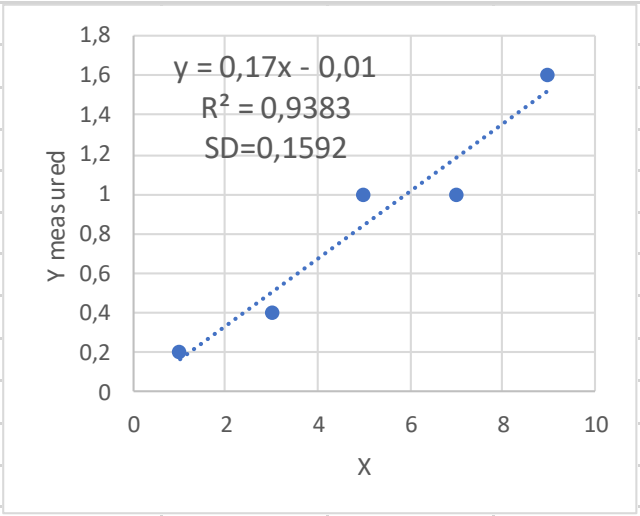
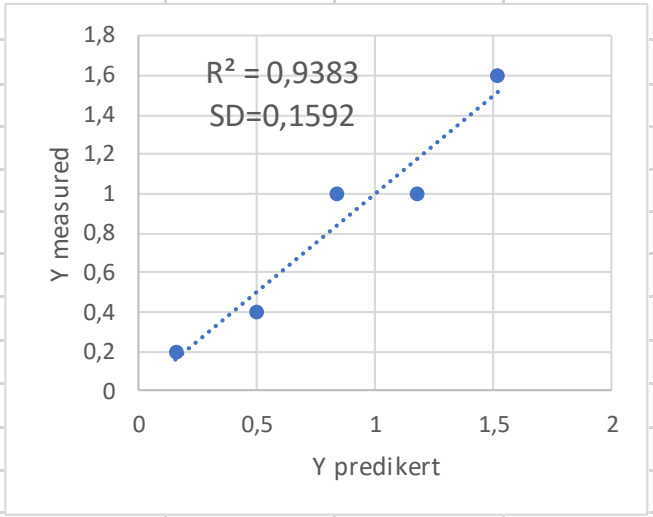
UiO

Two ‘similar’ ways of presenting X-Y correlation

X, mg/ml	Y (Measured abs)
1	0,2
3	0,4
5	1
7	1
9	1,6



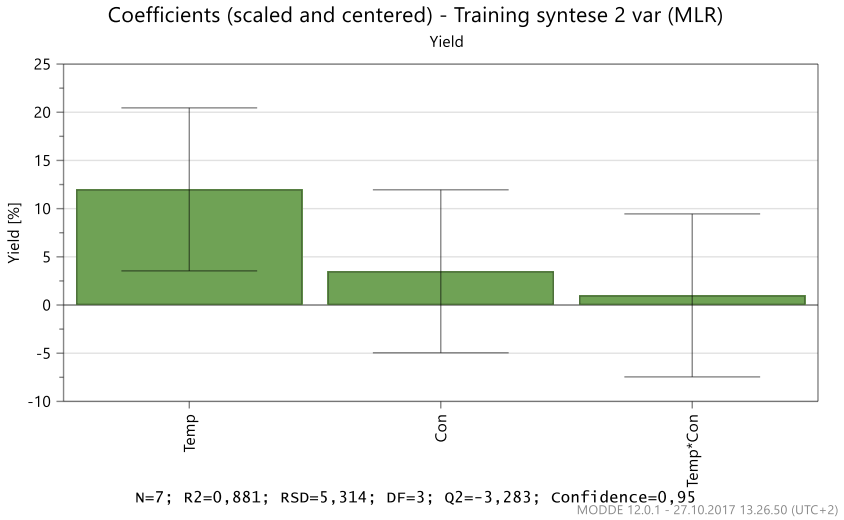
X, mg/ml	Slope (b)	Intercept (a)	Y predikert abs	Y (measured abs)
1	0,17	-0,01	0,16	0,2
3	0,17	-0,01	0,5	0,4
5	0,17	-0,01	0,84	1
7	0,17	-0,01	1,18	1
9	0,17	-0,01	1,52	1,6

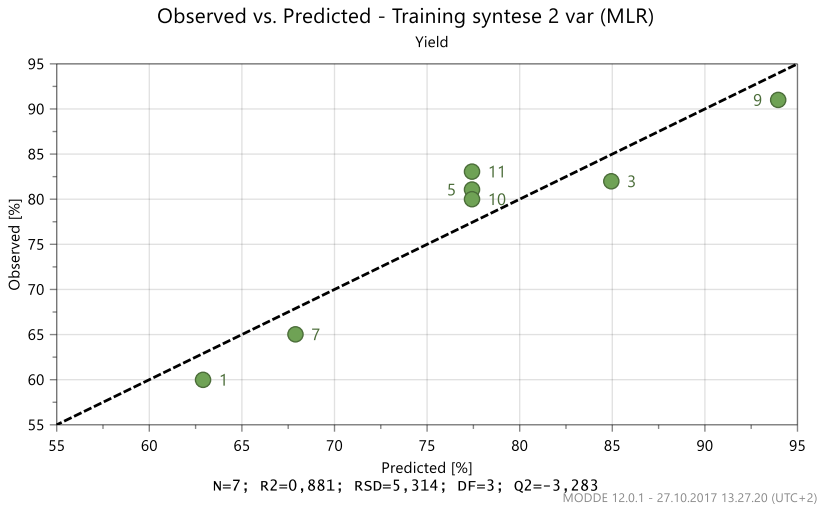


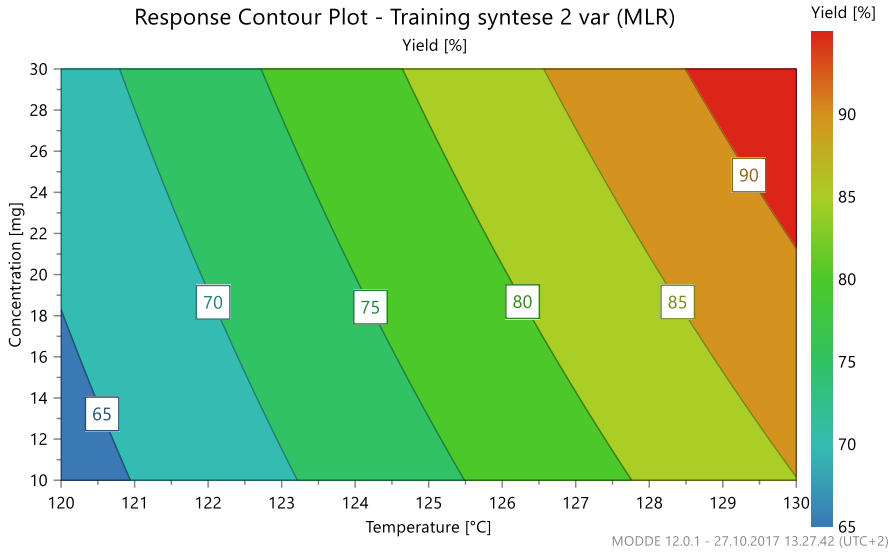
Knut Dyrstad

Regression analysis (MLR) of screening DOE

Experiment	Run	Temperature (°C)	Concentration (mg/ml)	Yield (%)
Screening	1	120	10	60
Screening	2	130	10	82
Screening	3	120	30	65
Screening	4	130	30	91
Screening, center	5	125	20	80
Screening, center	6	125	20	81
Screening, center	7	125	20	83







Regression analysis (MLR) of screening DOE, ANOVA table

Regression model sum of squares, SSmodel:
The sum of the squares of the deviations of the predicted values from the mean value of a response variable

	SS	DF	MS	F-ratio	p-value	B-coefficients	STDerr
Summary							
Model	629.000	3	209.667	7.425	0.0669		
Error	84.714	3	28.238				
Adjusted Total	713.714	6	118.952				
Variable							
Intercept	334.850	1	334.850	11.858	0.0411	-229.571	66.667
Temp	576.000	1	576.000	20.398	0.0203	2.400	0.531
Conc	49.000	1	49.000	1.735	0.2793	0.350	0.266
Temp*Conc	4.000	1	4.000	0.142	0.7317	0.667	1.771

Residual (error) sum of squares, SSerror:
The error sum of squares is the variation in the response Y that is not explained by the variables X. It is a measure of the discrepancy between the data and an the estimated model

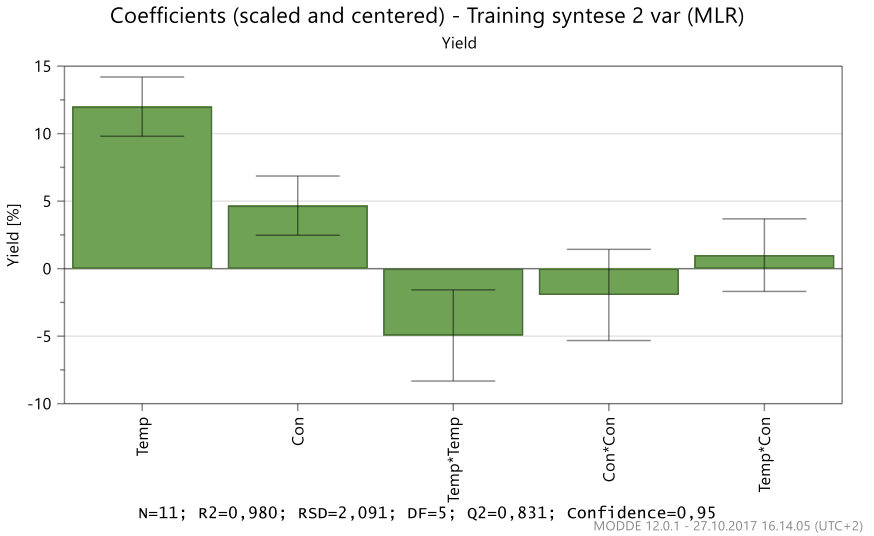
Total sum of squares, SStotal:
SStotal = SSmodel + SSerror = 713.7

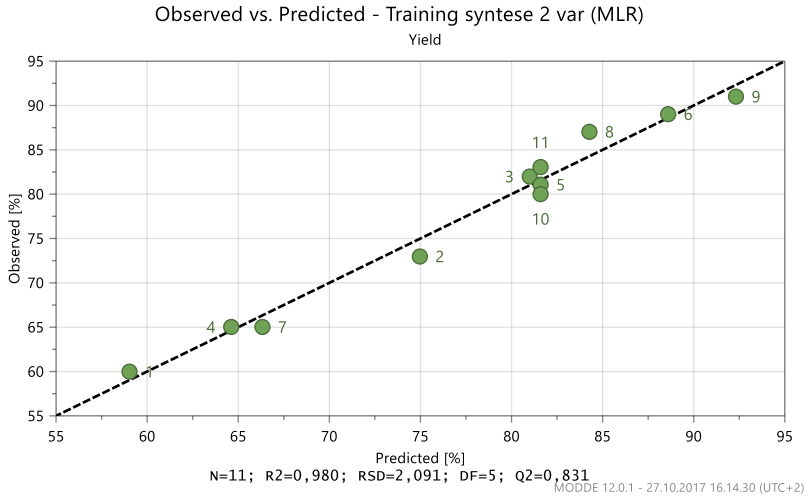
The mean square estimates:
MSmodel = SSmodel/df = 209.7
MSerror = SSerror/df =28.2
MStotal = SStotal = 119.0

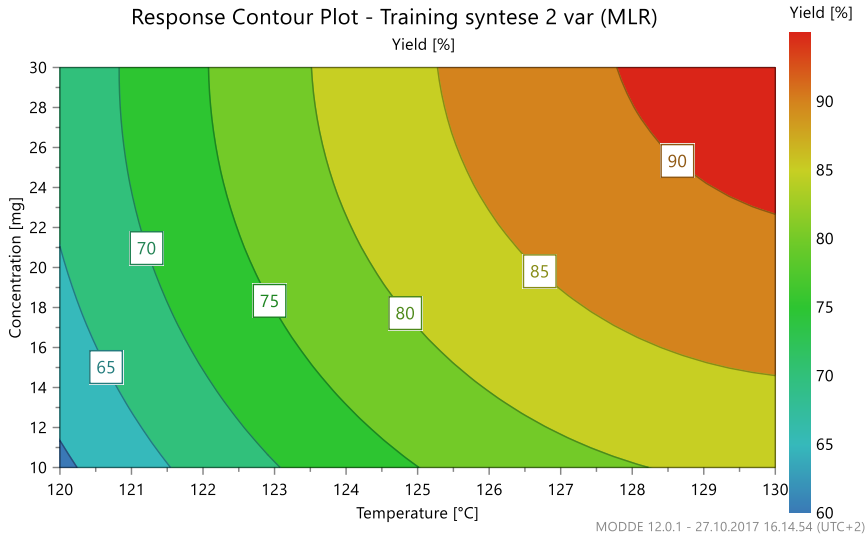
Significance of model (i.e. is the model describing much of Y variation or not):
F = MSmodel/MSerror =7.425
High F = significant model, i.e. MSmodel (and SSmodel) is large compared to MSerror (and SSerror)
Probability: If F is higher than table reference F for a selected probability level the model is ‘significant’

Regression analysis (MLR) of optimization DOE

Experiment	Run	Temperature (°C)	Concentration (mg/ml)	Yield (%)
Screening	1	120	10	60
Screening	2	130	10	82
Screening	3	120	30	65
Screening	4	130	30	91
Screening, center	5	125	20	80
Screening, center	6	125	20	81
Screening, center	7	125	20	83
Optimization	8	125	10	73
Optimization	9	125	30	87
Optimization	10	120	20	65
Optimization	11	130	20	89







Regression analysis (MLR) of optimization DOE, ANOVA table

	SS	DF	MS	F-ratio	p-value	B-coefficients	STDerr
Summary							
Model	1.090e+03	5	217.955	49.853	0.0003		
Error	21.860	5	4.372				
Adjusted Total	1.112e+03	10	111.164				
Variable							
Intercept	493.565	1	493.565	112.894	0.0001	-227.754	21.435
Temp	863.999	1	863.999	197.624	0.0000	2.400	0.171
Conc	130.666	1	130.666	29.888	0.0028	0.467	8.536e-02
Temp*Conc	4.000	1	4.000	0.915	0.3827	0.600	0.627
Temp**2	62.007	1	62.007	14.183	0.0131	-2.968	0.788
Conc**2	9.607	1	9.607	2.197	0.1983	-1.168	0.788

DOE randomization, block, fold-over

- Randomization: Randomize runs to avoid interference from a non-designed variables as build-up effects or other covariate effects
  - Nested design means that of different reasons, as practical ones, the DOE is not completely randomised
- Blocking: Divide the design in different blocks, often due to practical reasons (Different randomization within blocks may be selected, for example different randomization at day 1 and 2)
  - A block variable can be applied as a dummy variable checked for significance
- Fold-over: Extend the design and increase resolution by mirroring (multiply with -1) resolution III design
  - To understand the new confounding structure one can construct a reference design of similar size as the new design and study the confounding given by the stat software

One-way ANOVA (not mandatory, from excel)

	Replikat 1	Replikat 2	Replikat 3	Average
Analyse metode 1	80	79	80	79,66666667
Analyse metode 2	82	78	80	80
Analyse metode 3	86	85	83	84,66666667
Average	82,66666667	80,66666667	81	

One way anova

Variansanalyse: en-faktor

SAMMENDRAG

Grupper	*Antall*	*Sum*	*Gjennomsnitt*	*Varians*
Analyse metode 1	3	239	79,66666667	0,333333333
Analyse metode 2	3	240	80	4
Analyse metode 3	3	254	84,66666667	2,333333333

Variansanalyse

Variasjonskilde	*SK*	*fg*	*GK*	*F*	*P-verdi*	*F-krit*
Mellom grupper	46,88888889	2	23,44444444	10,55	0,010852902	5,14325285
Innenfor grupper	13,33333333	6	2,222222222			
Totalt	60,22222222	8				

The variation between the analytical methods is significant greater than the variation among replicates (P=0.0108)

Two-way ANOVA (not mandatory, from excel)

	Batch 1	Batch 2	Batch 3	Average
Analyse metode 1	80	79	80	79,66666667
Analyse metode 2	82	78	80	80
Analyse metode 3	86	85	83	84,66666667
Average	82,66666667	80,66666667	81	

Two way ANOVA

Variansanalyse: To-faktor uten tilbakelegging

SAMMENDRAG	*Antall*	*Sum*	*Gjennomsnitt*	*Varians*
Analyse metode 1	3	239	79,66666667	0,333333333
Analyse metode 2	3	240	80	4
Analyse metode 3	3	254	84,66666667	2,333333333
Batch 1	3	248	82,66666667	9,333333333
Batch 2	3	242	80,66666667	14,33333333
Batch 3	3	243	81	3

Variansanalyse

Variasjonskilde	*SK*	*fg*	*GK*	*F*	*P-verdi*	*F-krit*
Rader (analyse metode)	46,88888889	2	23,44444444	14,55172414	0,014600694	6,94427191
Kolonner (Batcher)	6,888888889	2	3,444444444	2,137931034	0,233611111	6,94427191
Feil	6,444444444	4	1,611111111			
Totalt	60,22222222	8				

The difference between the analytical methods is significant (P=0.0146) while the difference between batches is not (P=0.234) when compared to the error (given by GK, i.e. mean square)