

Modelling of Process History in intensified Design of Experiment Data

Verena Nold

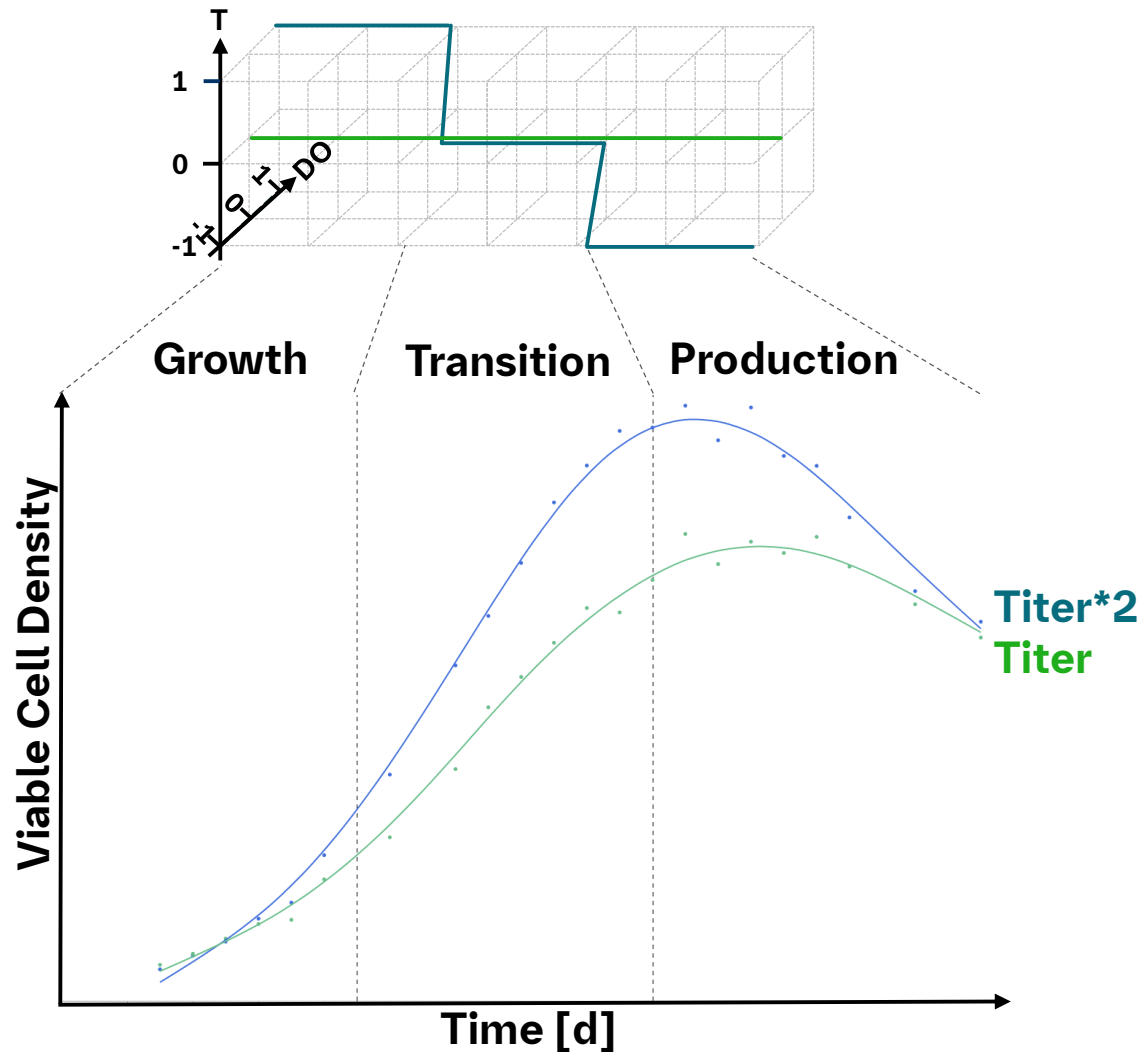


Bioprocess Dynamics Modulate Productivity

iDoE = Intensified design of experiment

T = Temperature

DO = Dissolved oxygen



- iDoE is a systematic approach to
 - Increase process understanding
 - Improve process performance
 - develop protocols containing input shifts
- iDoE requires
 - Biological feasibility ^{1,2}
 - Effect reversibility ³
 - Consideration of memory effects

Adequate design planning and modelling is
challenging

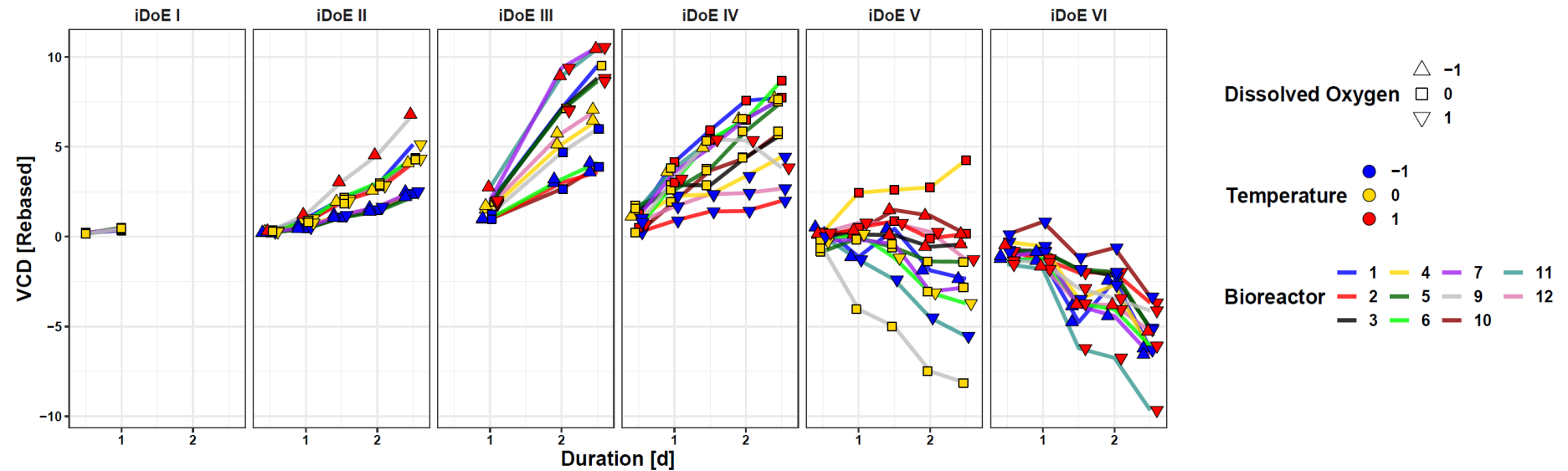
Published iDoE-Format – One-Factor-One-Column (OFOC)

Reactor	Time	T	DO	
1	0	-1	+1	Stage 1
1	2	-1	+1	
1	4	-1	+1	
1	6	0	-1	Stage 2
1	8	0	-1	
1	10	1	0	Stage 3
1	12	1	0	
1	14	1	0	
...	

- Requires time column to capture all information
- Model includes complex interactions between iDoE-mode factors * time

Stage-Wise Modelling of OFOC-Data

iDoE = Intensified design of experiment
 OFOC = one factor, one column
 VCD = viable cell density



- Time course modelling, separately for each stage (nested DoE) with overlaps at the end / beginning of a stage
- Initial values as additional (not explicitly plannable) input factor in the design to capture different states
- Re-basing to initial value per stage and bioreactor as offset correction
- Concatenation of models required to describe a given time point
- Error propagation for uncertainty bands

Initial values indirectly address the memory effect (influence of process history) by capturing the state of the culture

Alternative iDoE-Format - One Factor Multiple Columns (OFMC)

Reactor	Time	T	DO
1	0	-1	+1
1	2	-1	+1
1	4	-1	+1
1	6	0	-1
1	8	0	-1
1	10	1	0
1	12	1	0
1	14	1	0
...

Stage 1

Stage 2

Stage 3

One-Factor-One-Column (OFOC)

- Requires time column to capture all information
- Model includes interaction factor * time

Reactor	Time	Stage 1		Stage 2		Stage 3	
		T1	DO1	T2	DO2	T3	DO3
1	0	-1	+1	0	-1	1	0
1	2	-1	+1	0	-1	1	0
1	4	-1	+1	0	-1	1	0
1	6	-1	+1	0	-1	1	0
1	8	-1	+1	0	-1	1	0
1	10	-1	+1	0	-1	1	0
1	12	-1	+1	0	-1	1	0
1	14	-1	+1	0	-1	1	0
...

One Factor Multiple Columns (OFMC)

- More factors
- More potential interactions

iDoE-Formats Focus on Different Modelling

Reactor	Time	T	DO	
1	4	-1	+1	Stage 1
1	8	0	-1	Stage 2
1	14	1	0	Stage 3
2	

One-Factor-One-Column (OFOC)

- Requires time column to capture all information
- Model includes interaction factor * time
- End of fermentation prediction reached from (concatenated) time courses

Reactor	Time	Stage 1		Stage 2		Stage 3	
		T1	DO1	T2	DO2	T3	DO3
1	14	-1	+1	0	-1	1	0
2	14	0	+1	+1	-1	-1	0

One Factor Multiple Columns (OFMC)

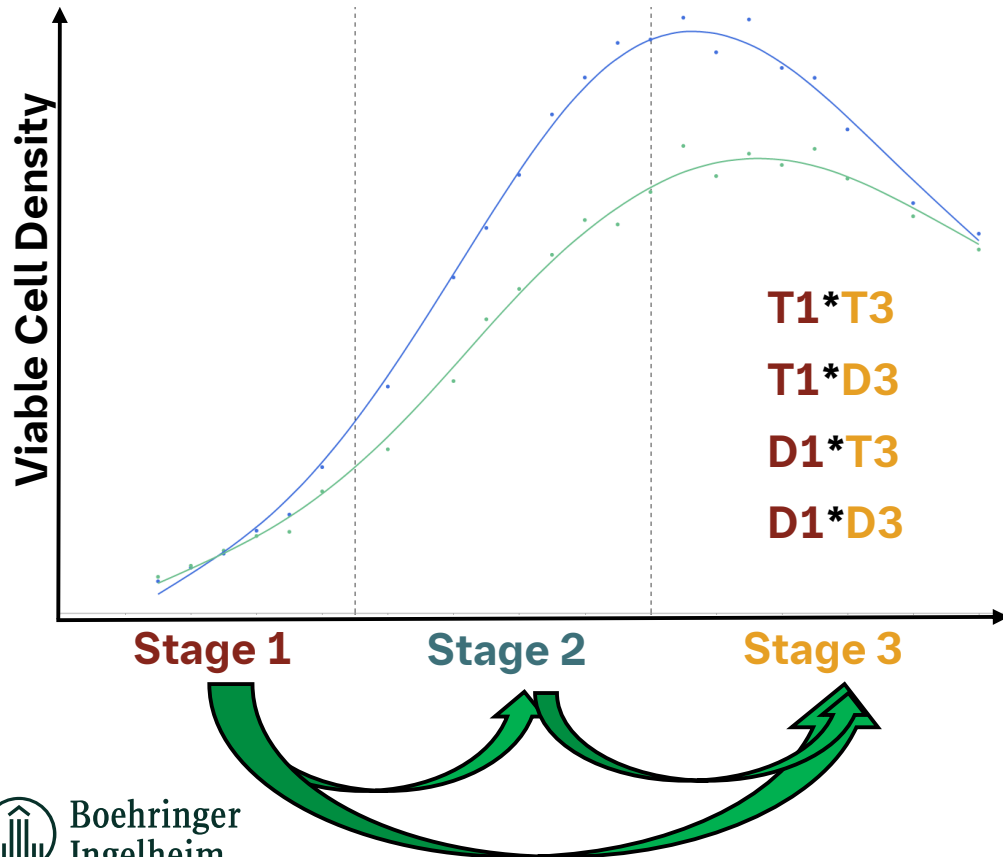
- More factors
- More potential interactions
- Direct modelling of a certain time point or time course models possible

Across-Stage Interaction Modelling of OFMC Data

OFMC = one factor, multiple columns
 T = Temperature
 D = Dissolved oxygen

T1*T2 T2*T3
 T1*D2 T2*D3
 D1*T2 D2*T3
 D1*D2 D2*D3

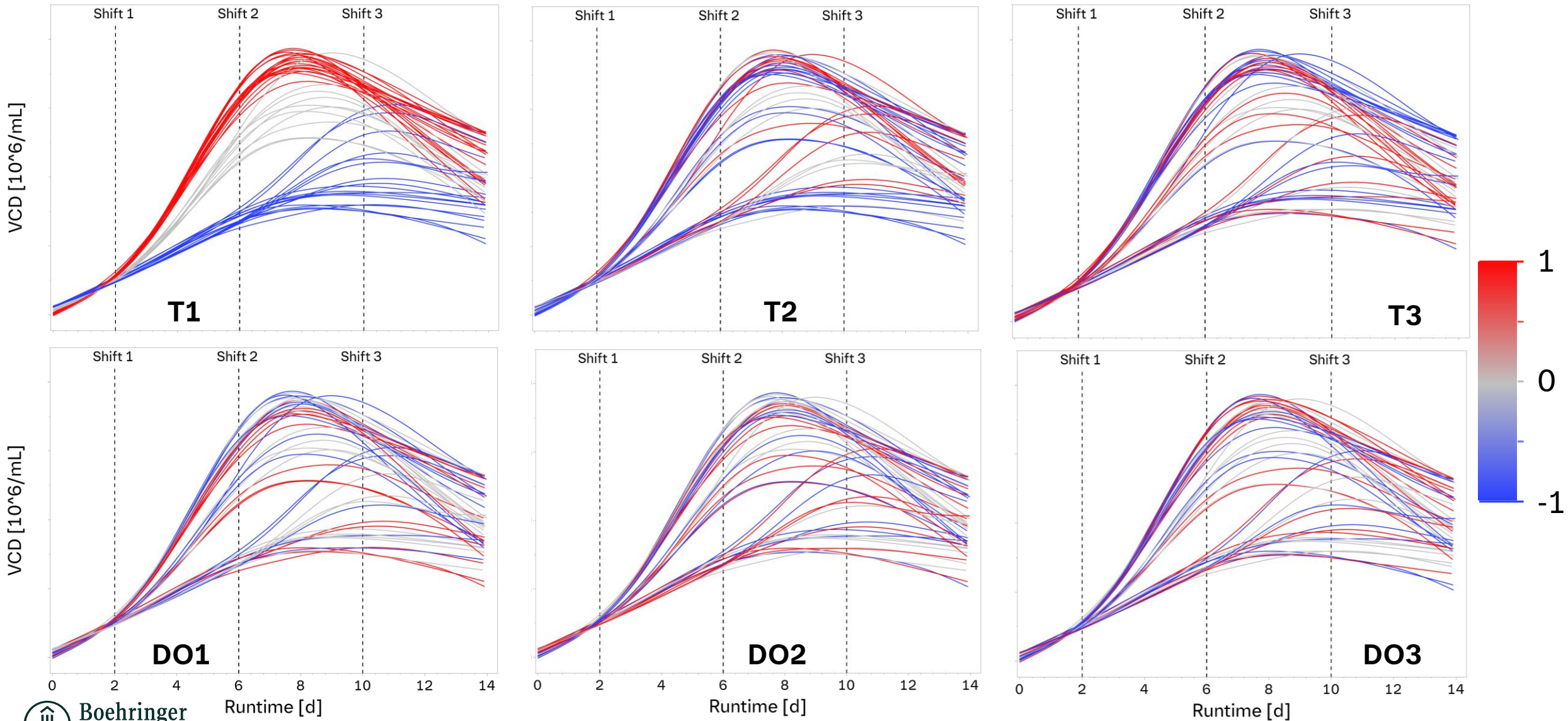
Reactor	Time	T1	D1	T2	D2	T3	D3
1	14	-1	+1	0	-1	1	0
2	14	0	+1	+1	-1	-1	0



Potential across-stage interactions capture memory effects (influence of process history causing the state of the culture)

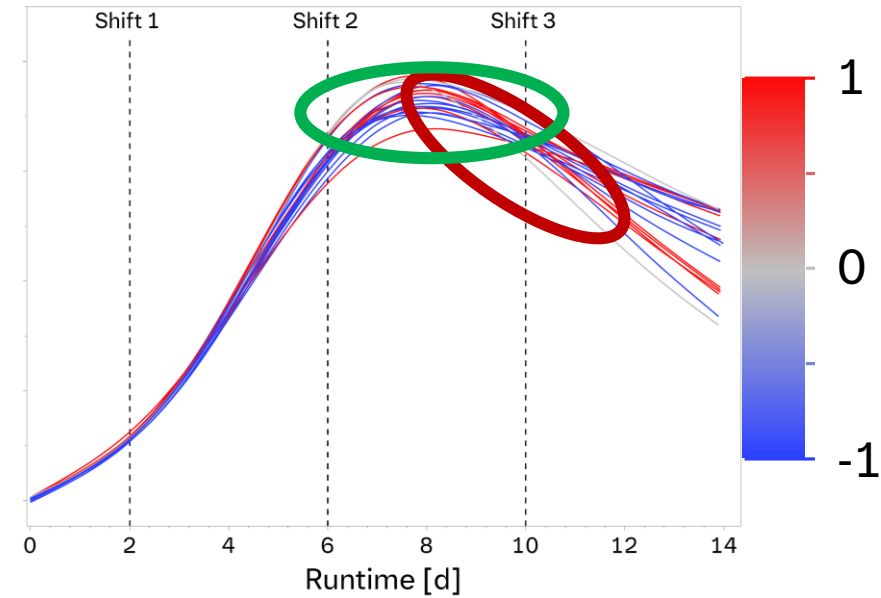
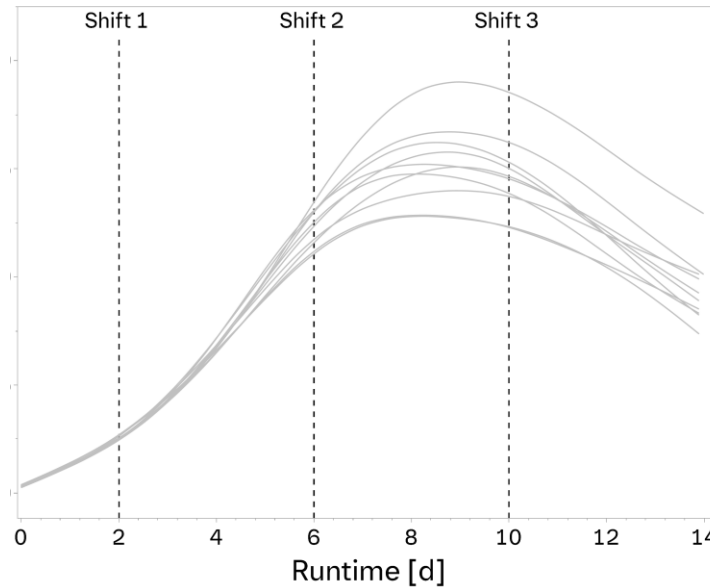
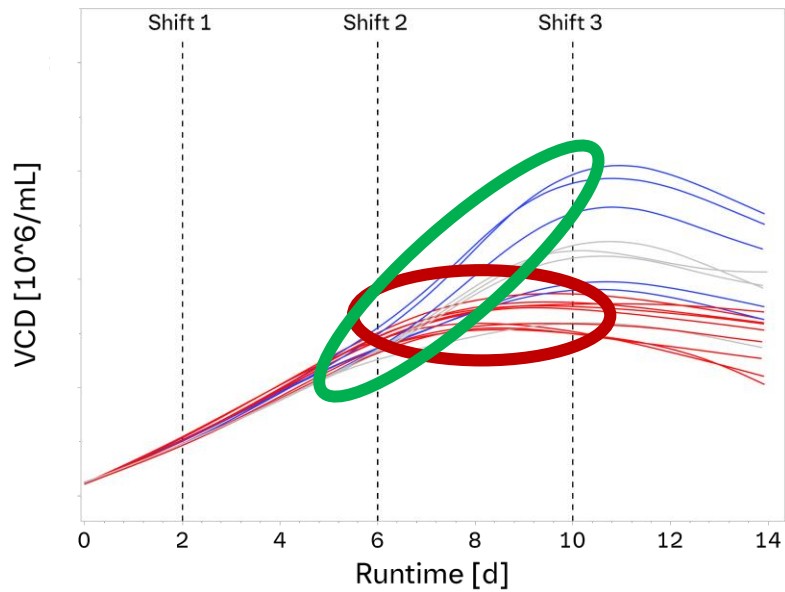
Impact of Stage 1 Settings is Obvious in the Raw Data

VCD= viable cell density
T = temperature
D = dissolved oxygen



T1 × T2 Suggests Change in Right Order to be Beneficial

T = temperature



$$T1 (-1) * T2 (-1) = 1$$

$$T1 (-1) * T2 (0) = 0$$

$$T1 (-1) * T2 (+1) = -1$$

$$T1 (0) * T2 (+1) = 0$$

$$T1 (0) * T2 (0) = 0$$

$$T1 (0) * T2 (-1) = 0$$

$$T1 (+1) * T2 (+1) = 1$$

$$T1 (+1) * T2 (0) = 0$$

$$T1 (+1) * T2 (-1) = -1$$

Determine Importance of Across-Stage Interactions

RSM = response surface model
 FI = factor interaction
 (n)ASI = (non-) across stage interaction
 Min = minimum
 Max = maximum
 T = temperature
 D = dissolved oxygen

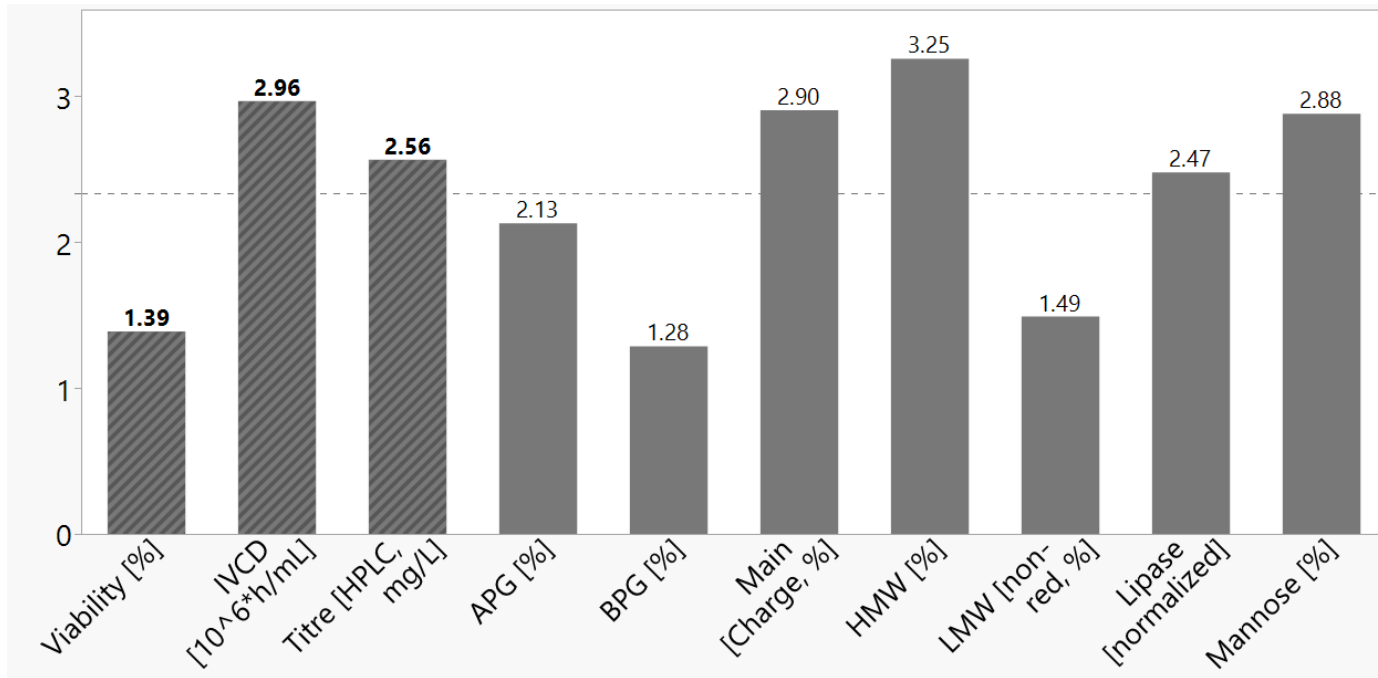
- Fit d14 data (without model selection) for
 - 3 performance-related responses
 - 5 product quality-related responses
- Full RSM (6 main, 6 quadratic, 2 2FI, 12 ASI)
- Scale the estimates within a model

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

T1	T1 ²	T1 * D1	T1 * T2	T1 * D2
T2	T2 ²	T2 * D2	T1 * T3	T1 * D3
T3	T3 ²		T2 * T3	T2 * D1
D1	D1 ²		D1 * D2	T2 * D3
D2	D2 ²		D1 * D3	T3 * D1
D3	D3 ²		D2 * D3	T3 * D2

ASIs are on Average 3-Times Smaller than Other Terms

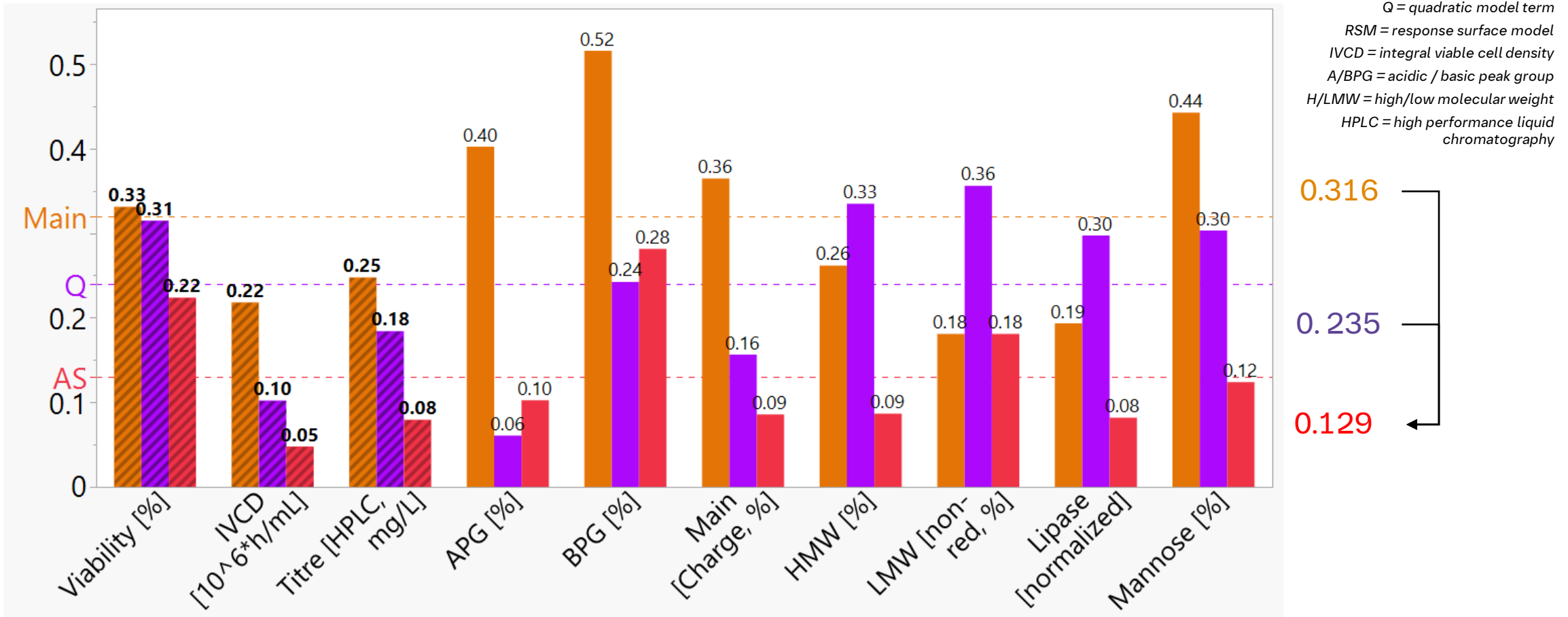
(n)ASI = (non-) across stage interaction
 RSM = response surface model
 IVCD = integral viable cell density
 A/BPG = acidic / basic peak group
 H/LMW = high/low molecular weight
 HPLC = high performance liquid chromatography



Response	Average (ASI)	Average (nASI)	Average(nASI)/Average(ASI)
Viability [%]	0.22385	0.31032	1.386317933
IVCD [10 ⁶ *h/mL]	0.04781	0.14158	2.961487678
Titre [HPLC, mg/L]	0.07953	0.20369	2.561337202
APG [%]	0.10258	0.21804	2.125656086
BPG [%]	0.28131	0.36123	1.284107375
Main [Charge, %]	0.08583	0.24889	2.899644596
HMW [%]	0.0866	0.2816	3.251553617
LMW [non-red, %]	0.18103	0.26923	1.487224736
Lipase [normalized]	0.08183	0.2024	2.473446365
Mannose [%]	0.12396	0.35644	2.875504561

Which model term classes are most influential?

Close-Up Relative Estimate Sizes per Model Term Class



- (Across-Stage) Interactions are on average the least influential (except APG, BPG)
- Main effects or quadratic effects are on average most influential

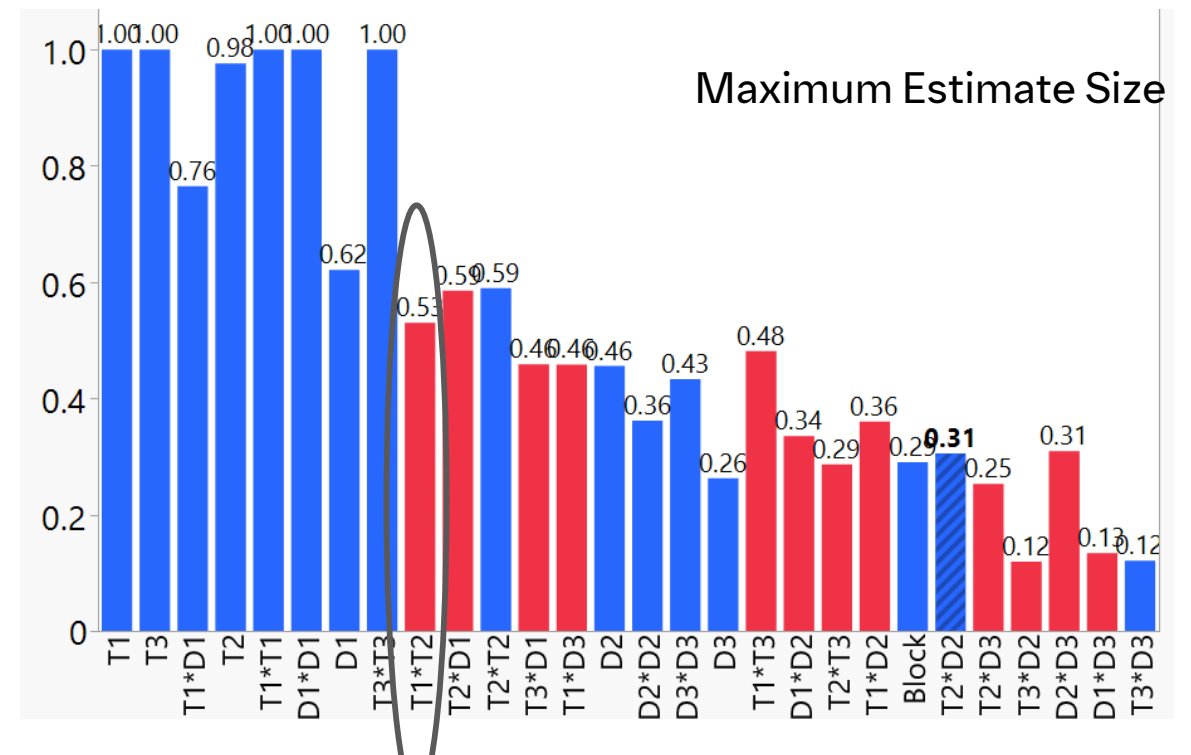
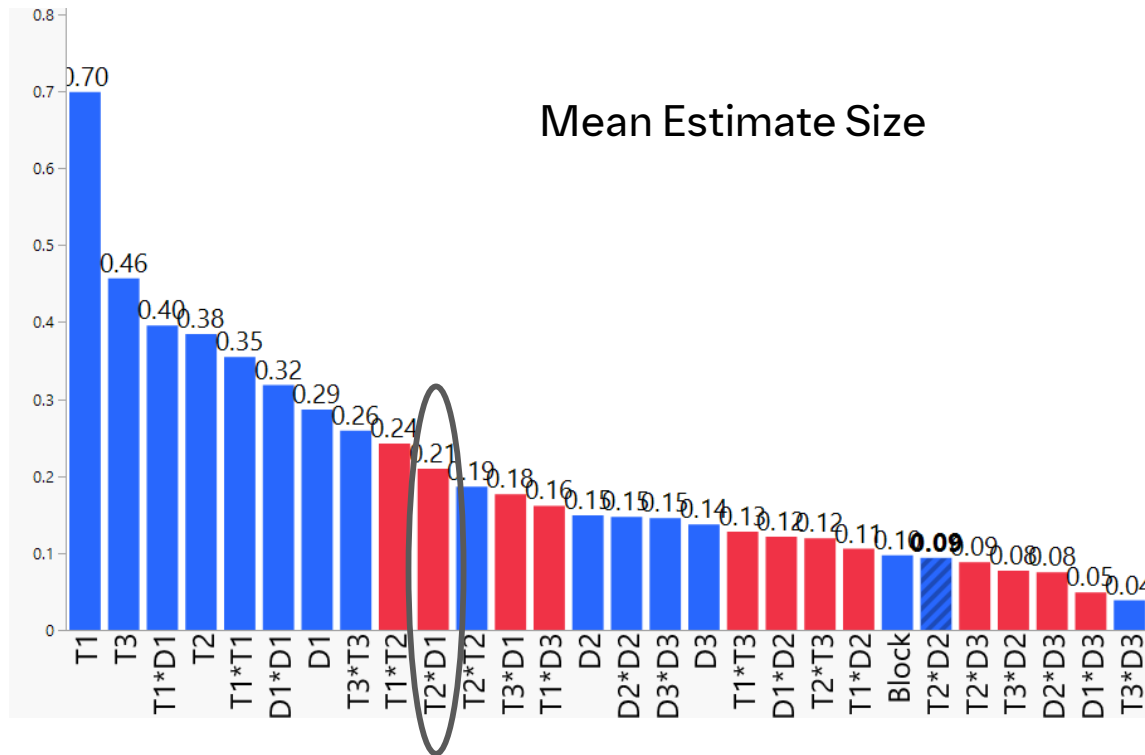
Are there Extremely Large ASIs?

ASI = across stage interaction
 RSM = response surface model
 IVCD = integral viable cell density
 A/BPG = acidic / basic peak group
 H/LMW = high/low molecular weight
 T = temperature
 D = dissolved oxygen

- Some ASIs are large in some responses while not so important in others

Term	IVCD	Viability	Titre	Mannose	APG	BPG	Main Charge	Lipase	LMW	HMW
T2 [°C]*DO1 [%]	0.05221	0.585356	0.015884	0.20953	0.189233	0.486571	0.070739	0.032133	0.380571	0.072729

- Some ASI are even the most important model term for individual models, e.g. T1*T2 is third largest effect for LMW



'Dilution' of ASI importance if few ASI have a large effect estimate

Determine Importance of Across-Stage Interactions

RSM = response surface model
 FI = factor interaction
 (n)ASI = (non-) across stage interaction
 Min = minimum
 Max = maximum
 T = temperature
 D = dissolved oxygen

- Fit d14 data (without model selection) for
 - 3 performance-related responses
 - 5 product quality-related responses
- Scale the estimates within a model

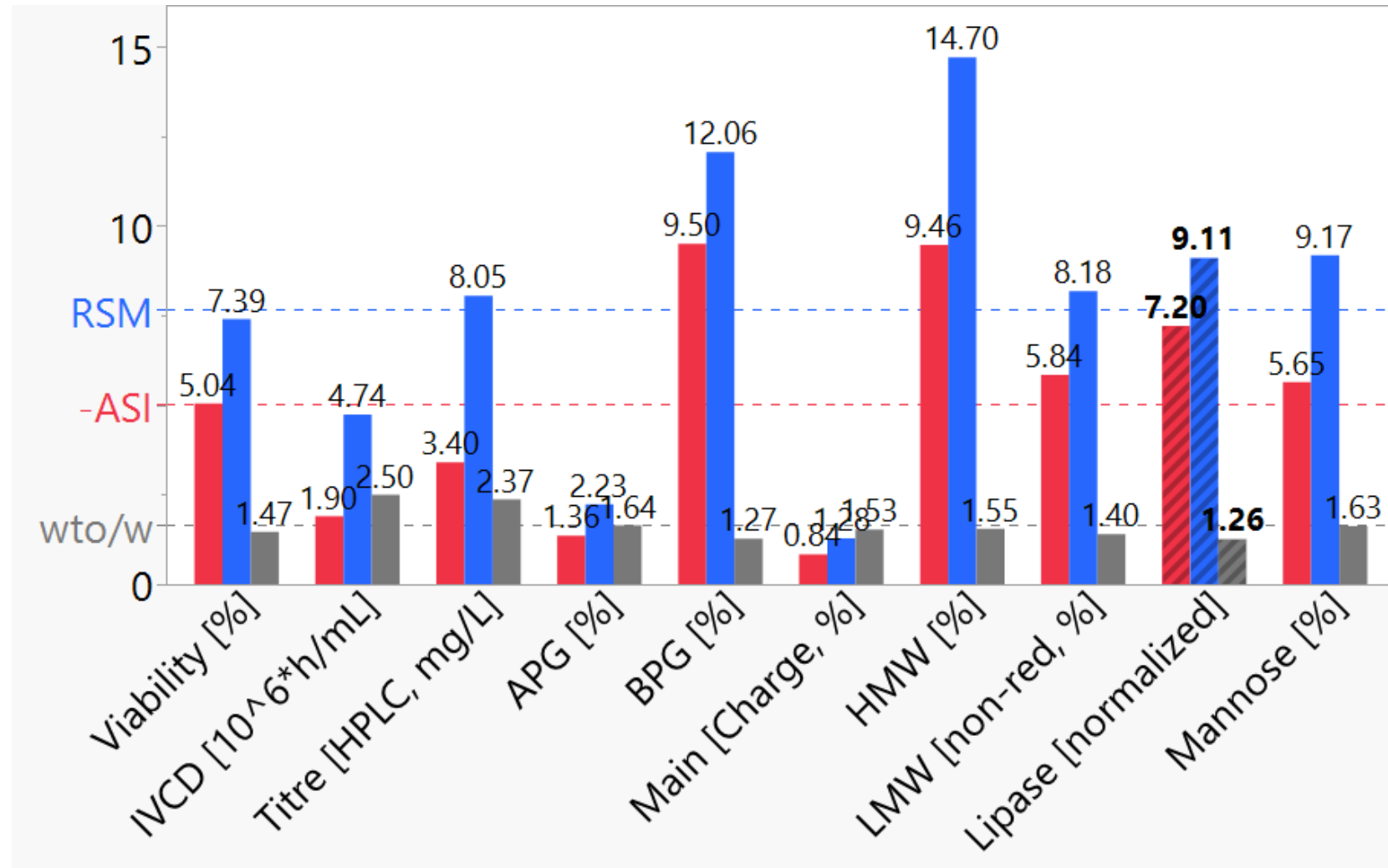
$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

- Compare model complexities
 - Full RSM (6 main, 6 quadratic, 2 2FI, 12 ASI)
 - RSM without ASI

T1	T1 ²	T1 * D1	T1 * T2	T1 * D2
T2	T2 ²	T2 * D2	T1 * T3	T1 * D3
T3	T3 ²		T3	T2 * D1
D1	D1 ²		D1 * D2	T2 * D3
D2	D2 ²		D1 * D3	T3 * D1
D3	D3 ²		D2 * D3	T3 * D2

T1	T1 ²	T1 * D1		
T2	T2 ²	T2 * D2		
T3	T3 ²			
D1	D1 ²			
D2	D2 ²			
D3	D3 ²			

Omitting ASIs Leads to Strongly Increased MAPEs



MAPE = Mean absolute percentage error
 ASI = across stage interaction
 RSM = response surface model
 W(to) = with(out)
 IVCD = integral viable cell density
 A/BPG = acidic / basic peak group
 H/LMW = high/low molecular weight
 A_t = Actual value
 F_t = Forecast value
 n = fitted points
 N = number observations

Mean increase in MAPE is 66.16%

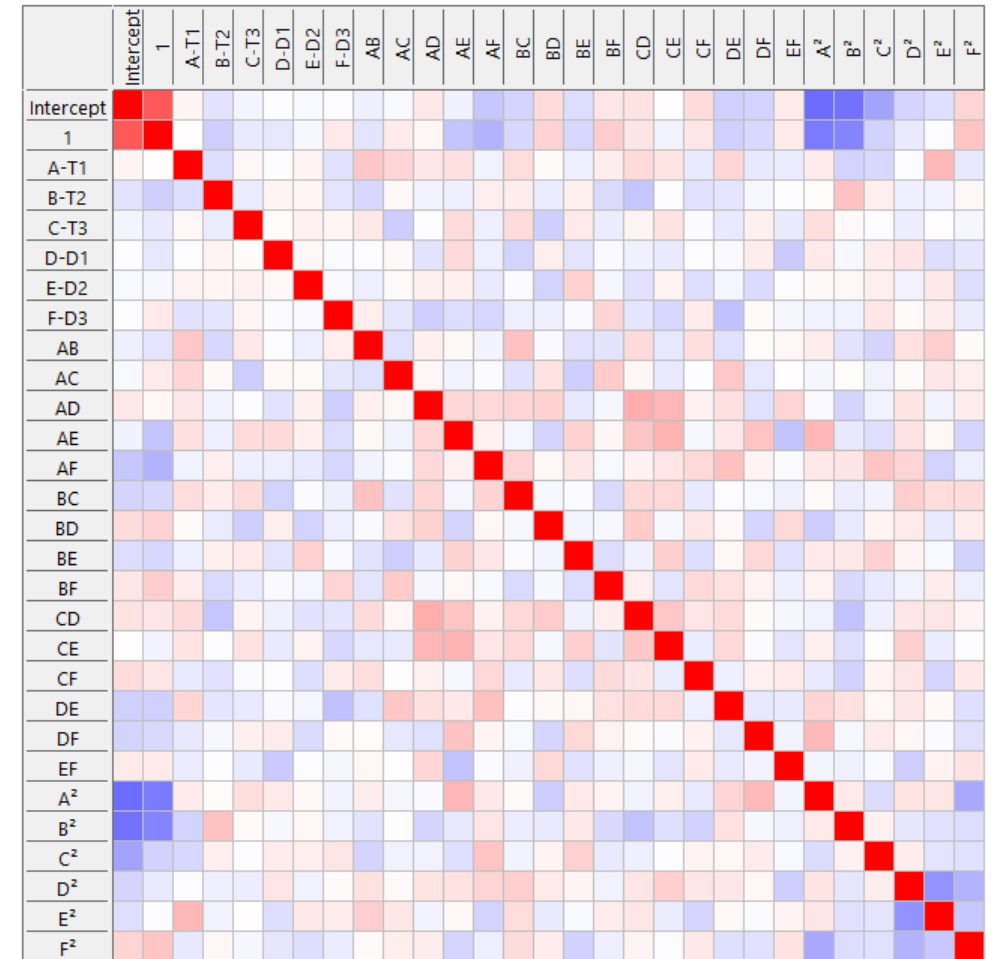
$$MAPE = 100 \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Can ASI be Neglected in iDoE Planning?

ASI = across stage interaction
 iDoE = intensified design of experiments
 MAPE = Mean absolute percentage error

- Only relatively small effect of ASIs on average
- Few ASI have a high effect estimate
- Multiple ASIs jointly have a strong impact on the model (MAPE)
- Designs with minimal correlation structure enable to estimate model terms independently
 - minimal bias of estimated terms (if not selecting model structure)
 - Difference in model would be more visible for „worse“ design

Models must be able to estimate relevant ASI without biasing non-ASI model terms



Questions about Statistics?

IU CMC Statistics



Dr. Beate Presser
Head of CMC Statistics



Dr. Armin Böhler
ADB, Device, VTC



Dr. Alexandra
Erdmann
ADB



Dr. Anja Wittmann
BDB, ADB



Marco Kunzelmann
BDB, PDB, ADB,



Jan Müller
BDB



Dr. Verena Nold
BDB, VTC



Lena Peter
PDB, BDB



Alessa Schurr
PDB



Samuel Kienzle
PhD-Student, PIDs



Dr. Sonja Schach
PostDoc, PIDs



Viva Engage:

CMC Statistics Development Biologicals

Website:

- CMC Statistics (boehringer.com)
- R & D and CMC Statistics Community (boehringer.com)

Training and Seminars:

- Statistik Seminar
- DoE Seminar
- DoE Training („Kurzschulung“ or 2-day course on demand)

Back Up

Supplementary Material



Orient Choice of Stages on Bioprocess Dynamics

- Study all factors at 3 levels
- 2x Ambr250 = 48 experiments
- Stage 0 at center point to allow accommodation to split and seeding into new bioreactor
- 4 days for each shift-induced change to provide sufficient time for adaptation to new settings and measurable change of cellular behaviour

