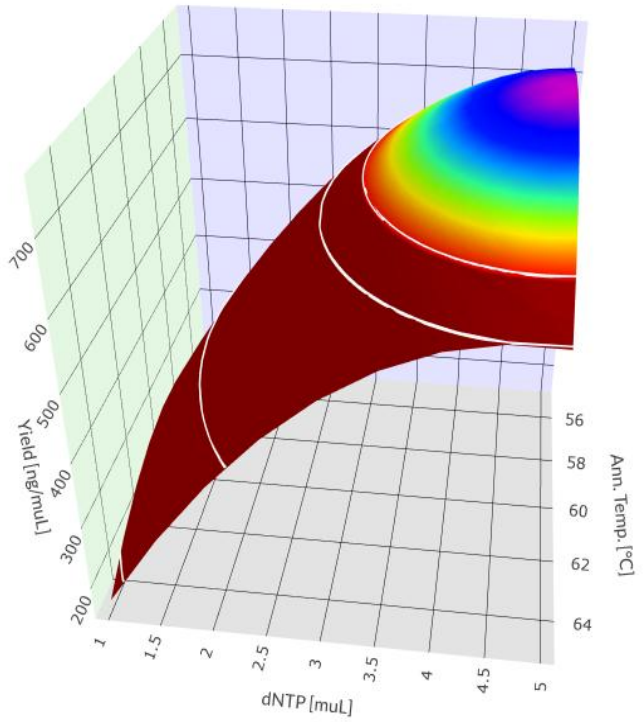


Subject matter expert:

- domain knowledge + reference literature
  ➔ reducing factors
- realistic factor ranges
- setting priorities
  (too expensive, too long)

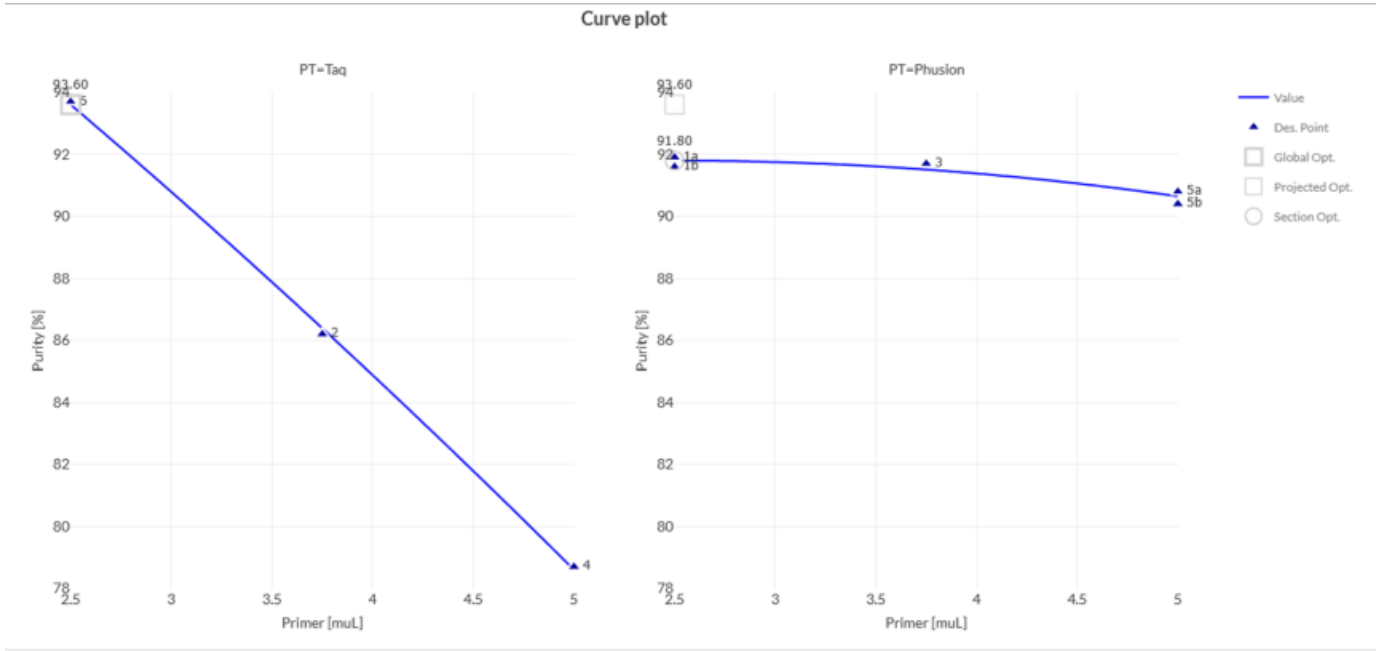
Knowledge

Desirability



Comparison TAQ ⇔ Phusion:

Less purity with increasing primer



Benefits: BO ⇔ Classic

BO:

- Running process improvement
  (“A process should be routinely operated so as to produce not only product but information on how to improve the product.”)
- Convincing people
  (⇔ “trial & error”, OFAT)

“Classic“

- Lab scale experiments starting from 0
  ➔ statistical tools
  (model reduction ⇔ factor importance)
  ➔ visualisation
  (few parameters: response surface)
- Convincing people (⇔ OFAT…?)

automated lab systems
⇔ depending on the aim

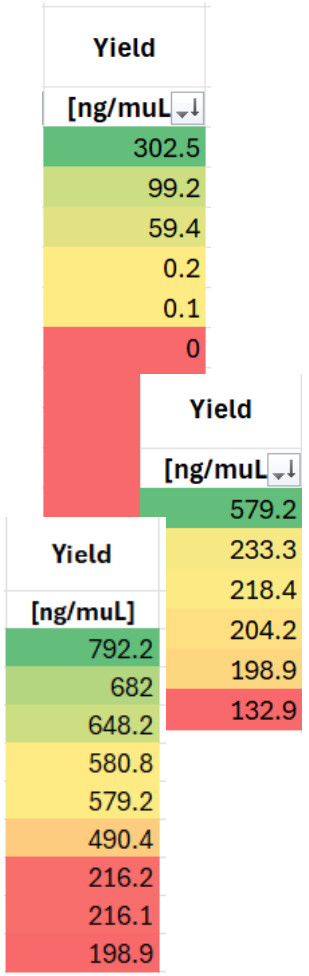
Validation context
(Design space, robustness ⇔ FDA?)

Number of runs?

Solution:

➔ radical risk-based approach (including subject matter expertise)

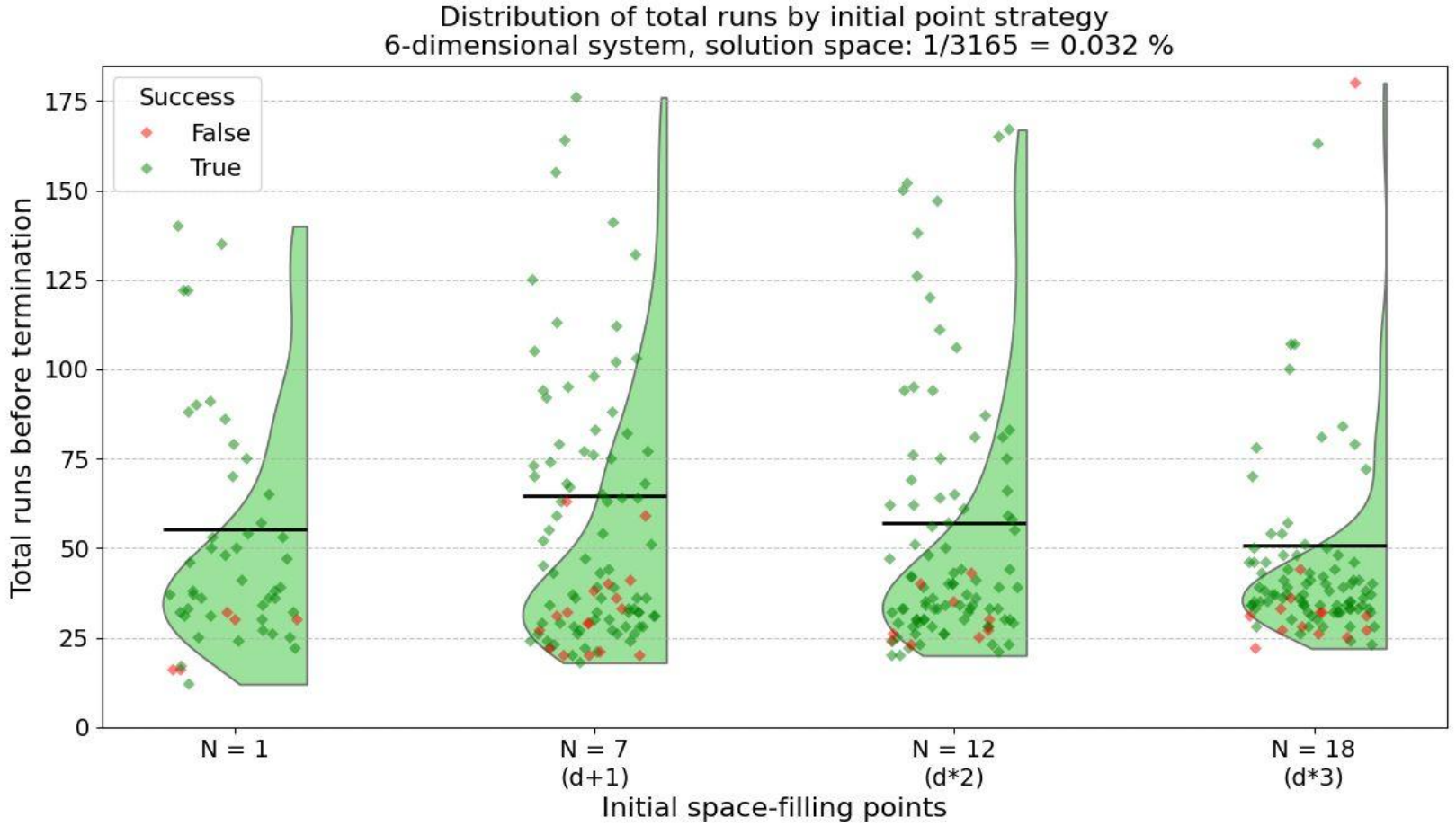
⇔ ICH / FDA! ✓



Ongoing

Expected Run Time (ERT) for BO (work of Morten Borman Nielsen)

➔ comparison to “classic” DoE



What is your opinion?

Is classic DoE obsolete?

If not, when to use which method?

Why does it feel so old-fashioned?