Application of central composite design for optimization of the removal of humic substances using coconut copra

The objective of this research paper is to evaluate the optimum conditions for the removal of humic substances from peat swamp runoff using modified coconut copra by applying Design of experiments (DoE) methodology.

The factors (independent variables) examined are: X_1 = dosage (g), X_2 = contact time (min) and X_3 = temperature (°C). All the factors are continuous. The response (dependent variable) examined is: Y = removal efficiency (%). The applied DoE method is Inscribed Central Composite design.

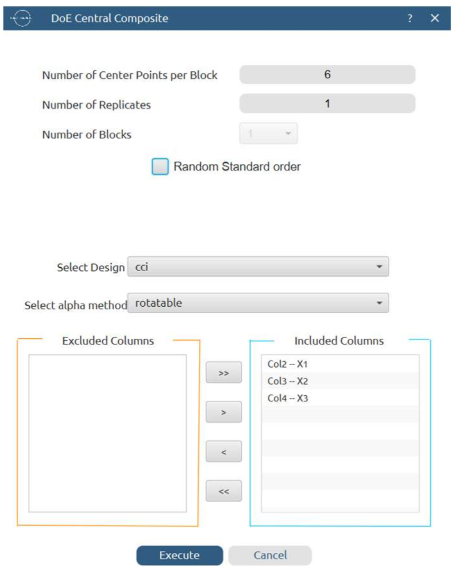
Isalos version used: 2.0.6

Scientific article: https://link.springer.com/article/10.1007/s40090-015-0041-0

Step 1: Central Composite Design

In the first tab named "Action" define the factors in the column headers and fill each column with the low and high levels of the corresponding factors. This tab can be renamed "CCI". Afterwards, apply the Inscribed Central Composite method: *DOE → Response Surface → Central Composite*

	Col1	Col2 (I)	Col3 (I)	Col4 (I)
User Header	User Row ID	X1	X2	X3
1		1	15	30
2		5	60	70



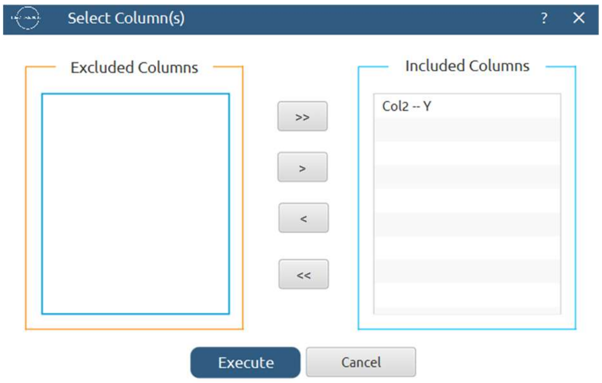
Results (right spreadsheet):

	Col1	Col2 (I)	Col3 (S)	Col4 (S)	Col5 (S)	Col6 (D)	Col7 (D)	Col8 (D)
User Header	User Row ID	Standard Order	Block Number	Replicate Number	Point Type	X1	X2	X3
1		1	Block: 1	Replicate: 1	Design Point	1.8107929	24.1214200	38.1079288
2		2	Block: 1	Replicate: 1	Design Point	4.1892071	24.1214200	38.1079288
3		3	Block: 1	Replicate: 1	Design Point	1.8107929	50.8785800	38.1079288
4		4	Block: 1	Replicate: 1	Design Point	4.1892071	50.8785800	38.1079288
5		5	Block: 1	Replicate: 1	Design Point	1.8107929	24.1214200	61.8920712
6		6	Block: 1	Replicate: 1	Design Point	4.1892071	24.1214200	61.8920712
7		7	Block: 1	Replicate: 1	Design Point	1.8107929	50.8785800	61.8920712
8		8	Block: 1	Replicate: 1	Design Point	4.1892071	50.8785800	61.8920712
9		9	Block: 1	Replicate: 1	Design Point	1.0	37.5	50.0
10		10	Block: 1	Replicate: 1	Design Point	5.0	37.5	50.0
11		11	Block: 1	Replicate: 1	Design Point	3.0	15.0	50.0
12		12	Block: 1	Replicate: 1	Design Point	3.0	60.0	50.0
13		13	Block: 1	Replicate: 1	Design Point	3.0	37.5	30.0
14		14	Block: 1	Replicate: 1	Design Point	3.0	37.5	70.0
15		15	Block: 1	----	Center Point	3.0	37.5	50.0
16		16	Block: 1	----	Center Point	3.0	37.5	50.0
17		17	Block: 1	----	Center Point	3.0	37.5	50.0
18		18	Block: 1	----	Center Point	3.0	37.5	50.0
19		19	Block: 1	----	Center Point	3.0	37.5	50.0
20		20	Block: 1	----	Center Point	3.0	37.5	50.0

Step 2: Definition of response variables

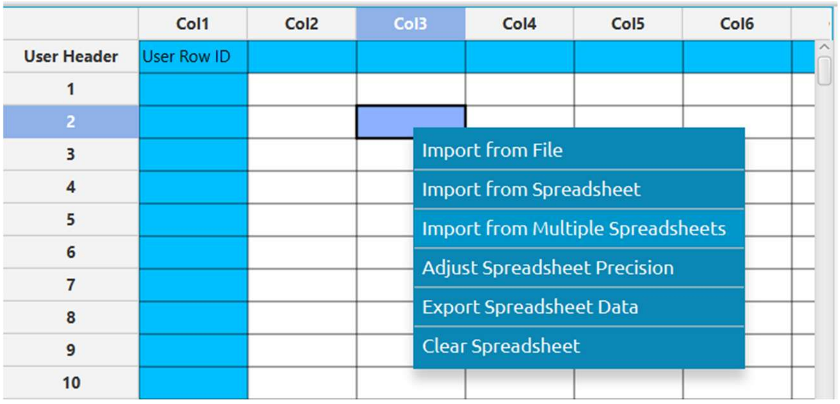
Create a new tab named "Responses" and define the responses in the column headers. Fill each column with the values of the corresponding responses that were observed and make sure the values follow the order of the experiments as given by the Inscribed Central Composite method. Then, select all columns to be transferred to the right spreadsheet: *Data Transformation → Data Manipulation → Select Column(s)*

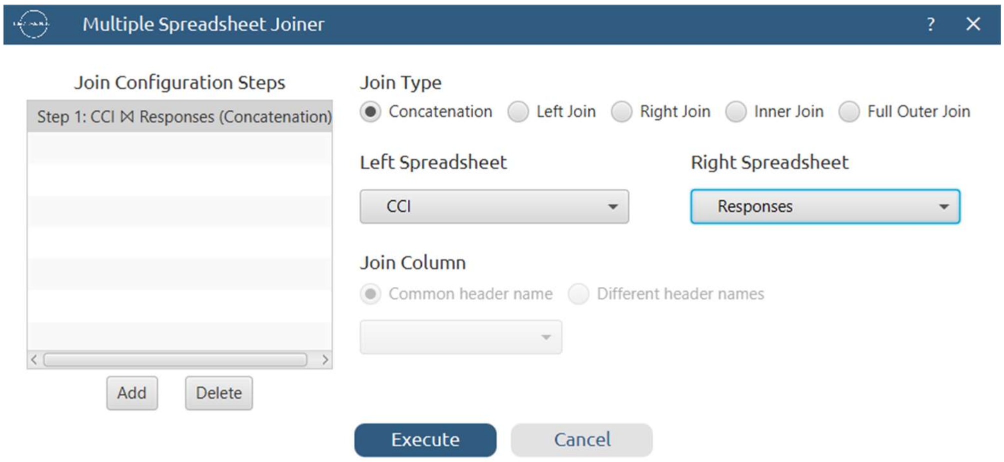
	Col1	Col2 (D)
User Header	User Row ID	Y
1		55.8
2		81.07
3		59.39
4		79.01
5		60.74
6		84.94
7		60.36
8		85.53
9		40.05
10		89.03
11		69.71
12		79.93
13		72.13
14		73.46
15		78.08
16		78.19
17		77.98
18		80.41
19		80.3
20		80.23

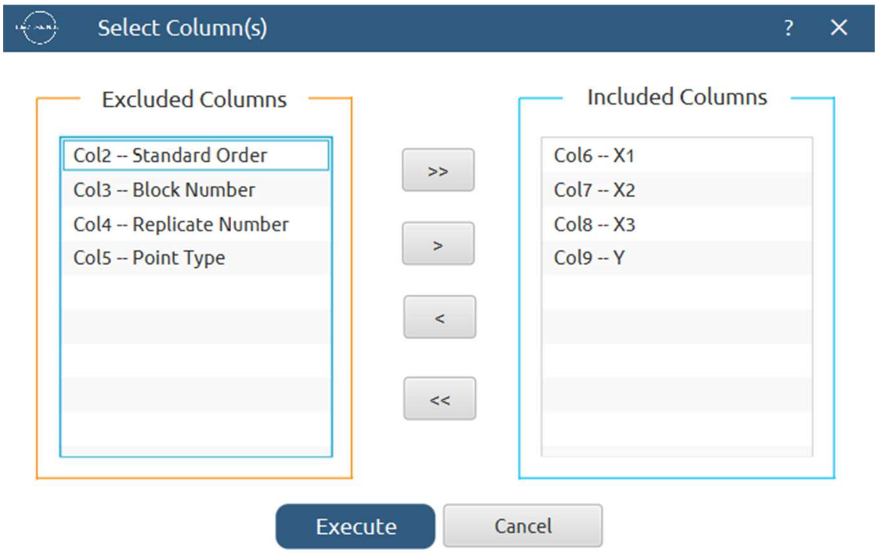


Step 3: Data isolation

Create a new tab named “Data” and import the results from the “CCI” and “Responses” spreadsheets by right clicking on the left spreadsheet. Then, select only the factors and responses columns to be transferred to the right spreadsheet: *Data Transformation → Data Manipulation → Select Column(s)*





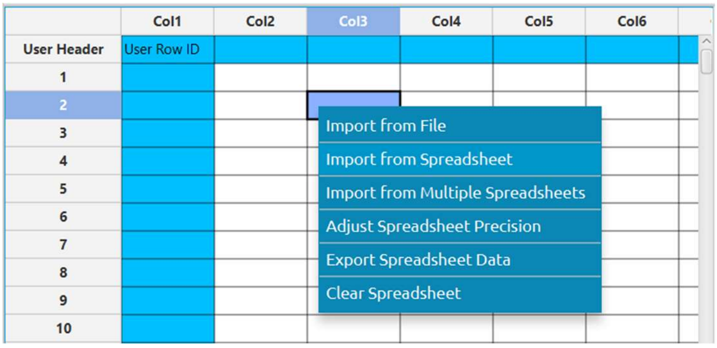


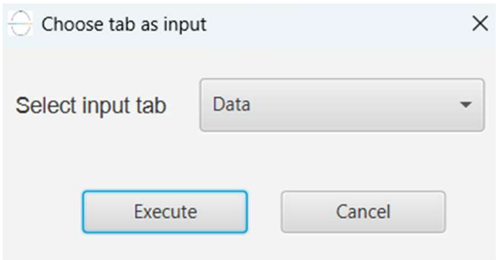
Results:

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	X1	X2	X3	Y
1		1.8107929	24.1214200	38.1079288	55.8
2		4.1892071	24.1214200	38.1079288	81.07
3		1.8107929	50.8785800	38.1079288	59.39
4		4.1892071	50.8785800	38.1079288	79.01
5		1.8107929	24.1214200	61.8920712	60.74
6		4.1892071	24.1214200	61.8920712	84.94
7		1.8107929	50.8785800	61.8920712	60.36
8		4.1892071	50.8785800	61.8920712	85.53
9		1.0	37.5	50.0	40.05
10		5.0	37.5	50.0	89.03
11		3.0	15.0	50.0	69.71
12		3.0	60.0	50.0	79.93
13		3.0	37.5	30.0	72.13
14		3.0	37.5	70.0	73.46
15		3.0	37.5	50.0	78.08
16		3.0	37.5	50.0	78.19
17		3.0	37.5	50.0	77.98
18		3.0	37.5	50.0	80.41
19		3.0	37.5	50.0	80.3
20		3.0	37.5	50.0	80.23

Step 4: Normalization

Create a new tab named “Normalized data” and import the results from the “Data” spreadsheet. Afterwards, normalize the factor columns to take values in the range [-1, 1]: *Data Transformation → Normalizers → Min-Max*





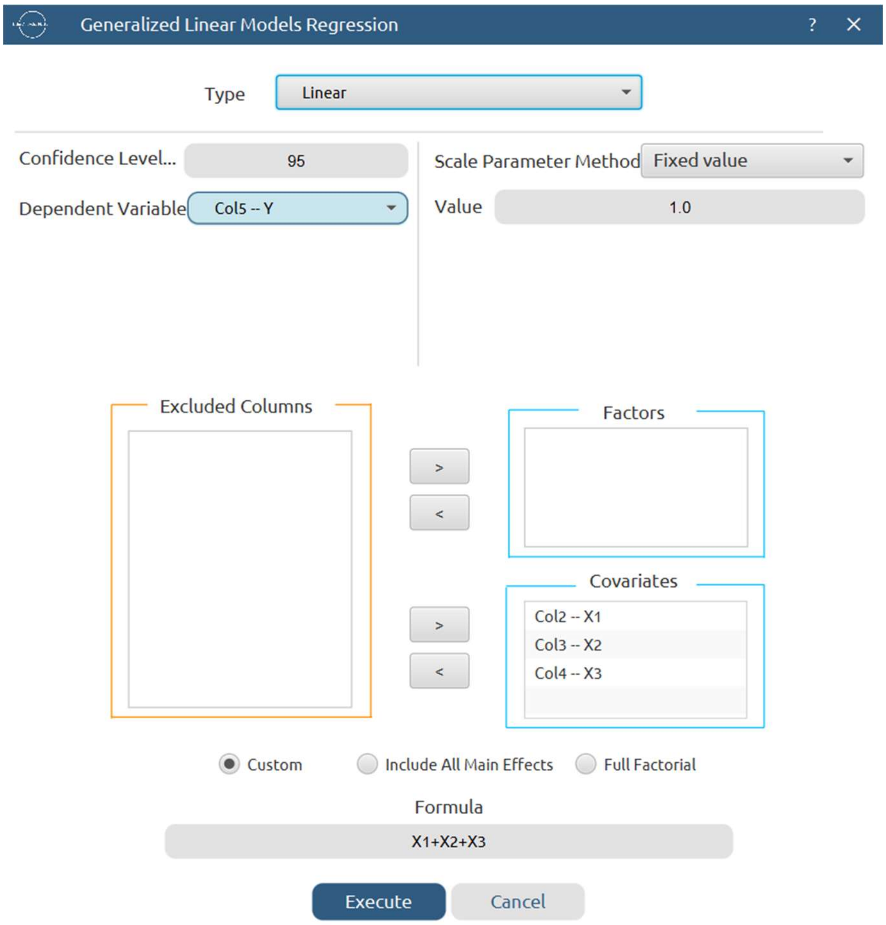
Results:

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	X1	X2	X3	Y
1		-0.5946036	-0.5946036	-0.5946036	55.8
2		0.5946036	-0.5946036	-0.5946036	81.07
3		-0.5946036	0.5946036	-0.5946036	59.39
4		0.5946036	0.5946036	-0.5946036	79.01
5		-0.5946036	-0.5946036	0.5946036	60.74
6		0.5946036	-0.5946036	0.5946036	84.94
7		-0.5946036	0.5946036	0.5946036	60.36
8		0.5946036	0.5946036	0.5946036	85.53
9		-1.0	0.0	0.0	40.05
10		1.0	0.0	0.0	89.03
11		0.0	-1.0	0.0	69.71
12		0.0	1.0	0.0	79.93
13		0.0	0.0	-1.0	72.13
14		0.0	0.0	1.0	73.46
15		0.0	0.0	0.0	78.08
16		0.0	0.0	0.0	78.19
17		0.0	0.0	0.0	77.98
18		0.0	0.0	0.0	80.41
19		0.0	0.0	0.0	80.3
20		0.0	0.0	0.0	80.23

Step 5: Regression – Linear

The goal here is to produce a regression equation that includes only the main effects for the response variable Y: $Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$

Create a new tab named "Regression – Linear" and import the results from the spreadsheet "Normalized data". Afterwards, fit a generalized linear model to the data: *Analytics → Regression → Statistical fitting → Generalized Linear Models*



Results:

Y	Prediction
55.8	57.6399695
81.07	83.5074511
59.39	60.4118996
79.01	86.2793812
60.74	60.3546188
84.94	86.2221004
60.36	63.1265489
85.53	88.9940305
40.05	51.5651275
89.03	95.0688725
69.71	70.9860939
79.93	75.6479061
72.13	71.0342612
73.46	75.5997388
78.08	73.3170000
78.19	73.3170000
77.98	73.3170000
80.41	73.3170000
80.3	73.3170000
80.23	73.3170000

Goodness of Fit	
	Value
Deviance	494.5088236
Scaled Deviance	494.5088236
Pearson Chi-Square	494.5088236
Scaled Pearson Chi-Square	494.5088236
Log Likelihood	-265.6331825
Akaike's Information Criterion (AIC)	539.2663649
Finite Sample Corrected AIC (AICC)	541.9330316
Bayesian Information Criterion (BIC)	543.2492940
Consistent AIC (CAIC)	547.2492940

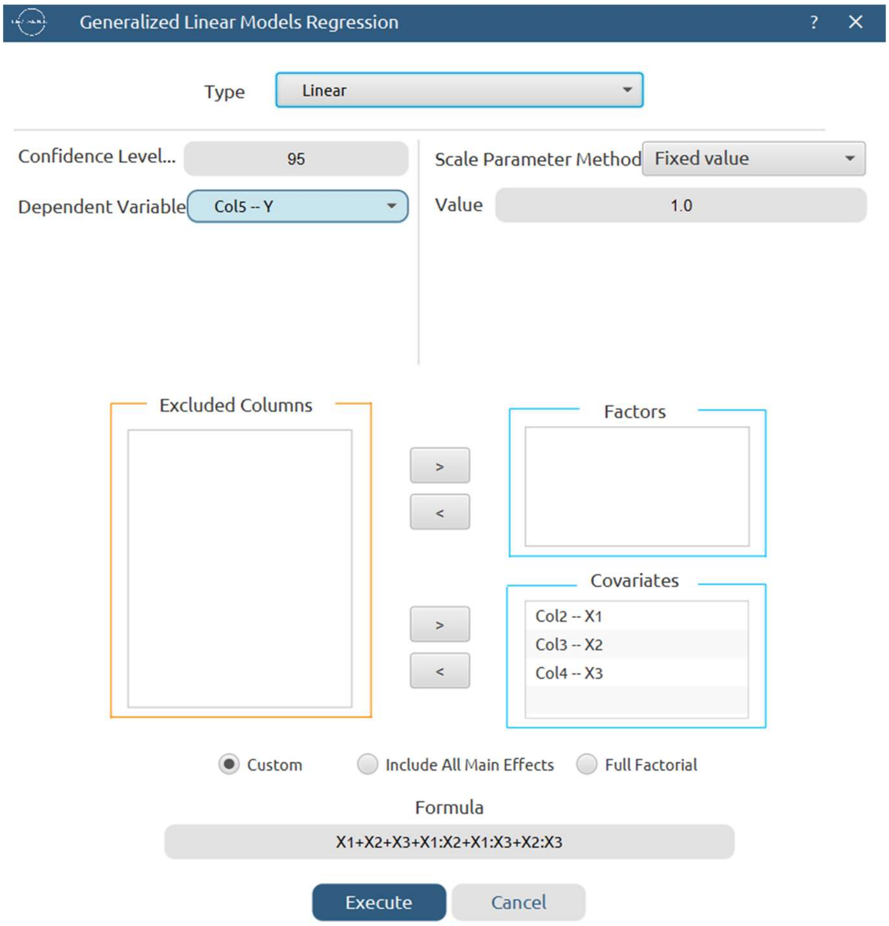
Parameter Estimates							
Variable	Coefficient	Std. Error	Lower CI	Upper CI	Test Statistic	df	p-value
intercept	73.3170000	0.2236068	72.8787387	73.7552613	107507.6497800	1	0.0
X1	21.7518725	0.4550899	20.8599128	22.6438323	2284.5411231	1	0.0
X2	2.3309061	0.4550899	1.4389464	3.2228658	26.2334394	1	3E-7
X3	2.2827388	0.4550899	1.3907791	3.1746985	25.1604339	1	5E-7

Step 6: Regression – Interactions

The goal here is to produce a regression equation that includes main effects and two-factor interactions for the response variable Y:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3$$

Create a new tab named "Regression – Int" and import the results from the spreadsheet "Normalized data". Afterwards, fit a generalized linear model to the data: *Analytics → Regression → Statistical fitting → Generalized Linear Models*



Results:

Y	Prediction
55.8	57.4499695
81.07	83.3674511
59.39	61.7218996
79.01	85.2993812
60.74	59.3746188
84.94	87.5321004
60.36	62.9865489
85.53	88.8040305
40.05	51.5651275
89.03	95.0688725
69.71	70.9860939
79.93	75.6479061
72.13	71.0342612
73.46	75.5997388
78.08	73.3170000
78.19	73.3170000
77.98	73.3170000
80.41	73.3170000
80.3	73.3170000
80.23	73.3170000

Goodness of Fit	
	Value
Deviance	489.0444236
Scaled Deviance	489.0444236
Pearson Chi-Square	489.0444236
Scaled Pearson Chi-Square	489.0444236
Log Likelihood	-262.9009825
Akaike's Information Criterion (AIC)	539.8019649
Finite Sample Corrected AIC (AICC)	549.1352983
Bayesian Information Criterion (BIC)	546.7720908
Consistent AIC (CAIC)	553.7720908

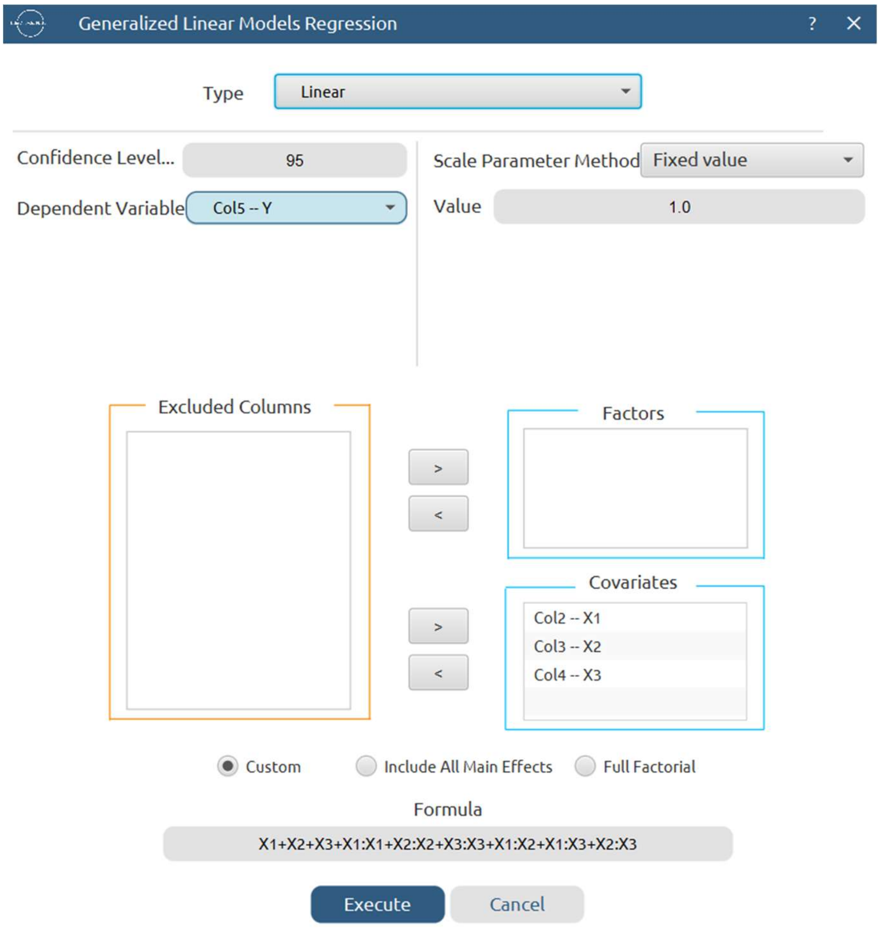
Parameter Estimates							
Variable	Coefficient	Std. Error	Lower CI	Upper CI	Test Statistic	df	p-value
intercept	73.3170000	0.2236068	72.8787387	73.7552613	107507.6497800	1	0.0
X1	21.7518725	0.4550899	20.8599128	22.6438323	2284.5411231	1	0.0
X2	2.3309061	0.4550899	1.4389464	3.2228658	26.2334394	1	3E-7
X3	2.2827388	0.4550899	1.3907791	3.1746985	25.1604339	1	5E-7
X1*X3	1.5839192	1.0000000	-0.3760448	3.5438832	2.5088000	1	0.1132121
X1*X2	-1.6546299	1.0	-3.6145939	0.3053341	2.7378000	1	0.0979996
X2*X3	-0.4666905	1.0	-2.4266545	1.4932735	0.2178000	1	0.6407213

Step 7: Regression – Quadratic

The goal here is to produce a regression equation that includes main effects, two-factor interactions and quadratic effects for the response variable Y:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{11}{X_1}^2 + b_{22}{X_2}^2 + b_{33}{X_3}^2$$

Create a new tab named "Regression – Quad" and import the results from the spreadsheet "Normalized data". Afterwards, fit a generalized linear model to the data: *Analytics → Regression → Statistical fitting → Generalized Linear Models*



Results:

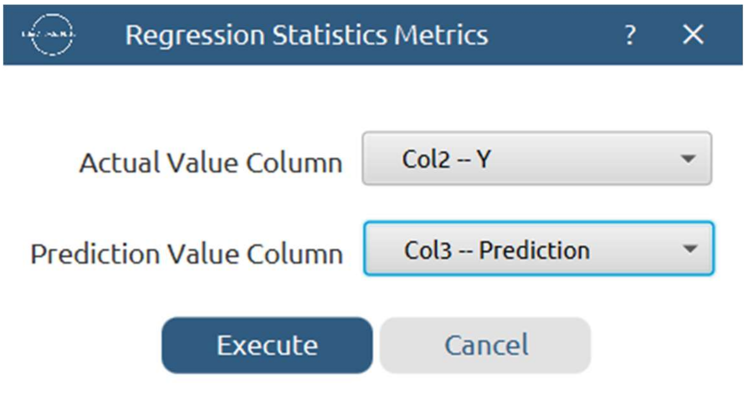
Y	Prediction
55.8	54.7280537
81.07	80.6455352
59.39	58.9999838
79.01	82.5774653
60.74	56.6527029
84.94	84.8101845
60.36	60.2646330
85.53	86.0821146
40.05	43.1557040
89.03	86.6594491
69.71	72.8566705
79.93	77.5184827
72.13	70.8798378
73.46	75.4453154
78.08	79.1773113
78.19	79.1773113
77.98	79.1773113
80.41	79.1773113
80.3	79.1773113
80.23	79.1773113

Goodness of Fit	
	Value
Deviance	75.2317139
Scaled Deviance	75.2317139
Pearson Chi-Square	75.2317139
Scaled Pearson Chi-Square	75.2317139
Log Likelihood	-55.9946276
Akaike's Information Criterion (AIC)	131.9892552
Finite Sample Corrected AIC (AICC)	156.4336997
Bayesian Information Criterion (BIC)	141.9465779
Consistent AIC (CAIC)	151.9465779

Parameter Estimates							
Variable	Coefficient	Std. Error	Lower CI	Upper CI	Test Statistic	df	p-value
intercept	79.1773113	0.4078483	78.3779433	79.9766792	37688.0971269	1	0.0
X1	21.7518725	0.4550899	20.8599128	22.6438323	2284.5411231	1	0.0
X2	2.3309061	0.4550899	1.4389464	3.2228658	26.2334394	1	3E-7
X3	2.2827388	0.4550899	1.3907791	3.1746985	25.1604339	1	5E-7
X1*X3	1.5839192	1.0000000	-0.3760448	3.5438832	2.5088000	1	0.1132121
X1*X2	-1.6546299	1.0	-3.6145939	0.3053341	2.7378000	1	0.0979996
X2*X3	-0.4666905	1.0	-2.4266545	1.4932735	0.2178000	1	0.6407213
X1*X1	-14.2697347	0.7450640	-15.7300332	-12.8094362	366.8129730	1	0.0
X2*X2	-3.9897347	0.7450640	-5.4500332	-2.5294362	28.6748349	1	1E-7
X3*X3	-6.0147347	0.7450640	-7.4750332	-4.5544362	65.1697180	1	0E-7

Step 8: Regression Metrics

Create a tab named "Metrics – Linear" and import the results from the spreadsheet "Regression – Linear". Then, produce the regression metrics for the linear regression equation: *Statistics → Model Metrics → Regression Metrics*



Results:

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared
1		24.7254412	4.9724683	4.1051214	0.8252893

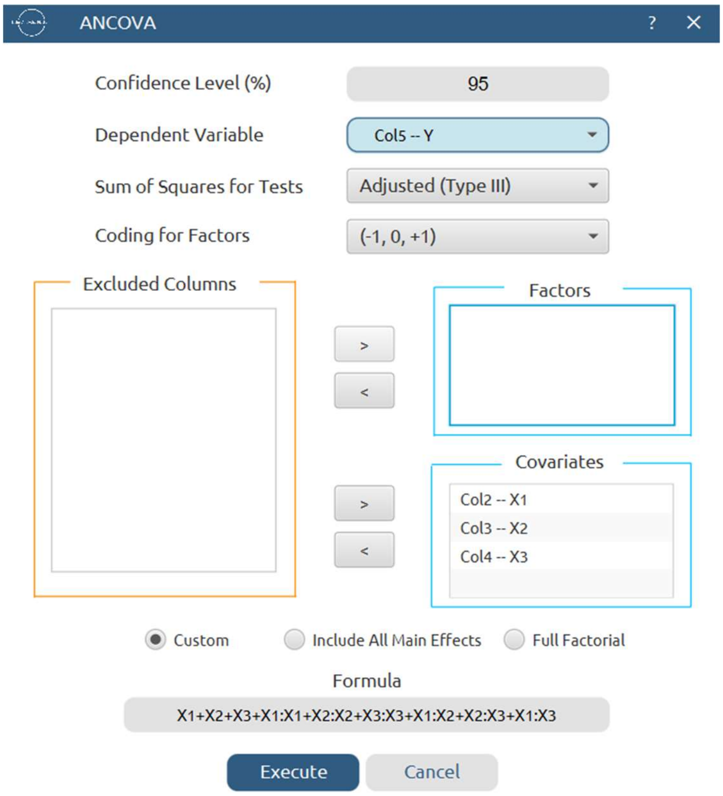
Repeat this step for the interactions and quadratic regression equations. Results:

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared
1		24.4522212	4.9449187	4.2031214	0.8272199

	Col1	Col2 (D)	Col3 (D)	Col4 (D)	Col5 (D)
User Header	User Row ID	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R Squared
1		3.7615857	1.9394808	1.5639204	0.9734205

Step 9: Analysis of Covariance

Create a new tab named "ANCOVA – Quad" and import the results from the spreadsheet "Normalized data". Afterwards perform analysis of covariance for Y using the formula for the quadratic equation: *Statistics → Analysis of (Co)Variance → ANCOVA*

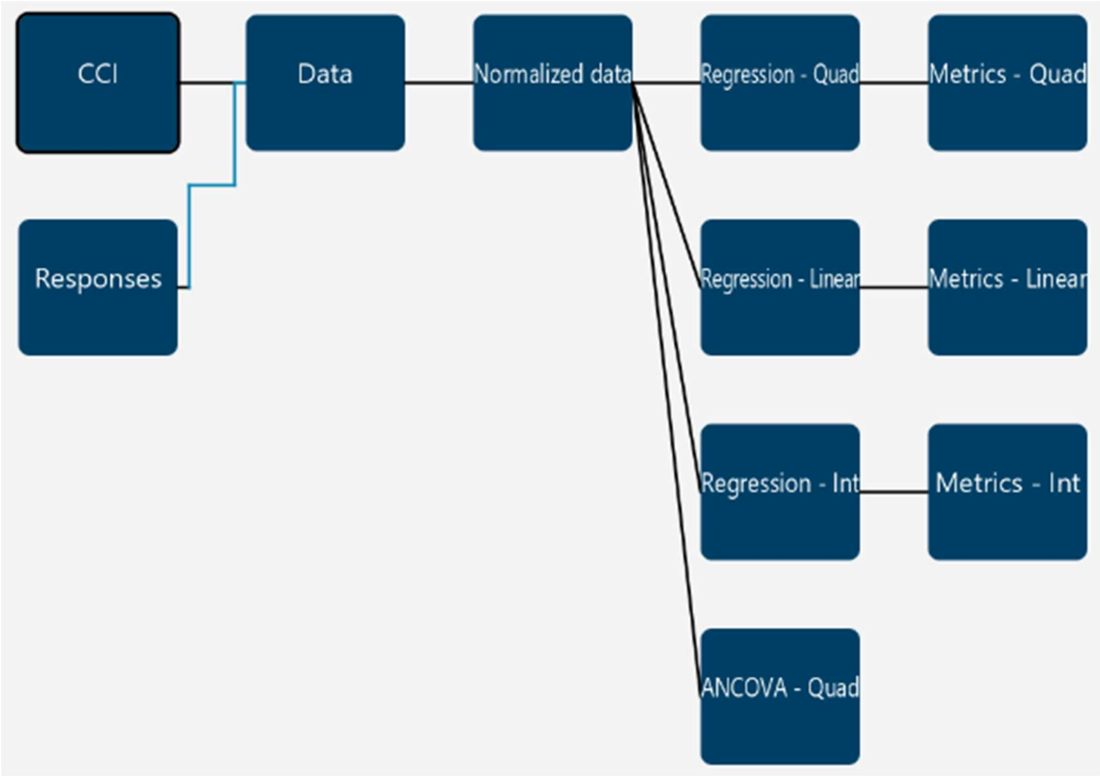


Results:

	Col1	Col2 (S)	Col3 (I)	Col4 (D)	Col5 (D)	Col6 (D)	Col7 (D)
User Header	User Row ID	Source	DF	Adj SS	Adj MS	F-Value	P-Value
1		X1	1	2284.5411231	2284.5411231	303.6672974	0E-7
2		X2	1	26.2334394	26.2334394	3.4870187	0.0914042
3		X3	1	25.1604339	25.1604339	3.3443920	0.0973718
4		X1*X1	1	366.8129730	366.8129730	48.7577584	0.0000379
5		X2*X2	1	28.6748349	28.6748349	3.8115355	0.0794427
6		X3*X3	1	65.1697180	65.1697180	8.6625327	0.0147043
7		X1*X2	1	2.7378000	2.7378000	0.3639157	0.5597725
8		X1*X3	1	2.5088000	2.5088000	0.3334764	0.5763892
9		X2*X3	1	0.2178000	0.2178000	0.0289506	0.8682863
10		Error	10	75.2317139	7.5231714		
11		Total	19	2830.4438200			

Final Isalos Workflow

The final workflow is presented below:



References

(1) Lee, T. Z. E.; Krongchai, C.; Mohd Irwn Lu, N. A. L.; Kittiwachana, S.; Sim, S. F. Application of Central Composite Design for Optimization of the Removal of Humic Substances Using Coconut Copra. *Int J Ind Chem* **2015**, *6* (3), 185–191. https://doi.org/10.1007/s40090-015-0041-0.