Multivariate Analysis and Design of Experiments in practice using The Unscrambler® X

Frank Westad

CAMO Software

fw@camo.com

Pat Whitcomb

Stat-Ease

pat@statease.com

Stat-Ease
statistics made easy®

Agenda

Goal: Show how Multivariate Analysis (MVA) and Design of Experiments (DOE) can be used together.

Part 1: *Frank Westad from CAMO Software*
Use Unscrambler® X version 10.4 to model properties of 45 organic solvents using Principal Component Analysis (PCA).

Part 2: *Pat Whitcomb from Stat-Ease*
Use Design-Expert® version 10 to build an optimal design using the principle components, simulate results, analyze and optimize.

Stat-Ease statistics made easy®

CAMO Bring data to life

Part 1: Properties for 45 Organic Solvents

Common Organic Solvents:

- Table of Properties[1,2,3]

https://www.organicdivision.org/orig/organic_solvents.html

Solvent	formula	MW	boiling point (°C)	melting point (°C)	density (g/mL)	solubility in water (g/100g)	Dielectric Constant[3,4]	flash point (°C)
acetic acid	$C_2H_4O_2$	60.052	118	16.6	1.0446	Miscible	6.20	39
acetone	C_3H_6O	58.079	56.05	-94.7	0.7845	Miscible	21.01	-20
acetonitrile	C_2H_3N	41.052	81.65	-43.8	0.7857	Miscible	36.64	6
benzene	C_6H_6	78.11	80.1	5.5	0.8765	0.18	2.28	-11
1-butanol	$C_4H_{10}O$	74.12	117.7	-88.6	0.8095	6.3	17.8	37
2-butanol	$C_4H_{10}O$	74.12	99.5	-88.5	0.8063	15	17.26	24
2-butanone	C_4H_8O	72.11	79.6	-86.6	0.7999	25.6	18.6	-9
t-butyl alcohol	$C_4H_{10}O$	74.12	82.4	25.7	0.7887	Miscible	12.5	11
carbon tetrachloride	CCl_4	153.82	76.8	-22.6	1.594	0.08	2.24	--
chlorobenzene	C_6H_5Cl	112.56	131.7	-45.3	1.1058	0.05	5.69	28
chloroform	$CHCl_3$	119.38	61.2	-63.4	1.4788	0.795	4.81	--
cyclohexane	C_6H_{12}	84.16	80.7	6.6	0.7739	<0.1	2.02	-20
1,2-dichloroethane	$C_2H_4Cl_2$	98.96	83.5	-35.7	1.245	0.861	10.42	13
diethylene glycol	$C_4H_{10}O_3$	106.12	246	-10	1.1197	10	31.8	124
diethyl ether	$C_4H_{10}O$	74.12	34.5	-116.2	0.713	7.5	4.267	-45
diglyme (diethylene glycol dimethyl ether)	$C_6H_{14}O_3$	134.17	162	-68	0.943	Miscible	7.23	67
1,2-dimethoxy-ethane (glyme, DME)	$C_4H_{10}O_2$	90.12	84.5	-69.2	0.8637	Miscible	7.3	-2
dimethyl-formamide (DMF)	C_3H_7NO	73.09	153	-60.48	0.9445	Miscible	38.25	58
dimethyl sulfoxide (DMSO)	C_2H_6OS	78.13	189	18.4	1.092	25.3	47	95
1,4-dioxane	$C_4H_8O_2$	88.11	101.1	11.8	1.033	Miscible	2.21(25)	12
ethanol	C_2H_6O	46.07	78.5	-114.1	0.789	Miscible	24.6	13
ethyl acetate	$C_4H_8O_2$	88.11	77	-83.6	0.895	8.7	6(25)	-4
ethylene glycol	$C_2H_6O_2$	62.07	195	-13	1.115	Miscible	37.7	111
glycerin	$C_3H_8O_3$	92.09	290	17.8	1.261	Miscible	42.5	160

Notes:

1. This table was originally from: Prof. Murov's Orgsoltab, which was edited and reposted by Erowid
2. You can find more detailed information (Health & Safety, Physical, Regulatory, Environmental) on various organic solvents from NCMS
3. The values in the table above were obtained from the CRC (87th edition), or Vogel's *Practical Organic Chemistry* (5th ed.).
4. T = 20 °C unless specified otherwise.

The data table, Solubility was represented as three dummy variables

Solvents [Solvents no results.unsb] - The Unscrambler X

File Edit View Insert Plot Tasks Tools Help

Solvents
Table
Columnset
Properties

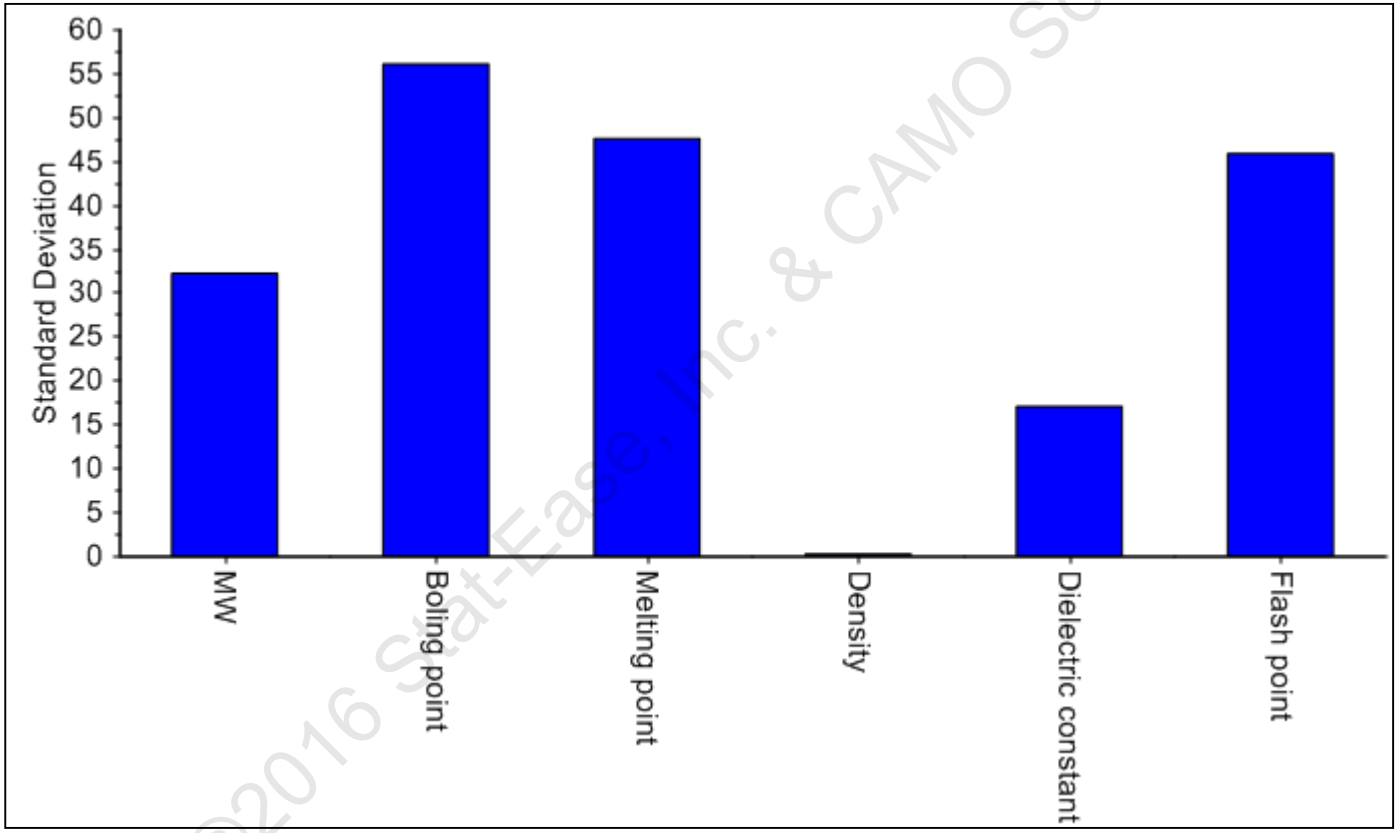
Table			MW	Boiling point	Melting point	Density	Solubility	Insoluble	Miscible	Soluble	Solubility typ	Dielectric con	Flash point
			1	2	3	4	5	6	7	8	9	10	11
acetic acid	C2H4O2	1	60.052000	118.000000	16.600000	1.044600	Miscible	0.000000	1.000000	0.000000	Miscible	6.200000	39.000000
acetone	C3H6O	2	58.079000	56.050000	-94.700000	0.784500	Miscible	0.000000	1.000000	0.000000	Miscible	21.010000	-20.000000
acetonitrile	C2H3N	3	41.052000	81.650000	-43.800000	0.785700	Miscible	0.000000	1.000000	0.000000	Miscible	36.640000	6.000000
benzene	C6H6	4	78.110000	80.100000	5.500000	0.876500	0.18	0.000000	0.000000	1.000000	Soluble	2.280000	-11.000000
1-butanol	C4H10O	5	74.120000	117.700000	-88.600000	0.809500	6.3	0.000000	0.000000	1.000000	Soluble	17.800000	37.000000
2-butanol	C4H10O	6	74.120000	99.500000	-88.500000	0.806300	15	0.000000	0.000000	1.000000	Soluble	17.260000	24.000000
2-butanone	C4H8O	7	72.110000	79.600000	-86.600000	0.799900	25.6	0.000000	0.000000	1.000000	Soluble	18.600000	-9.000000
t-butyl alcoho	C4H10O	8	74.120000	82.400000	25.700000	0.788700	Miscible	0.000000	1.000000	0.000000	Miscible	12.500000	11.000000
carbon tetrac	CCl4	9	153.820000	76.800000	-22.600000	1.594000	0.08	0.000000	0.000000	1.000000	Soluble	2.240000	
chlorobenzen	C6H5Cl	10	112.560000	131.700000	-45.300000	1.105800	0.05	0.000000	0.000000	1.000000	Soluble	5.690000	28.000000
chloroform	CHCl3	11	119.380000	61.200000	-63.400000	1.478800	0.795	0.000000	0.000000	1.000000	Soluble	4.810000	
cyclohexane	C6H12	12	84.160000	80.700000	6.600000	0.773900	<0.1	0.000000	0.000000	1.000000	Soluble	2.020000	-20.000000
1,2-dichloroe	C2H4Cl2	13	98.960000	83.500000	-35.700000	1.245000	0.861	0.000000	0.000000	1.000000	Soluble	10.420000	13.000000
diethylene gl	C4H10O3	14	106.120000	246.000000	-10.000000	1.119700	10	0.000000	0.000000	1.000000	Soluble	31.800000	124.000000
diethyl ether	C4H10O	15	74.120000	34.500000	-116.200000	0.713000	7.5	0.000000	0.000000	1.000000	Soluble	4.267000	-45.000000
diglyme (diet	C6H14O3	16	134.170000	162.000000	-68.000000	0.943000	Miscible	0.000000	1.000000	0.000000	Miscible	7.230000	67.000000
1,2-dimethox	C4H10O2	17	90.120000	84.500000	-69.200000	0.863700	Miscible	0.000000	1.000000	0.000000	Miscible	7.300000	-2.000000
dimethyl-	C3H7NO	18	73.090000	153.000000	-60.480000	0.944500	Miscible	0.000000	1.000000	0.000000	Miscible	38.250000	58.000000
dimethyl sulf	C2H6OS	19	78.130000	189.000000	18.400000	1.092000	25.3	0.000000	0.000000	1.000000	Soluble	47.000000	95.000000
1,4-dioxane	C4H8O2	20	88.110000	101.100000	11.800000	1.033000	Miscible	0.000000	1.000000	0.000000	Miscible		12.000000
ethanol	C2H6O	21	46.070000	78.500000	-114.100000	0.789000	Miscible	0.000000	1.000000	0.000000	Miscible	24.600000	13.000000
ethyl acetate	C4H8O2	22	88.110000	77.000000	-83.600000	0.895000	8.7	0.000000	0.000000	1.000000	Soluble		-4.000000
ethylene glyc	C2H6O2	23	62.070000	195.000000	-13.000000	1.115000	Miscible	0.000000	1.000000	0.000000	Miscible	37.700000	111.000000
glycerin	C3H8O3	24	92.090000	290.000000	17.800000	1.261000	Miscible	0.000000	1.000000	0.000000	Miscible	42.500000	160.000000
heptane	C7H16	25	100.200000	98.000000	-90.600000	0.684000	0.01	0.000000	0.000000	1.000000	Soluble	1.920000	-4.000000
Hexamethylp	C6H18N3OP	26	179.200000	232.500000	7.200000	1.030000	Miscible	0.000000	1.000000	0.000000	Miscible	31.300000	105.000000
Hexamethylp	C6H18N3P	27	163.200000	150.000000	-44.000000	0.898000	Miscible	0.000000	1.000000	0.000000	Miscible		26.000000
hexane	C6H14	28	86.180000	69.000000	-95.000000	0.659000	0.014	0.000000	0.000000	1.000000	Soluble	1.890000	-22.000000
methanol	CH4O	29	32.040000	64.600000	-98.000000	0.791000	Miscible	0.000000	1.000000	0.000000	Miscible	32.600000	12.000000
methyl t-butyl	C5H12O	30	88.150000	55.200000	-109.000000	0.741000	5.1	0.000000	0.000000	1.000000	Soluble		-28.000000
methylene ch	CH2Cl2	31	84.930000	39.800000	-96.700000	1.326000	1.32	0.000000	0.000000	1.000000	Soluble	9.080000	1.600000
N-methyl-2-p	CH5H9NO	32	99.130000	202.000000	-24.000000	1.033000	10	0.000000	0.000000	1.000000	Soluble	32.000000	91.000000
nitromethane	CH3NO2	33	61.040000	101.200000	-29.000000	1.382000	9.5	0.000000	0.000000	1.000000	Soluble	35.900000	35.000000

Looking at the raw data

- Visualizing the raw data as scatter plots, histograms and summarized as descriptive statistics is recommended to decide if some variables need to be transformed and how to scale them before modelling
  - Histograms: Reveal the distribution of the samples for the variables
  - Descriptive statistics: Plot of the standard deviation indicated that the variables should be scaled to unit variance
  - Note that PCA in itself does not require normally distributed variables, however variables might be transformed based on underlying theory and/or background knowledge (a skewed distribution in the score plot will indicate non-linearity)

Plot of the standard deviation

- "Mice and elephants": The variables must be scaled to unit variance

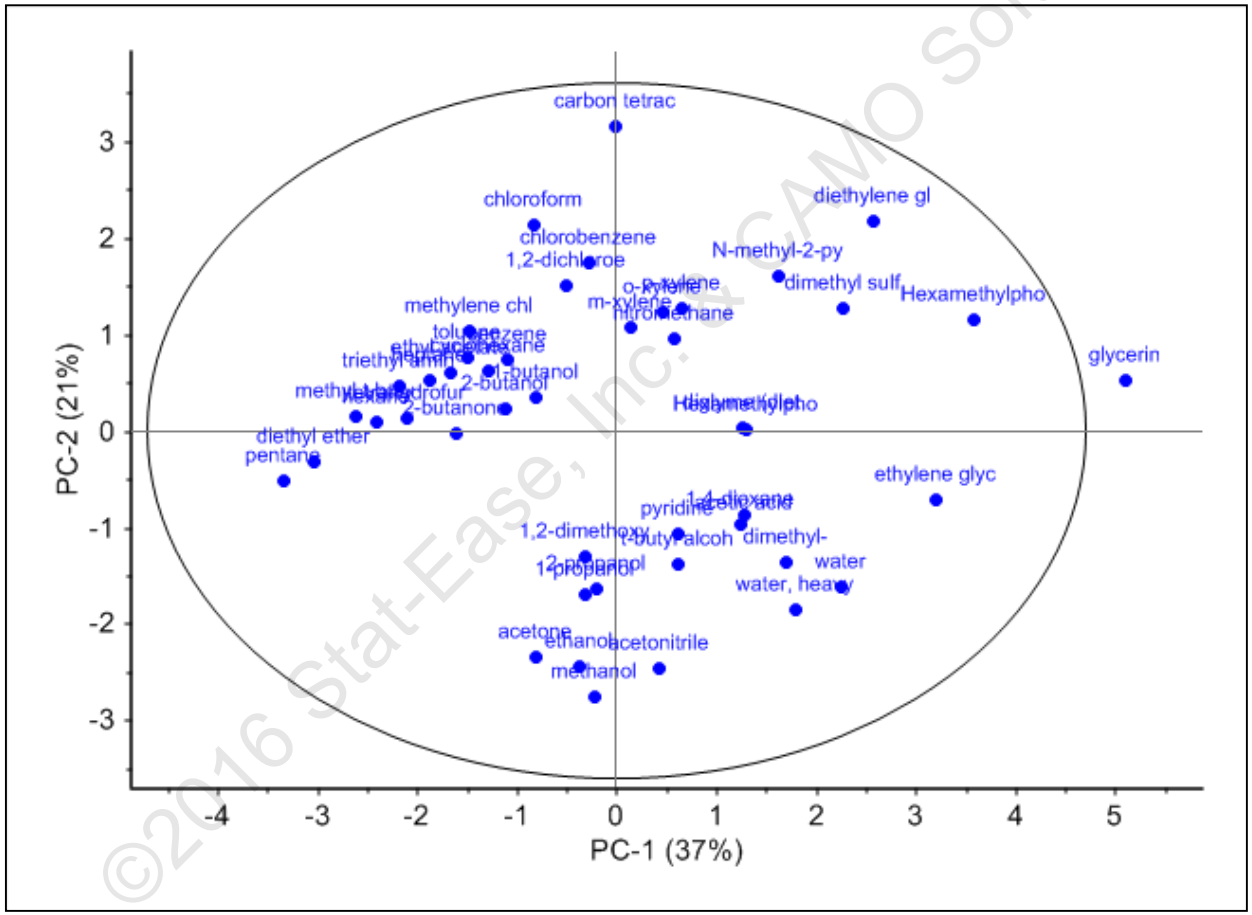


Principal Component Analysis (PCA)

- The “mother” of all multivariate methods, representing the data in terms of latent variables (principal components, PCs).
  - Objective: Find a new coordinate system that maximizes the variance in the data.
  - Score plot: Gives a map of the samples. A confidence ellipse may be added for outlier detection.
  - Loadings or Correlation Loadings: A map of the variables. Correlation loadings give a direct interpretation of the explained variance for the variables and their correlation.
  - NB!: Although this is “only mathematics”, the PCs often describe inherent underlying structures such as polarity etc.
  - Model validation is important to find the optimal model rank.

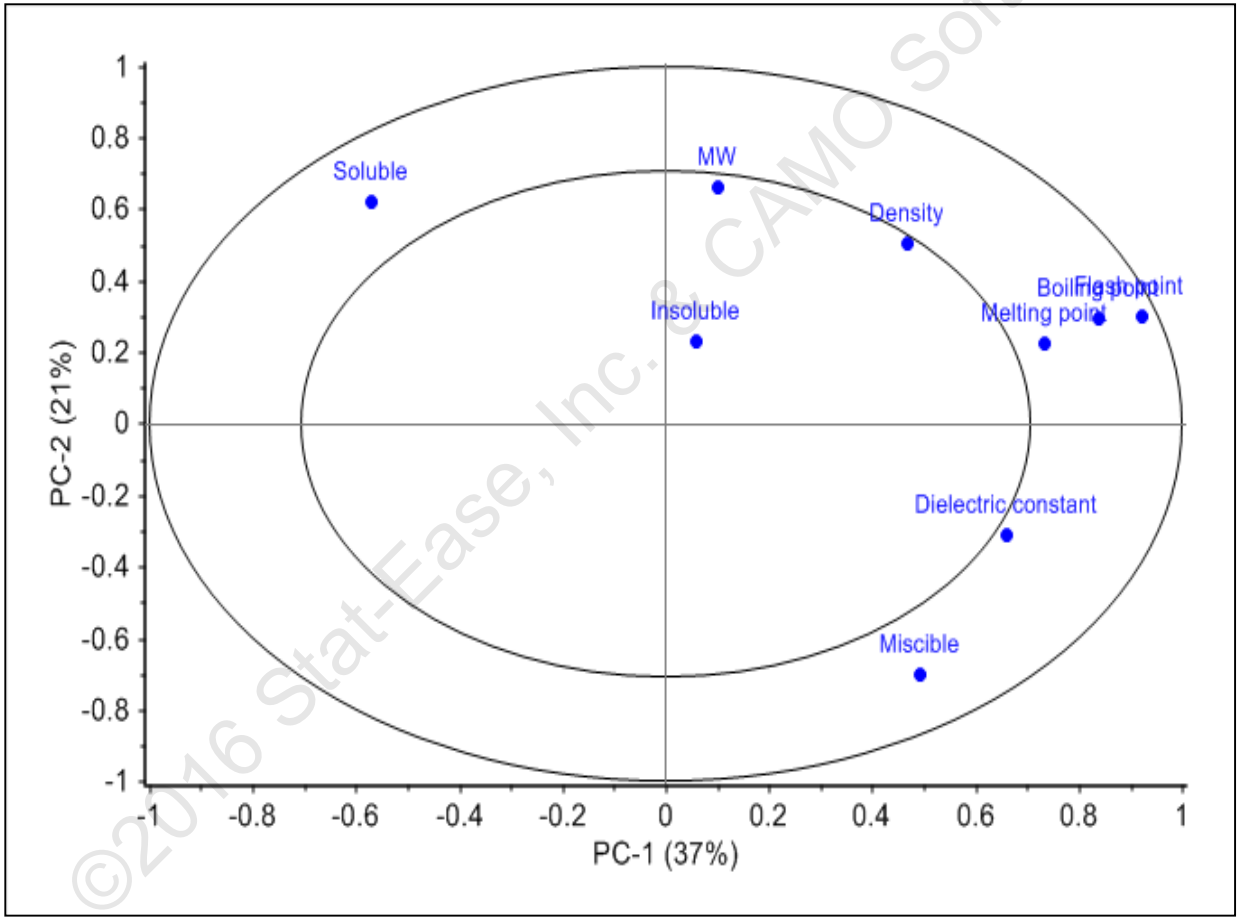
Properties for 45 Organic Solvents

- Score plot, PC2 vs. PC1



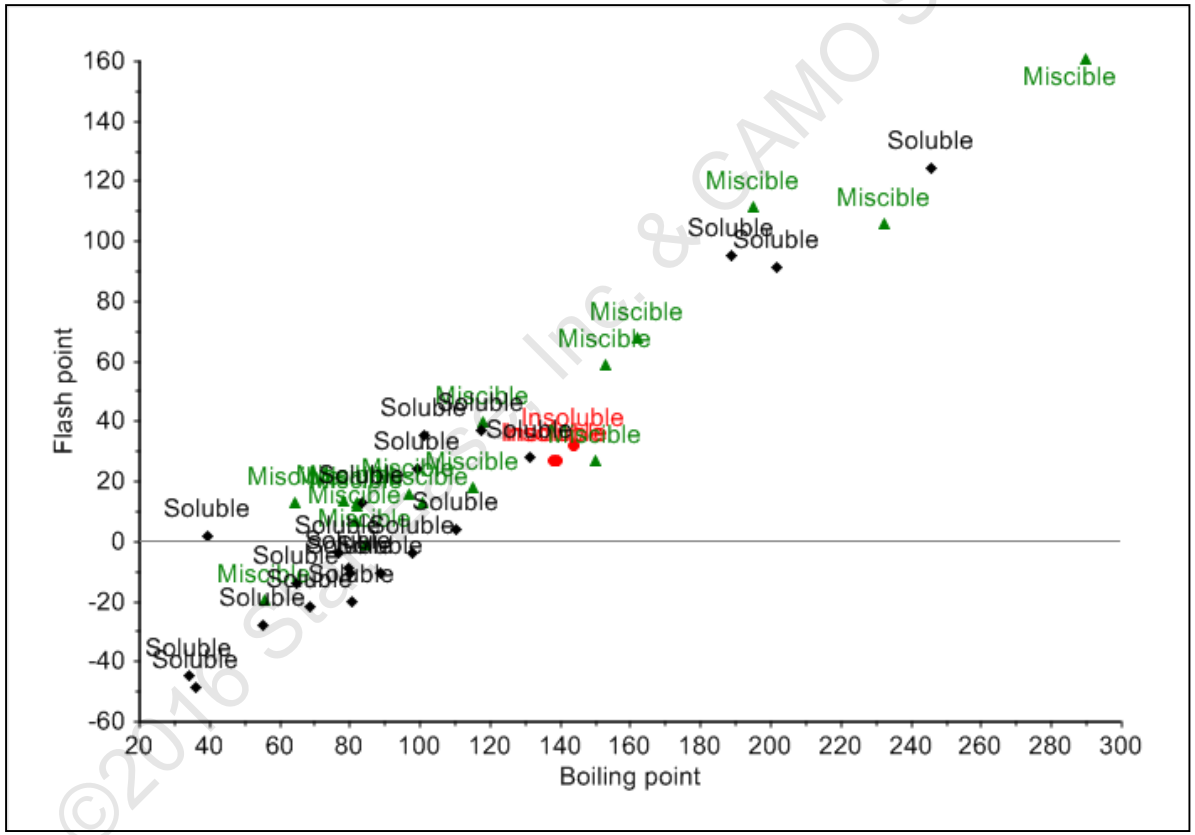
Properties for 45 Organic Solvents

- Correlation loadings plot, PC2 vs. PC3.



Properties for 45 Organic Solvents

- Scatter plot of Flash point vs. Boiling point; Correlation 0.95
  - Grouped after solubility



Part 2: Design of Experiments & Optimization

Stat-Ease
statistics made easy

CAMO
Bring data to life

How to select design points

- Given how many factors (k) you want to study and the number of coefficients (p) in the model you select, the design will be built as follows:
  - Model: p points using an optimal criteria
  - Lack-of-Fit: 5 points, based on distance – an approach that fills in the gaps *(see notes below for detail on this criteria)*
  - Replicates: 5 points, using the model optimality criteria

Principal Components

45 Organic Solvents

Scores			PC-1	PC-2	PC-3	PC-4	PC-5
			1	2	3	4	5
acetic acid	C2H4O2	1	1.2810	-0.8660	0.3134	0.1983	-1.2152
acetone	C3H6O	2	-0.8132	-2.3525	0.2582	0.2811	-0.3140
acetonitrile	C2H3N	3	0.4190	-2.4691	-0.2696	-0.5065	-0.2256
benzene	C6H6	4	-1.0892	0.7330	-0.2791	-0.4935	-0.5640
1-butanol	C4H10O	5	-0.8133	0.3407	-0.7077	-0.3397	1.1777
2-butanol	C4H10O	6	-1.1190	0.2288	-0.6907	-0.3533	0.9395
2-butanone	C4H8O	7	-1.6080	-0.0179	-0.6880	-0.4483	0.5653
t-butyl alcoho	C4H10O	8	0.6162	-1.3759	0.6793	0.1297	-0.9274
carbon tetrac	CCl4	9	-0.0022	3.1558	-0.6049	1.2624	-2.1758
chlorobenzen	C6H5Cl	10	-0.2691	1.7377	-0.3348	0.4563	-0.0464
chloroform	CHCl3	11	-0.8219	2.1263	-0.8761	0.7641	-1.8005
cyclohexane	C6H12	12	-1.2895	0.6296	-0.0458	-0.4366	-0.3102
1,2-dichloroe	C2H4Cl2	13	-0.5049	1.4971	-0.8180	0.1012	-1.0365
diethylene gl	C4H10O3	14	2.5785	2.1794	-1.0339	-0.1434	1.4696
diethyl ether	C4H10O	15	-3.0278	-0.3221	-0.2313	-0.1082	0.3127
diglyme (diet	C6H14O3	16	1.2565	0.0315	1.1781	1.8833	0.4814
1,2-dimethox	C4H10O2	17	-0.3100	-1.2944	0.8380	1.0066	-0.4672
dimethyl-	C3H7NO	18	1.7073	-1.3572	-0.1881	0.2522	0.4604
dimethyl sulf	C2H6OS	19	2.2790	1.2779	-1.6568	-1.1060	0.7324
1,4-dioxane	C4H8O2	20	1.2434	-0.9737	0.1671	0.3353	-1.2609
ethanol	C2H6O	21	-0.3732	-2.4359	-0.0156	0.1406	0.2514
ethyl acetate	C4H8O2	22	-1.6552	0.5945	-0.3986	0.0491	0.1910
ethylene glyc	C2H6O2	23	3.2048	-0.7243	-0.5465	-0.0399	0.2458
glycerin	C3H8O3	24	5.1085	0.5255	-0.4613	0.4094	0.7950
heptane	C7H16	25	-1.8811	0.5191	0.1730	0.3105	1.0316
Hexamethylp	C6H18N3OP	26	3.5834	1.1446	1.0794	1.9795	0.7694
Hexamethylp	C6H18N3P	27	1.2995	0.0125	1.4651	2.3201	0.0652
hexane	C6H14	28	-2.4073	0.0831	0.0458	0.0485	0.7495
methanol	CH4O	29	-0.2125	-2.7572	-0.3642	-0.3336	0.0375
methyl t-butyl	C5H12O	30	-2.6218	0.1485	-0.0357	0.2153	0.4525

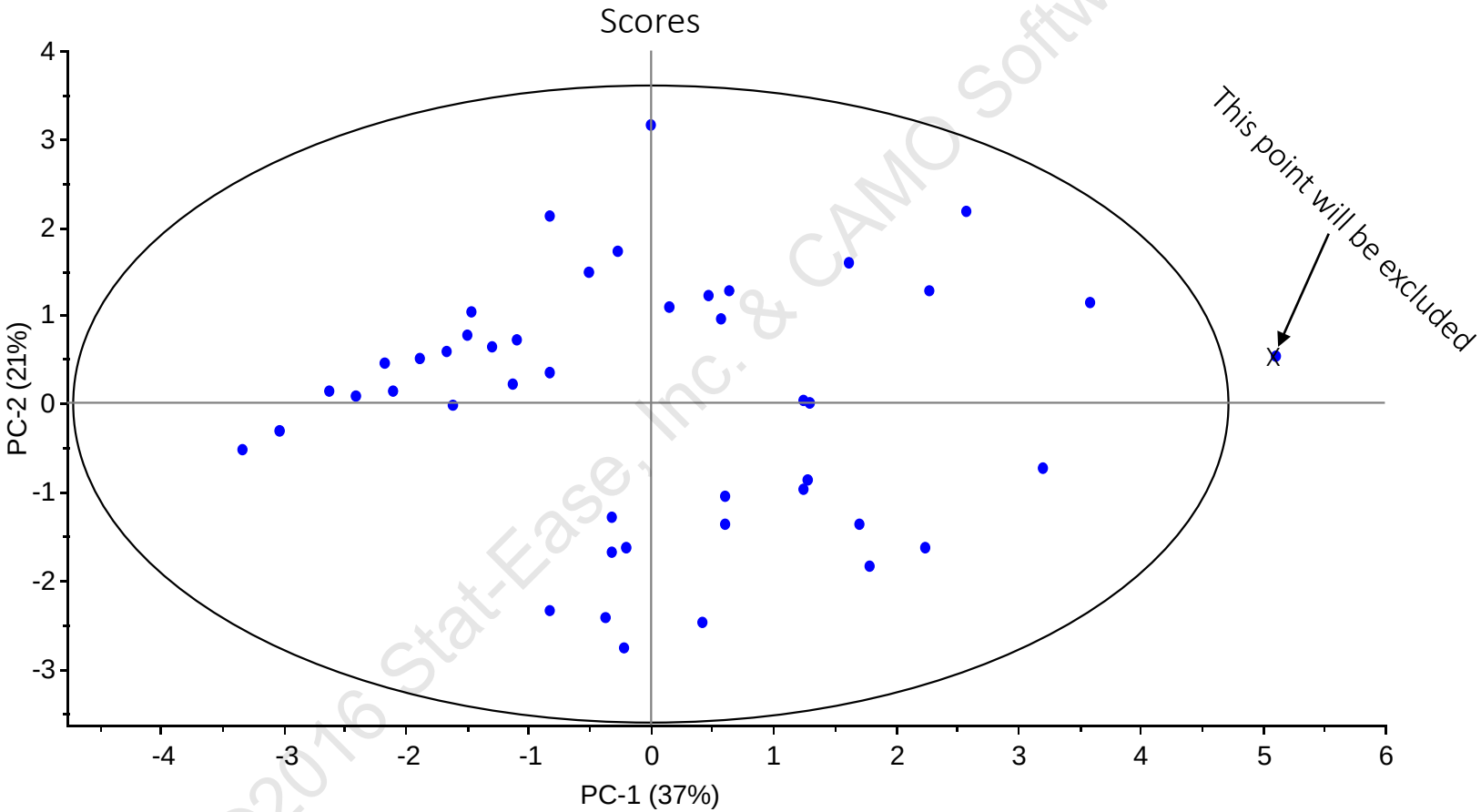
First Two Principal Components

45 Organic Solvents

			PC-1	PC-2
			1	2
acetic acid	C2H4O2	1	1.2810	-0.8660
acetone	C3H6O	2	-0.8132	-2.3525
acetonitrile	C2H3N	3	0.4190	-2.4691
benzene	C6H6	4	-1.0892	0.7330
1-butanol	C4H10O	5	-0.8133	0.3407
2-butanol	C4H10O	6	-1.1190	0.2288
2-butanone	C4H8O	7	-1.6080	-0.0179
t-butyl alcoho	C4H10O	8	0.6162	-1.3759
carbon tetrac	CCl4	9	-0.0022	3.1558
chlorobenzen	C6H5Cl	10	-0.2691	1.7377
chloroform	CHCl3	11	-0.8219	2.1263
cyclohexane	C6H12	12	-1.2895	0.6296
1,2-dichloroe	C2H4Cl2	13	-0.5049	1.4971
diethylene gl	C4H10O3	14	2.5785	2.1794
diethyl ether	C4H10O	15	-3.0278	-0.3221
diglyme (diet	C6H14O3	16	1.2565	0.0315
1,2-dimethox	C4H10O2	17	-0.3100	-1.2944
dimethyl-	C3H7NO	18	1.7073	-1.3572
dimethyl sulf	C2H6OS	19	2.2790	1.2779
1,4-dioxane	C4H8O2	20	1.2434	-0.9737
ethanol	C2H6O	21	-0.3732	-2.4359
ethyl acetate	C4H8O2	22	-1.6552	0.5945
ethylene glyc	C2H6O2	23	3.2048	-0.7243
glycerin	C3H8O3	24	5.1085	0.5255
heptane	C7H16	25	-1.8811	0.5191
Hexamethylp	C6H18N3OP	26	3.5834	1.1446
Hexamethylp	C6H18N3P	27	1.2995	0.0125
hexane	C6H14	28	-2.4073	0.0831
methanol	CH4O	29	-0.2125	-2.7572
methyl t-butyl	C5H12O	30	-2.6218	0.1485

Candidate points for DOE

44 Organic Solvents



Stat-Ease statistics made easy

CAMO Bring data to life

“Good” Design Properties

An Experimenter’s Wish List

- ☑ Allow the chosen polynomial to be estimated well.
- ☑ Give sufficient information to allow a test for lack of fit.
  - ☑ Have more unique design points than model coefficients.
  - ☑ Provide an estimate of “pure” error.
- ☑ Remain insensitive to outliers, influential values and bias from model misspecification.
- ☑ Provide a check on variance assumptions, e.g., studentized residuals are N(0, σ^2); normal, mean zero, constant variance.
- ☑ Generate useful information throughout the region of interest, i.e., provide a good distribution of standard error of prediction.
- ☑ Do not contain an excessively large number of trials.

Optimal Design

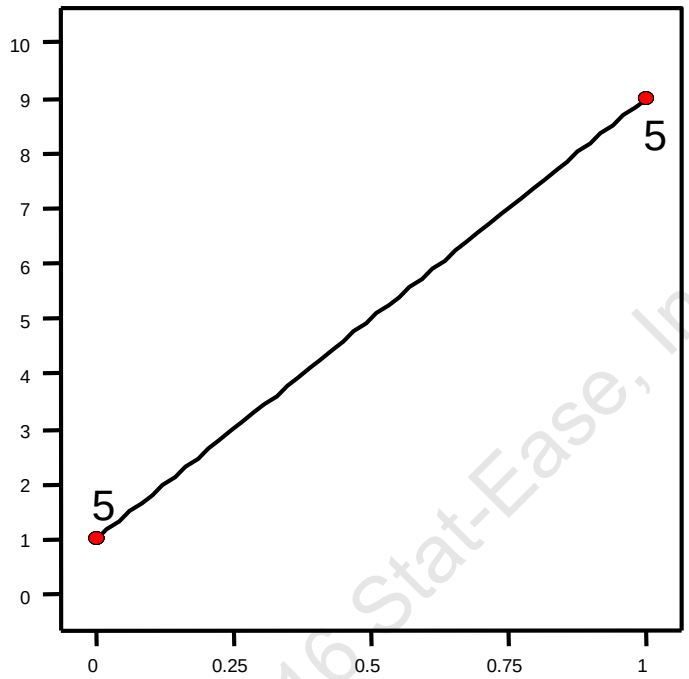
Design-Expert's modified algorithm

1. Select a polynomial that you think is needed to get a decent approximation of the actual response surface.
   - *Usually a quadratic.*
2. Select good points to estimate your model.
   (There are two basic algorithms: point exchange and coordinate exchange.)
3. Select design points for:
   - Model: *To allow estimation of all coefficients.*
   - Lack-of-fit: *To test how well the model represents actual behavior in our region of interest.*
   - Replicates: *To estimate pure error.*

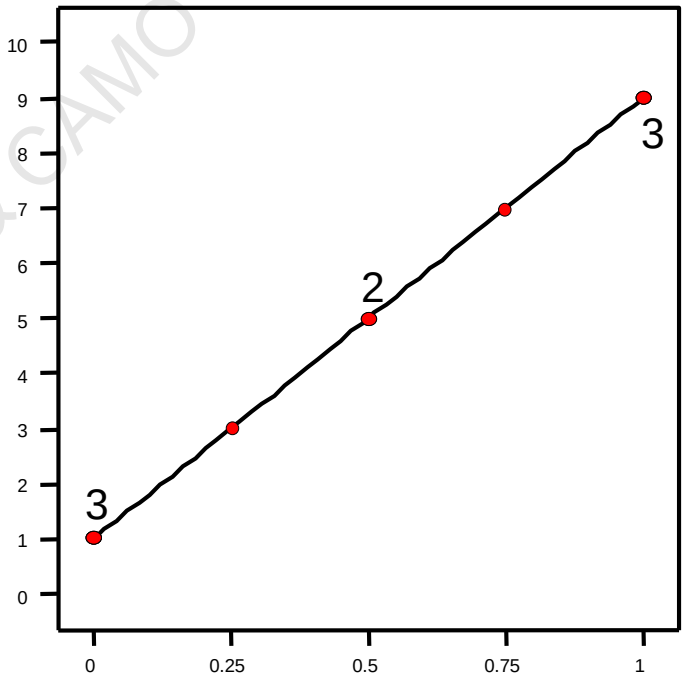
Point Selection by Computer vs Expert

Linear model with 10 runs allowed

- Optimal Selection



Good Selection



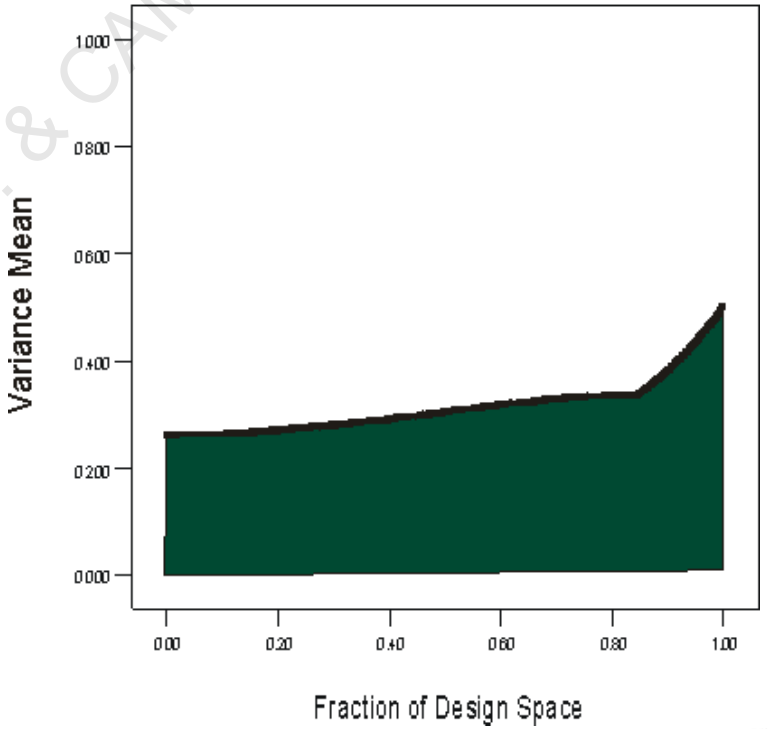
Optimal Point Exchange

Selecting Points from Candidate Set

- Given how many factors (k) you want to study and the number of coefficients (p) in the model you select, the design will be built as follows:
  - Model: p points using an optimal criteria
  - Lack-of-Fit: 5 points, based on distance – an approach that fills in the gaps *(see notes below for detail on this criteria)*
  - Replicates: 5 points, using the model optimality criteria

I-optimal Point Selection

- An I-optimal design seeks to minimize the integral of the prediction variance across the design space. These designs are built algorithmically to provide lower integrated prediction variance across the design space. This equates to minimizing the area under the FDS curve.



Design Point Selection

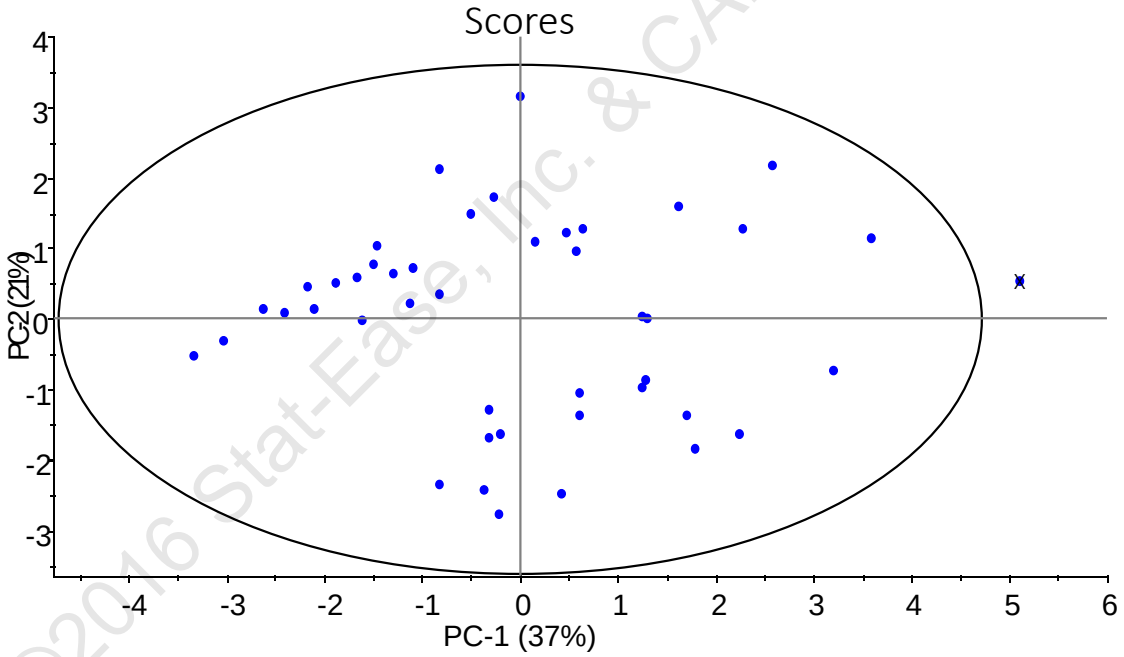
Point Exchange Algorithm

1. Start with a candidate list of points.
2. Randomly pick a nonsingular set of model points.
3. Perform 1-point exchange steps until there is no improvement in the design. Then perform 2-point exchange steps, and so on through a 5-point exchange. If at any time, there is improvement, start over with 1-point exchanges.
4. The exchanges continue until there is no further improvement in the optimality criterion all the way through the 5-point exchange.

Candidate set for DOE

44 Organic Solvents

- In our example point exchange is used to chose a design from the 44 organic solvents (represented by their principle components).

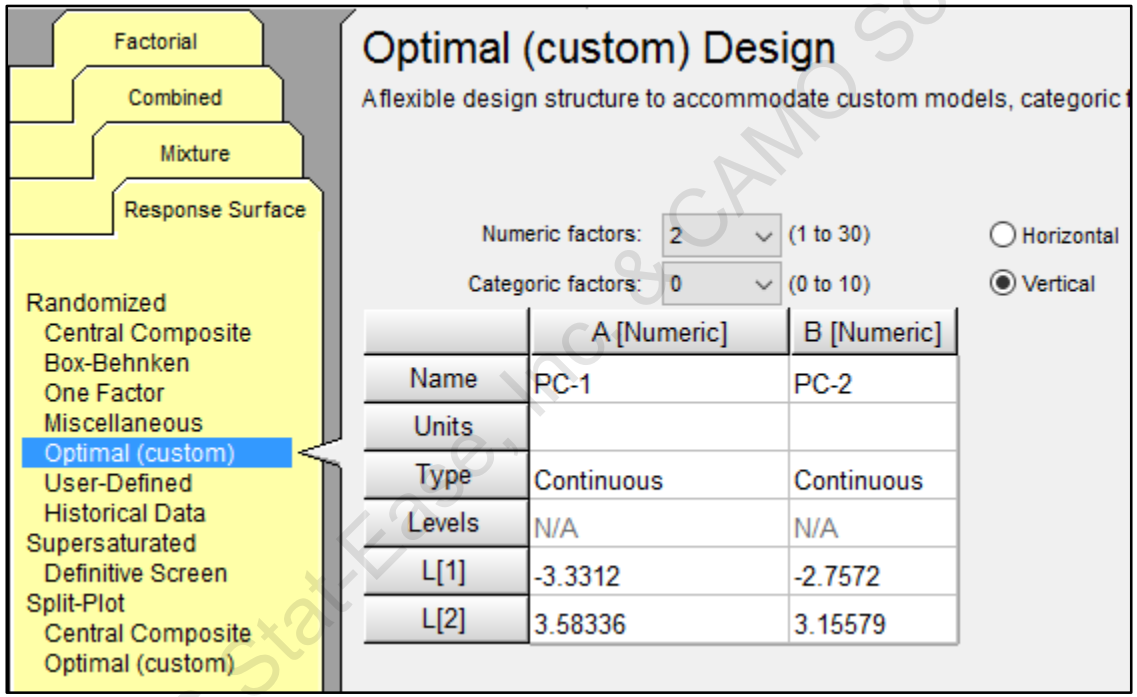


Stat-Ease statistics made easy

CAMO Bring data to life

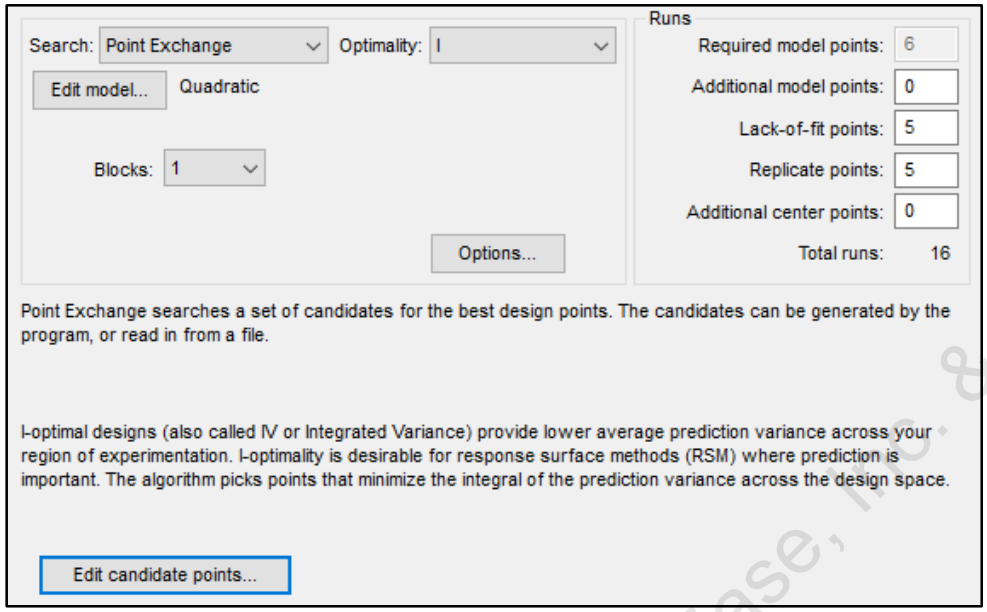
First Two Principal Components

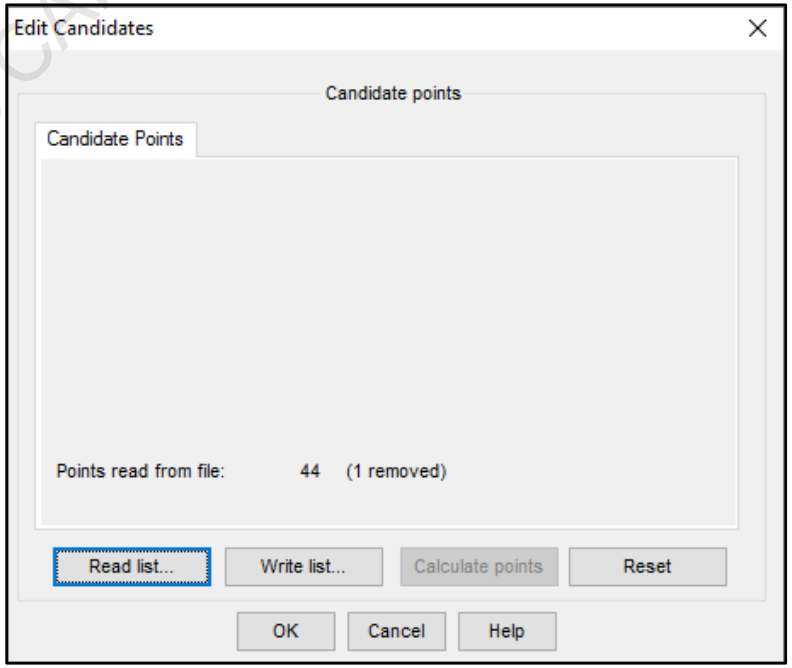
Build an Optimal Design *(page 1 of 3)*



Choose Design Points *(from 44 solvents)*

Build an Optimal Design *(page 2 of 3)*



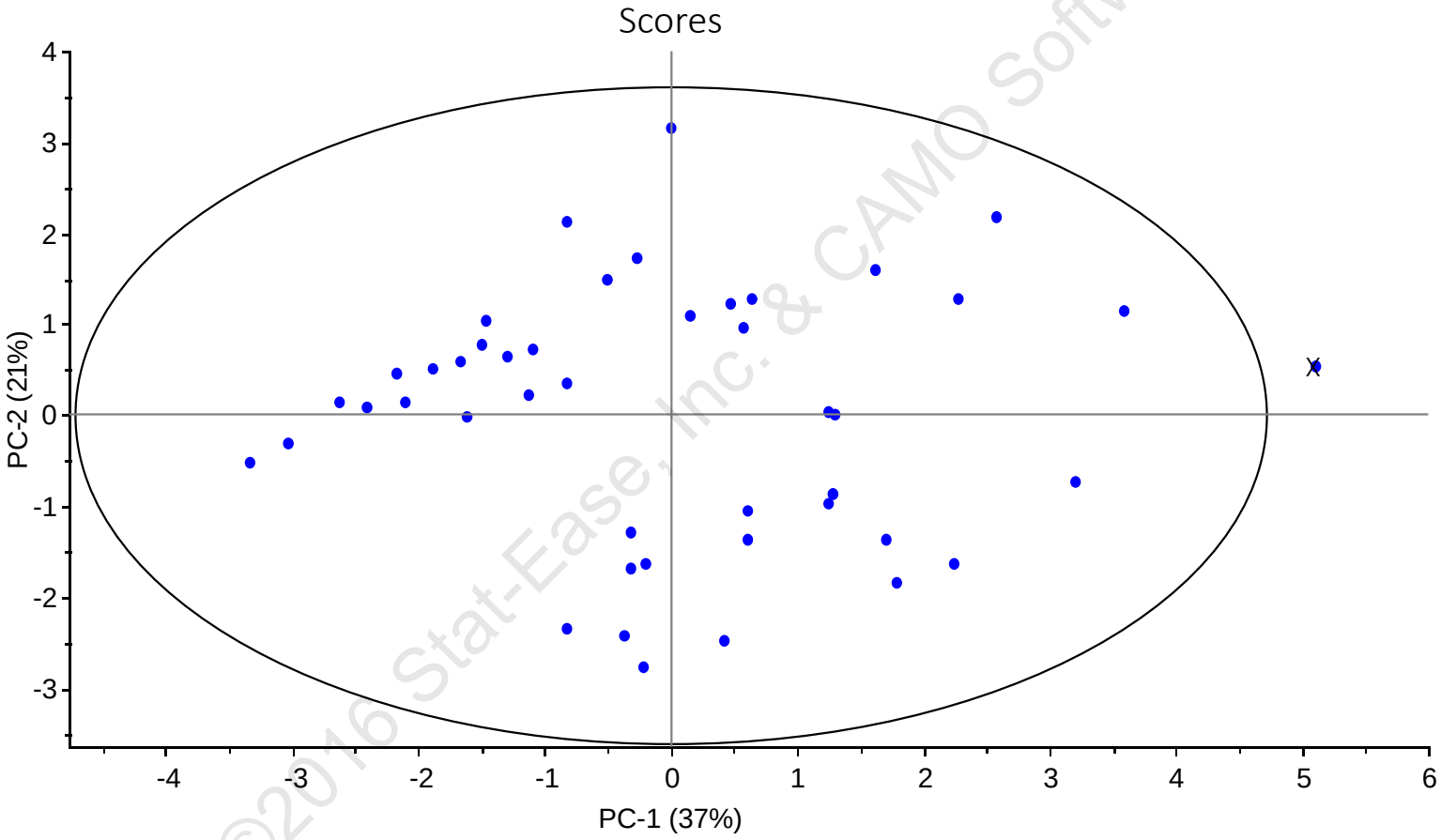


Stat-Ease statistics made easy®

CAMO Bring data to life

Candidate points for DOE

44 Organic Solvents

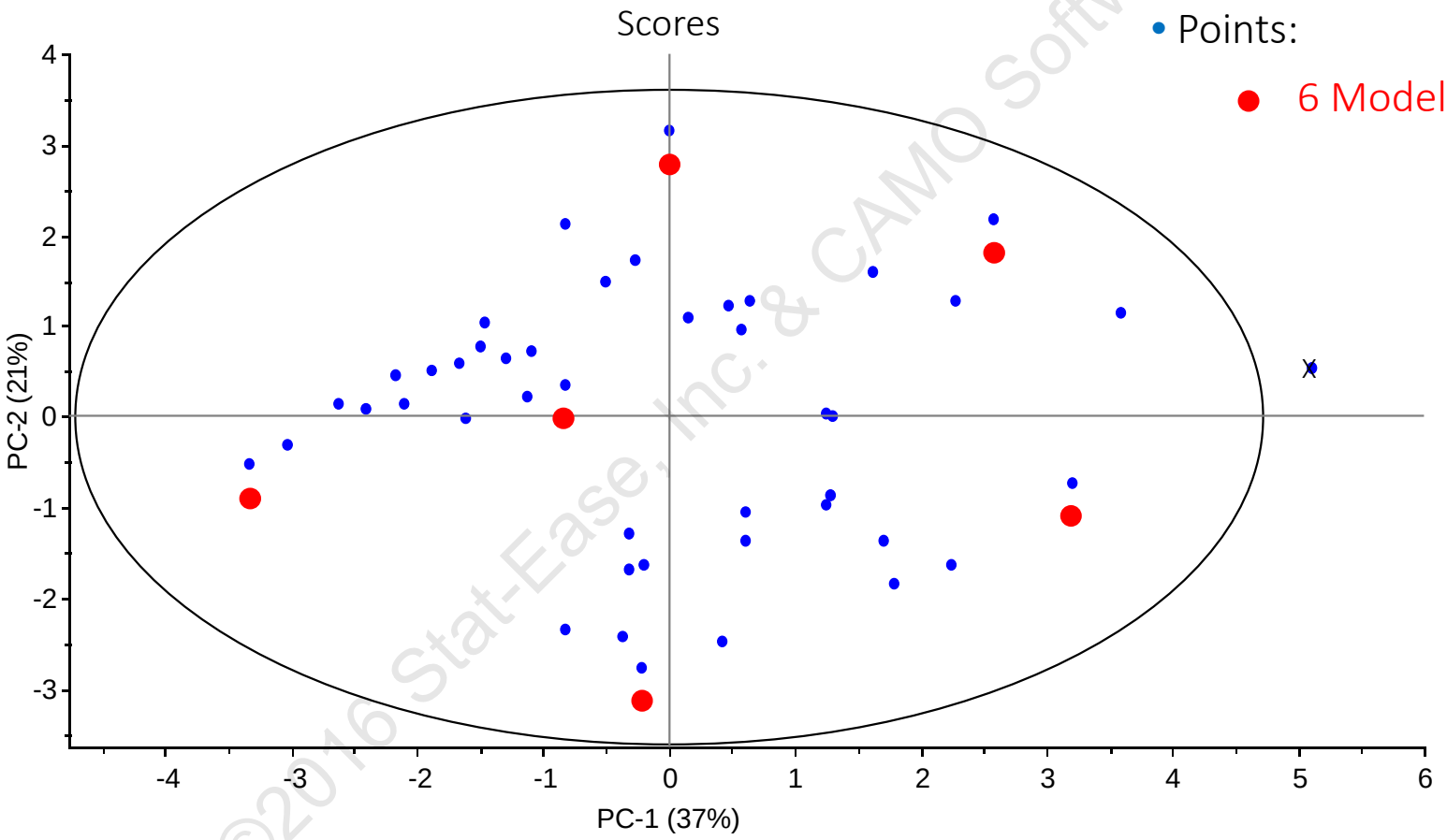


Stat-Ease statistics made easy

CAMO Bring data to life

Candidate points for DOE

6 Model points selected using I-optimality

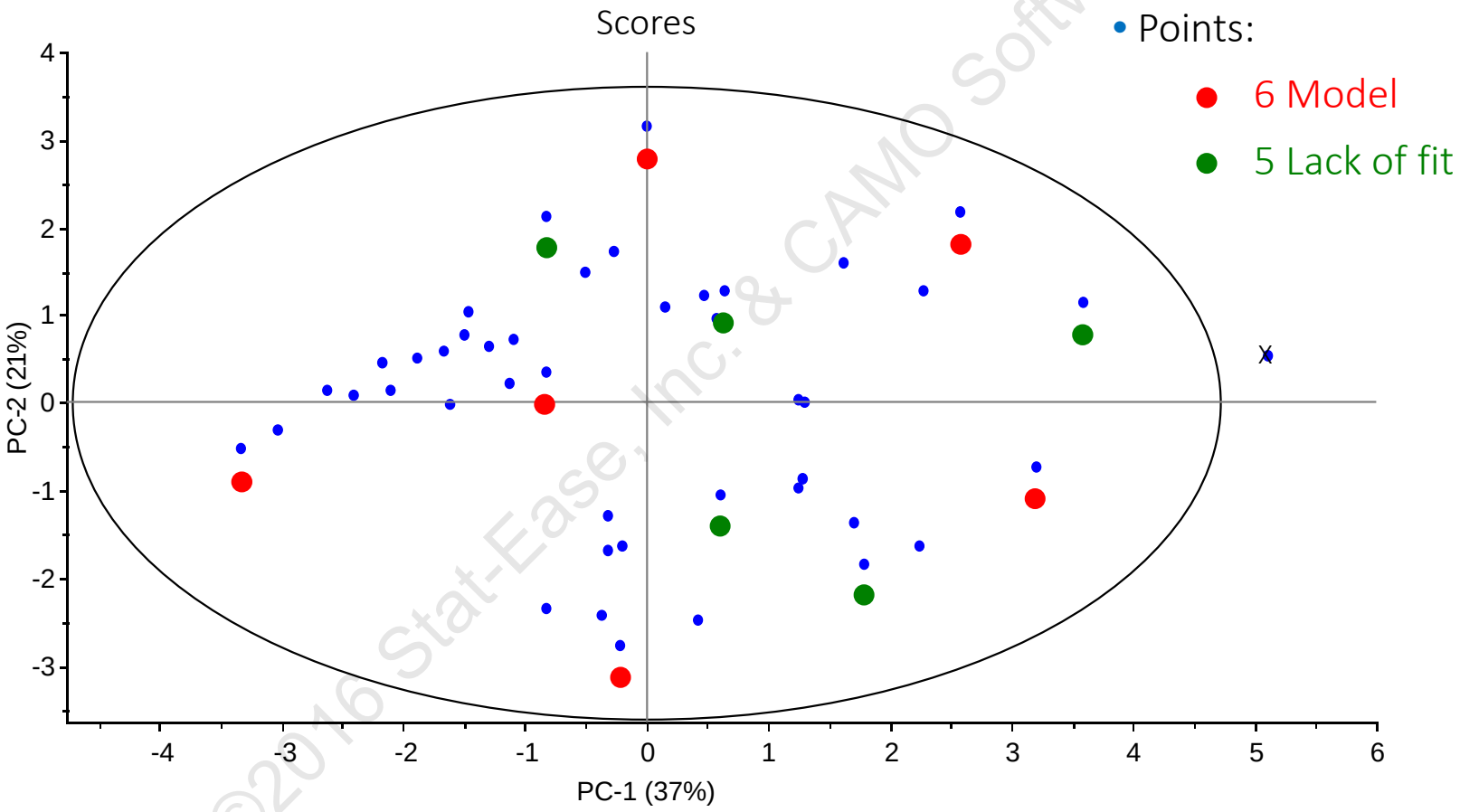


Stat-Ease statistics made easy

CAMO Bring data to life

Candidate points for DOE

5 Lack of Fit points selected using Distance

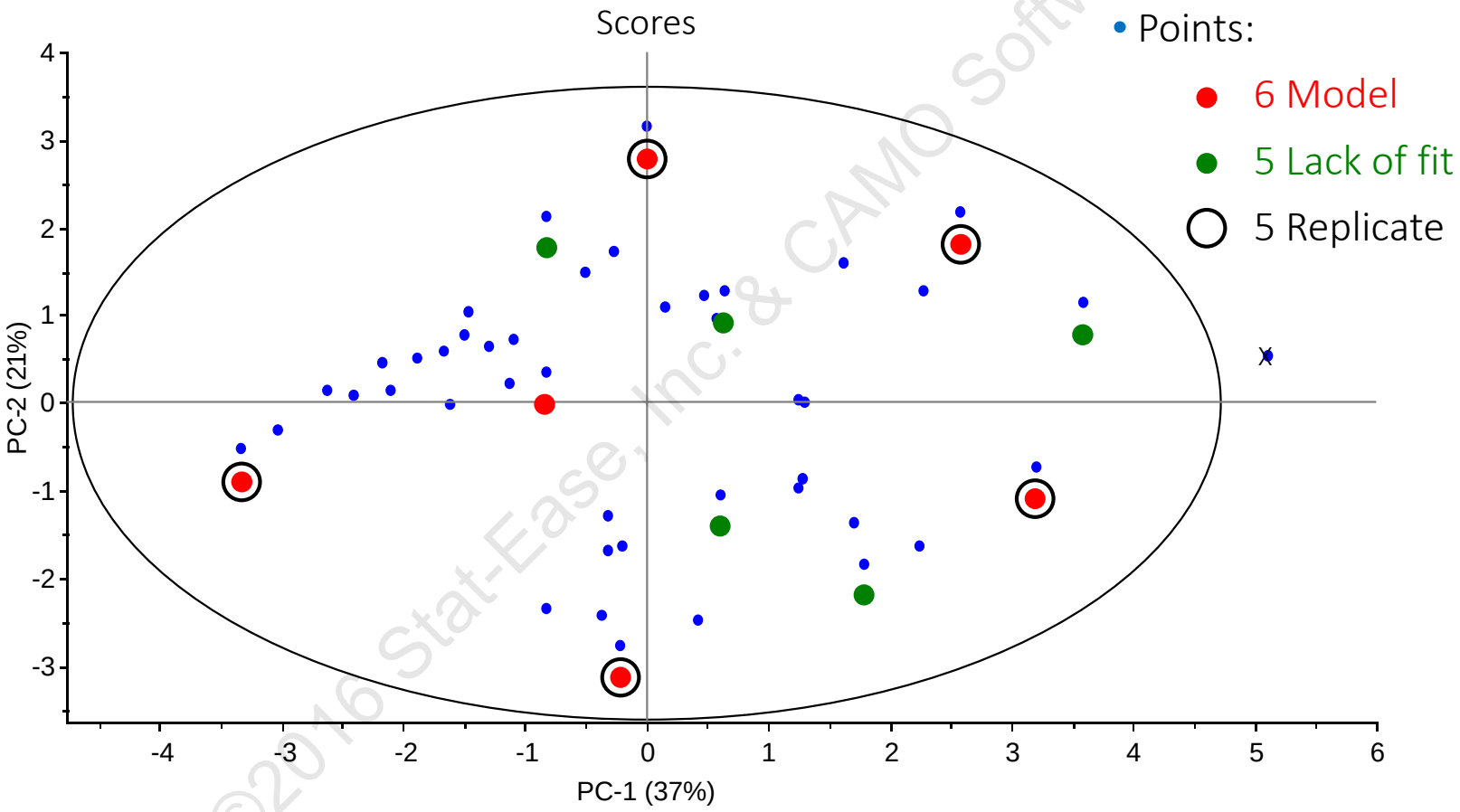


Stat-Ease statistics made easy

CAMO Bring data to life

Candidate points for DOE

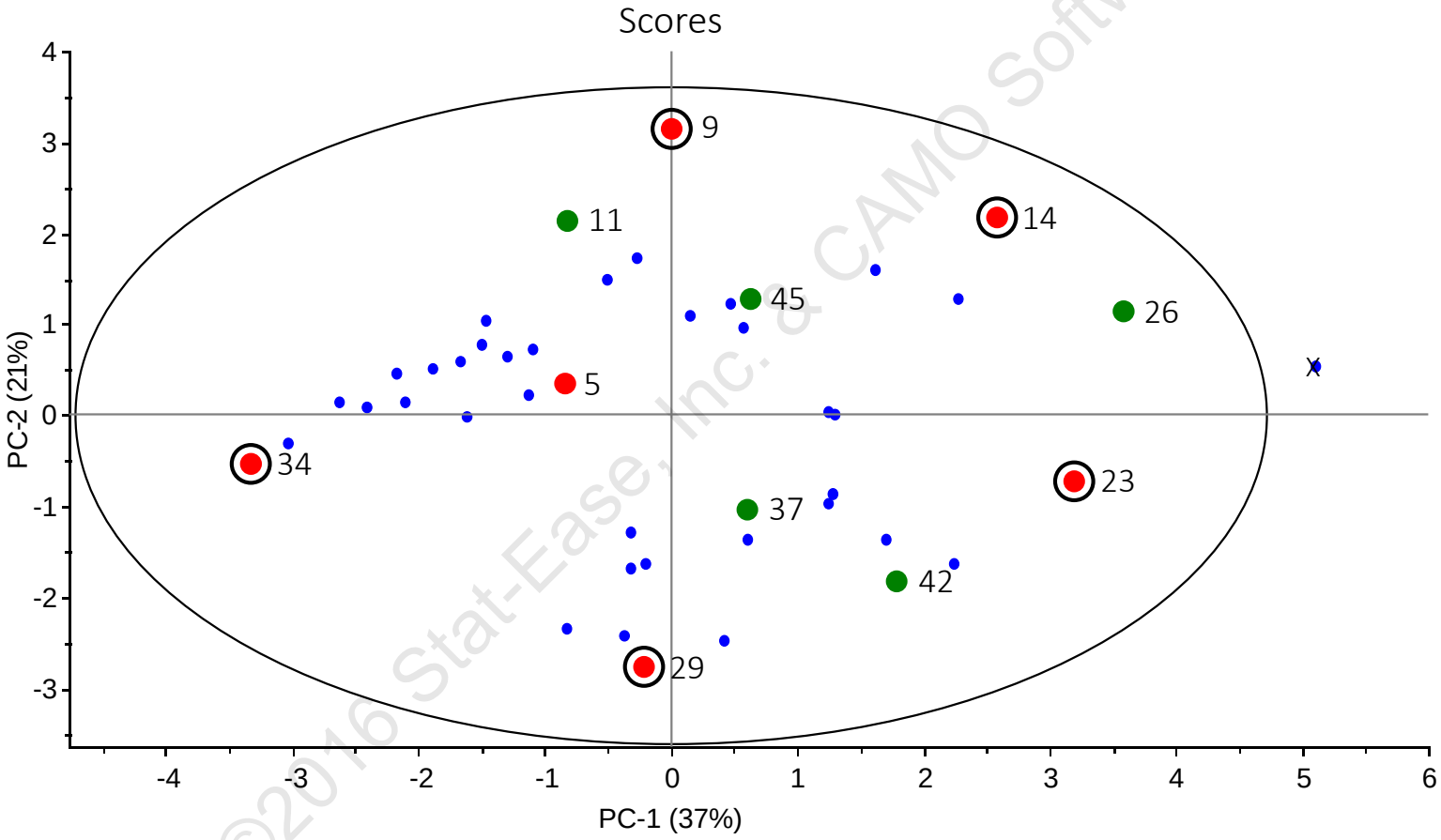
5 Replicates selected using I-optimality



Stat-Ease statistics made easy

CAMO Bring data to life

Solvents Used in DOE
Uses 11 solvents with 5 replicates

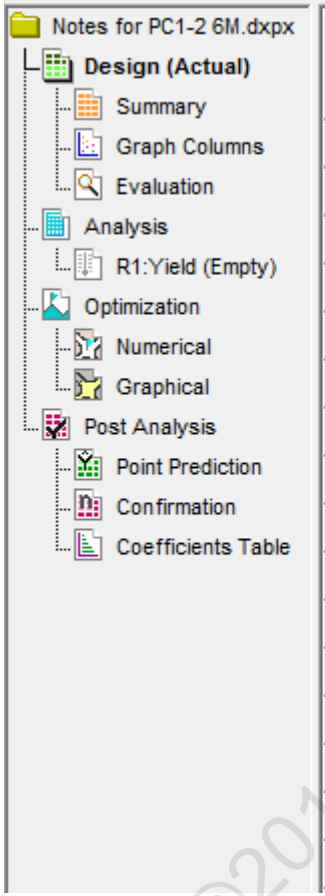


Stat-Ease statistics made easy

CAMO Bring data to life

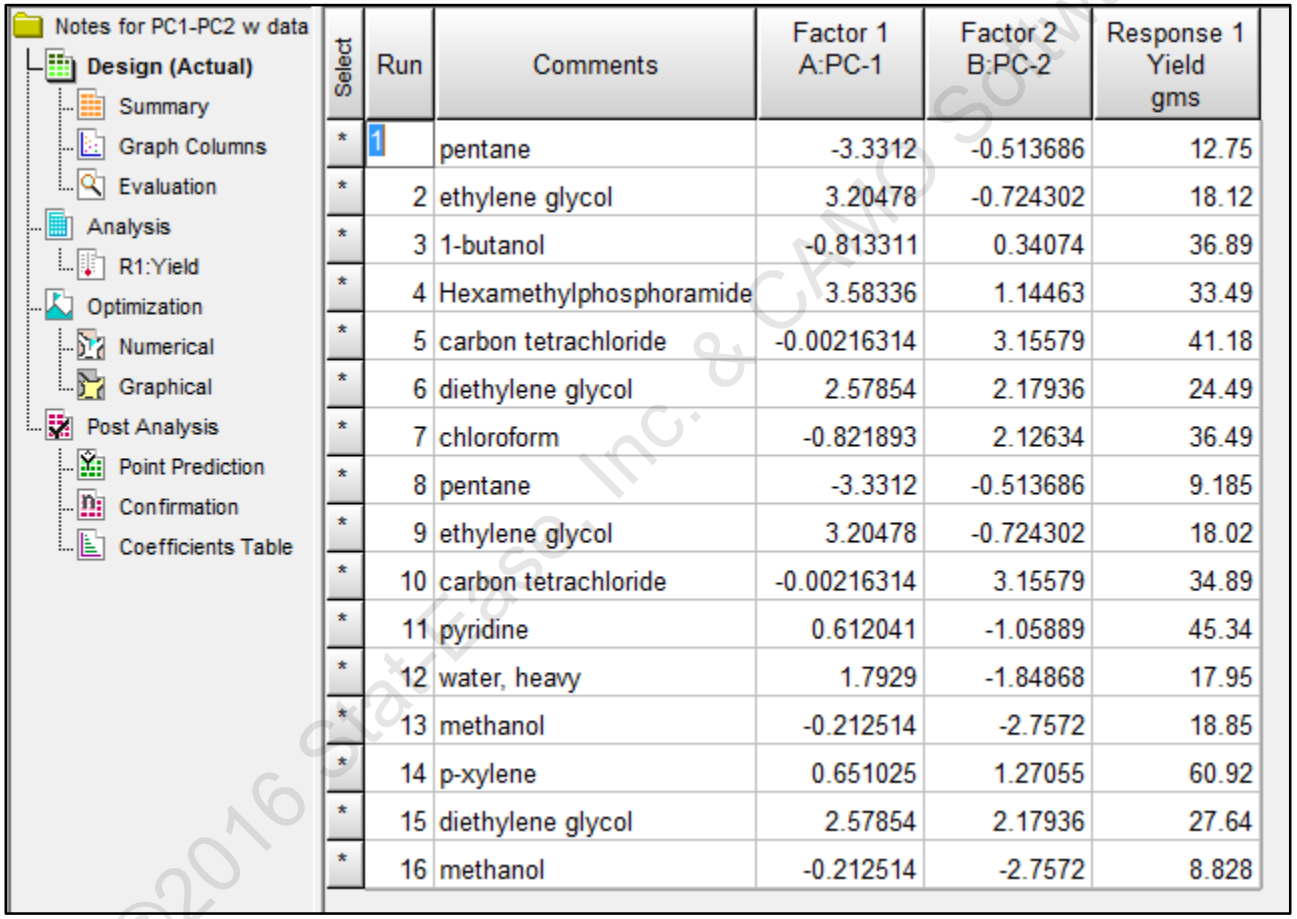
First Two Principal Components

Build an Optimal Design *(page 3 of 3)*



Select	Run	Build Type	Comments	Factor 1 A:PC-1	Factor 2 B:PC-2	Response 1 Yield gms
*	1	Replicate	pentane	-3.3312	-0.513686	
*	2	Replicate	ethylene glycol	3.20478	-0.724302	
*	3	Model	1-butanol	-0.813311	0.34074	
*	4	Lack of Fit	Hexamethylphosphoramide	3.58336	1.14463	
*	5	Model	carbon tetrachloride	-0.00216314	3.15579	
*	6	Replicate	diethylene glycol	2.57854	2.17936	
*	7	Lack of Fit	chloroform	-0.821893	2.12634	
*	8	Model	pentane	-3.3312	-0.513686	
*	9	Model	ethylene glycol	3.20478	-0.724302	
*	10	Replicate	carbon tetrachloride	-0.00216314	3.15579	
*	11	Lack of Fit	pyridine	0.612041	-1.05889	
*	12	Lack of Fit	water, heavy	1.7929	-1.84868	
*	13	Model	methanol	-0.212514	-2.7572	
*	14	Lack of Fit	p-xylene	0.651025	1.27055	
*	15	Model	diethylene glycol	2.57854	2.17936	
*	16	Replicate	methanol	-0.212514	-2.7572	

Simulate Response, Analyze and Maximize



Select	Run	Comments	Factor 1 A:PC-1	Factor 2 B:PC-2	Response 1 Yield gms
*	1	pentane	-3.3312	-0.513686	12.75
*	2	ethylene glycol	3.20478	-0.724302	18.12
*	3	1-butanol	-0.813311	0.34074	36.89
*	4	Hexamethylphosphoramide	3.58336	1.14463	33.49
*	5	carbon tetrachloride	-0.00216314	3.15579	41.18
*	6	diethylene glycol	2.57854	2.17936	24.49
*	7	chloroform	-0.821893	2.12634	36.49
*	8	pentane	-3.3312	-0.513686	9.185
*	9	ethylene glycol	3.20478	-0.724302	18.02
*	10	carbon tetrachloride	-0.00216314	3.15579	34.89
*	11	pyridine	0.612041	-1.05889	45.34
*	12	water, heavy	1.7929	-1.84868	17.95
*	13	methanol	-0.212514	-2.7572	18.85
*	14	p-xylene	0.651025	1.27055	60.92
*	15	diethylene glycol	2.57854	2.17936	27.64
*	16	methanol	-0.212514	-2.7572	8.828

- PC1-PC2 w data.dxpx

Stat-Ease statistics made easy

CAMO Bring data to life

Simulate Response, Analyze and Maximize

Transform | Fit Summary | Model | ANOVA

Response 1 Yield

ANOVA for Response Surface Reduced Quadratic model

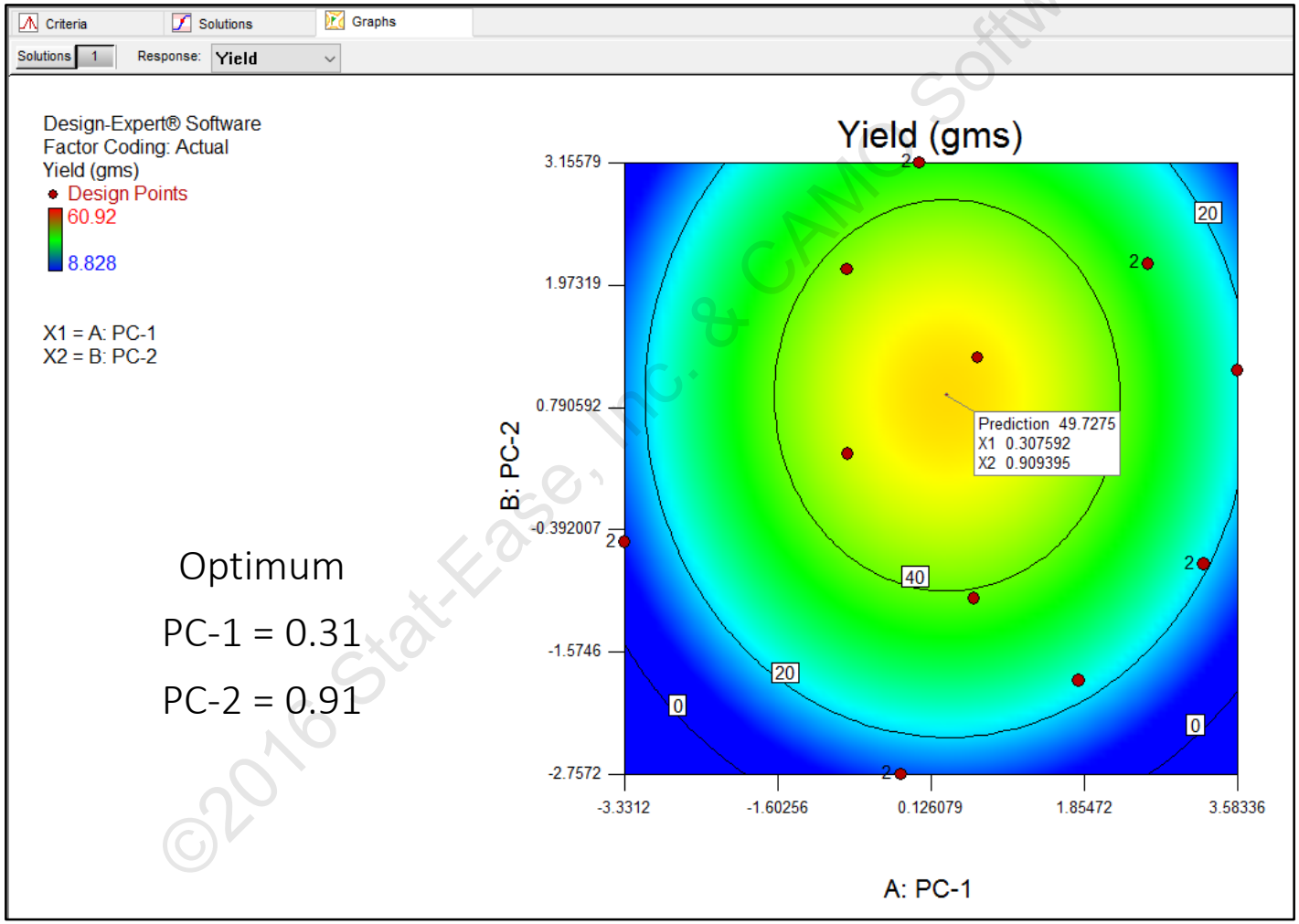
Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	2495.70	4	623.93	10.74	0.0009
A-PC-1	*57.40*	*1*	*57.40*	*0.99*	*0.3416*
B-PC-2	*828.47*	*1*	*828.47*	*14.26*	*0.0031*
A²	*1471.52*	*1*	*1471.52*	*25.33*	*0.0004*
B²	*784.03*	*1*	*784.03*	*13.49*	*0.0037*
Residual	639.11	11	58.10		
Lack of Fit	*557.79*	*6*	*92.96*	*5.72*	*0.0376*
Pure Error	*81.32*	*5*	*16.26*		
Cor Total	3134.81	15			

Std. Dev.	7.62	R-Squared	0.7961
Mean	27.81	Adj R-Squarec	0.7220
C.V. %	27.40	Pred R-Square	0.5727
PRESS	1339.55	Adeq Precisio	9.089

- Significant Lack of Fit
  - Perhaps two principal components are not enough to describe the response!
  - More on this later.

Simulate Response, Analyze and Maximize



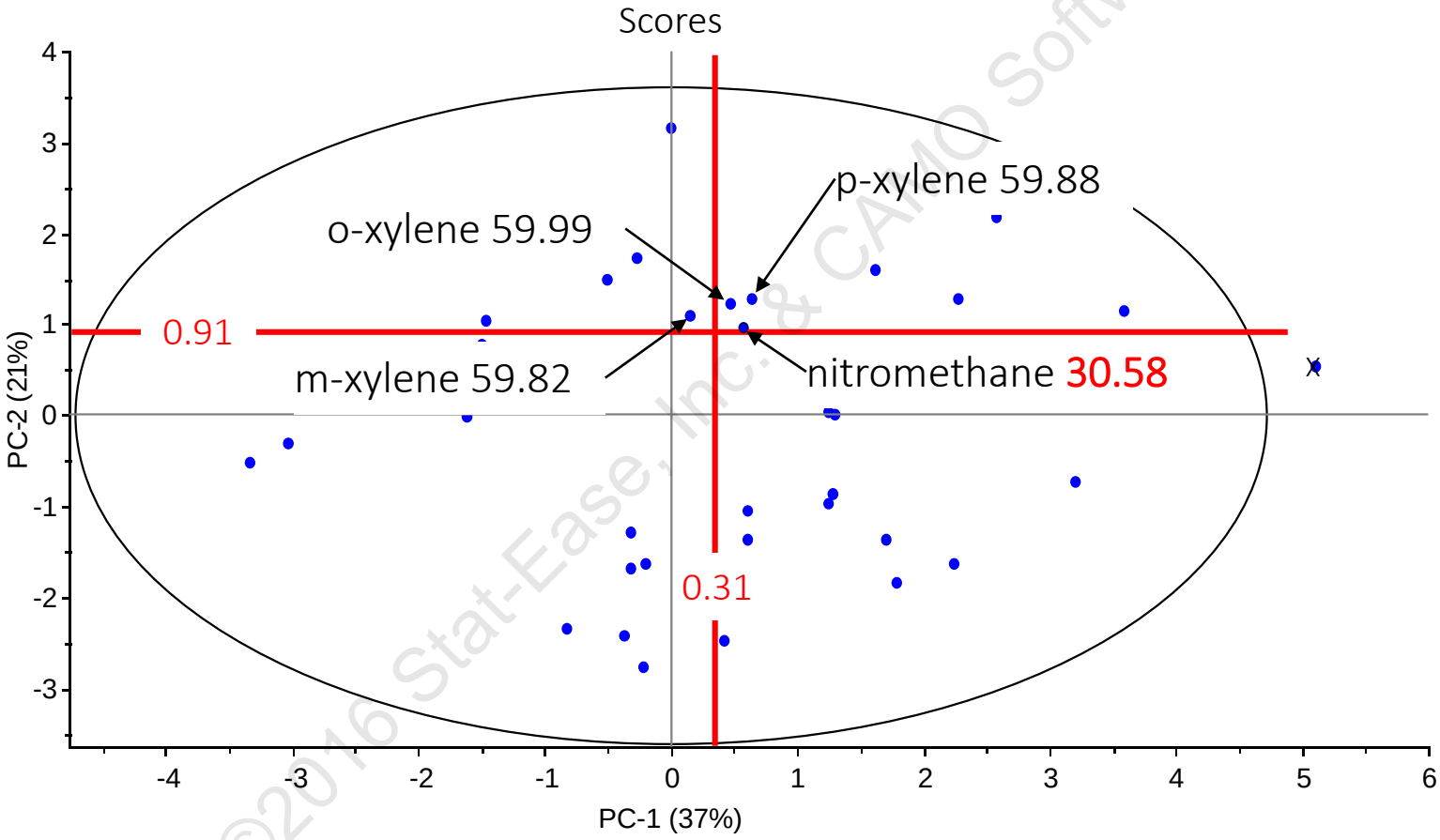
Optimum

PC-1 = 0.31

PC-2 = 0.91

Points near Optimum

PC-1 = 0.31 and PC-2 = 0.91



Stat-Ease statistics made easy

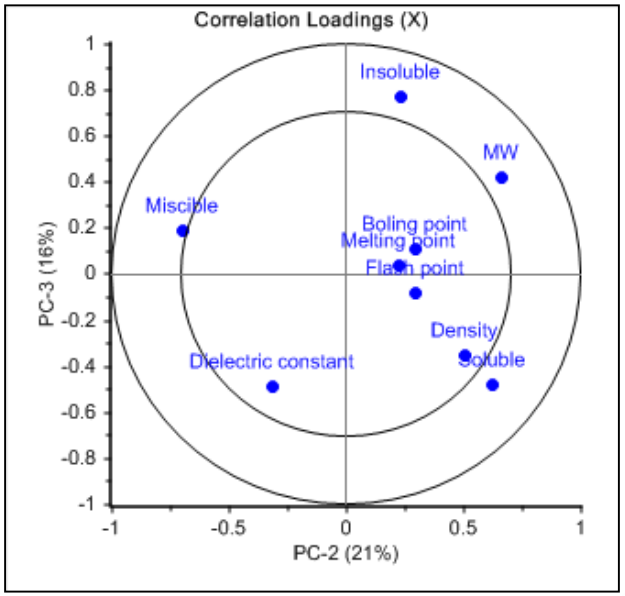
CAMO Bring data to life

First Three Principal Components

45 Organic Solvents

Scores			PC-1	PC-2	PC-3
			1	2	3
acetic acid	C2H4O2	1	1.2810	-0.8660	0.3134
acetone	C3H6O	2	-0.8132	-2.3525	0.2582
acetonitrile	C2H3N	3	0.4190	-2.4691	-0.2696
benzene	C6H6	4	-1.0892	0.7330	-0.2791
1-butanol	C4H10O	5	-0.8133	0.3407	-0.7077
2-butanol	C4H10O	6	-1.1190	0.2288	-0.6907
2-butanone	C4H8O	7	-1.6080	-0.0179	-0.6880
t-butyl alcoho	C4H10O	8	0.6162	-1.3759	0.6793
carbon tetrac	CCl4	9	-0.0022	3.1558	-0.6049
chlorobenzen	C6H5Cl	10	-0.2691	1.7377	-0.3348
chloroform	CHCl3	11	-0.8219	2.1263	-0.8761
cyclohexane	C6H12	12	-1.2895	0.6296	-0.0458
1,2-dichloroe	C2H4Cl2	13	-0.5049	1.4971	-0.8180
diethylene gl	C4H10O3	14	2.5785	2.1794	-1.0339
diethyl ether	C4H10O	15	-3.0278	-0.3221	-0.2313
diglyme (diet	C6H14O3	16	1.2565	0.0315	1.1781
1,2-dimethox	C4H10O2	17	-0.3100	-1.2944	0.8380
dimethyl-	C3H7NO	18	1.7073	-1.3572	-0.1881
dimethyl sulf	C2H6OS	19	2.2790	1.2779	-1.6568
1,4-dioxane	C4H8O2	20	1.2434	-0.9737	0.1671
ethanol	C2H6O	21	-0.3732	-2.4359	-0.0156
ethyl acetate	C4H8O2	22	-1.6552	0.5945	-0.3986
ethylene glyc	C2H6O2	23	3.2048	-0.7243	-0.5465
glycerin	C3H8O3	24	5.1085	0.5255	-0.4613
heptane	C7H16	25	-1.8811	0.5191	0.1730
Hexamethylp	C6H18N3OP	26	3.5834	1.1446	1.0794
Hexamethylp	C6H18N3P	27	1.2995	0.0125	1.4651
hexane	C6H14	28	-2.4073	0.0831	0.0458
methanol	CH4O	29	-0.2125	-2.7572	-0.3642
methyl t-butyl	C5H12O	30	-2.6218	0.1485	-0.0357
methylene ch	CH2Cl2	31	-1.4643	1.0366	-1.1397

- A correlation loading plot of PC3 vs. PC2 revealed that the binary variable Insoluble spanned the third PC

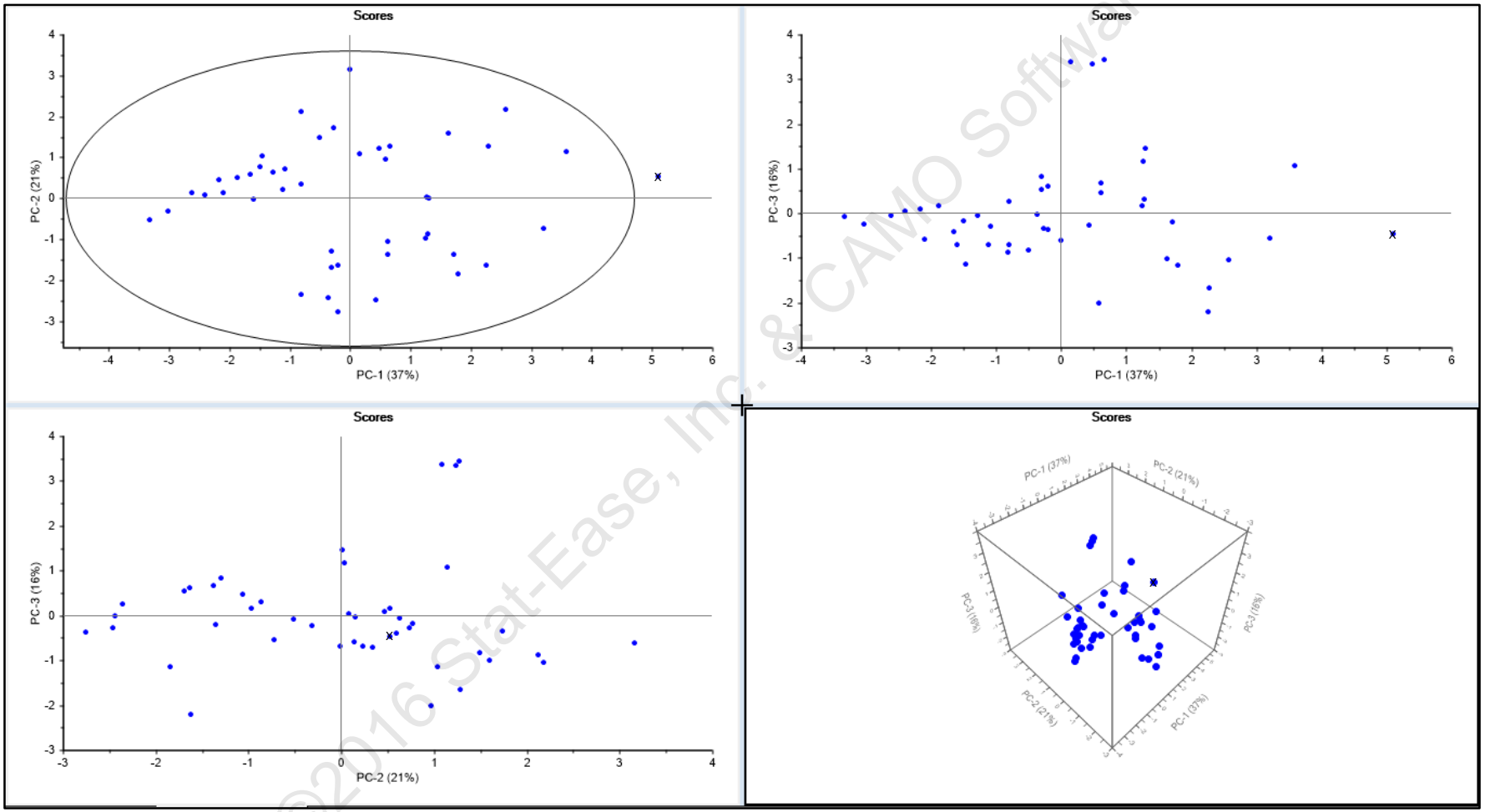


Stat-Ease statistics made easy

CAMO Bring data to life

First Three Principal Components

45 Organic Solvents

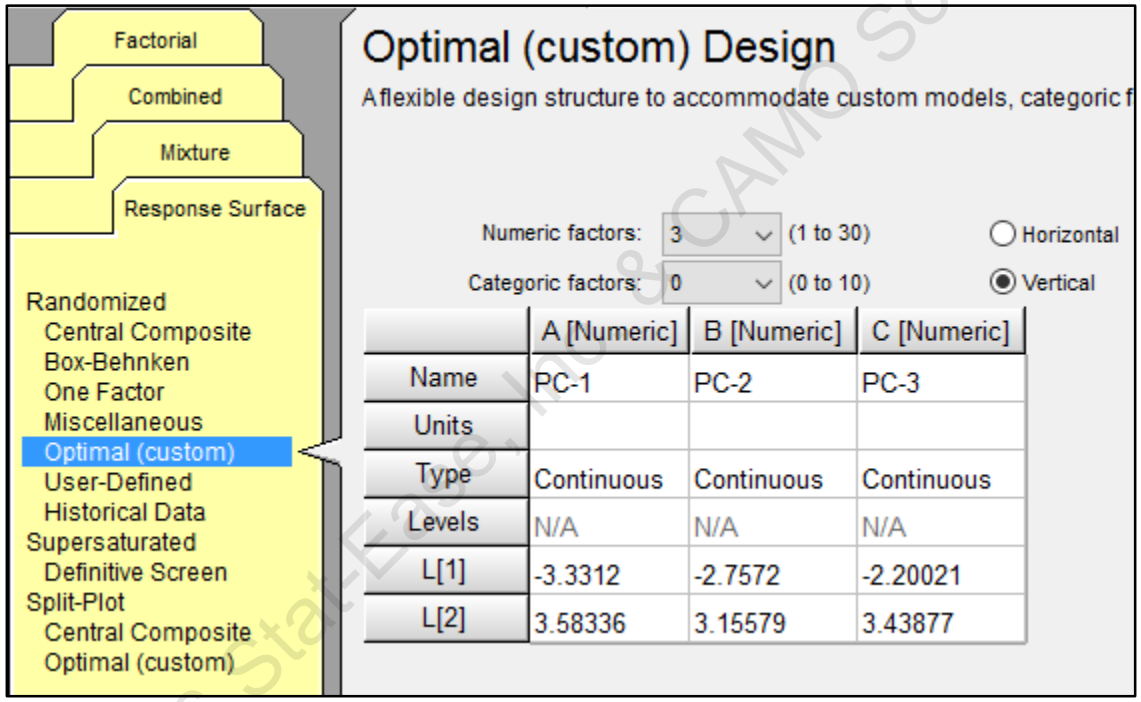


Stat-Ease statistics made easy

CAMO Bring data to life

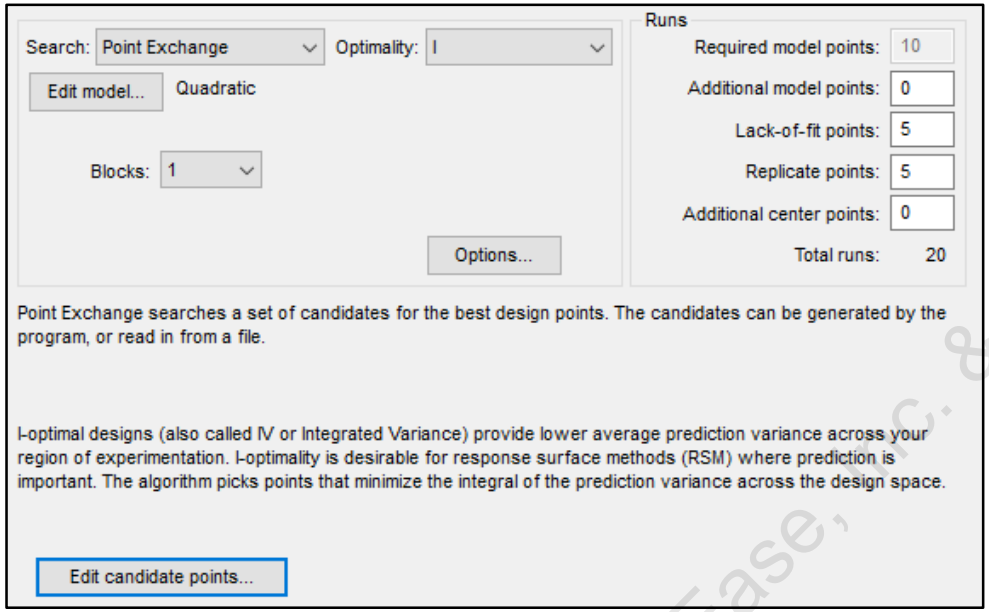
First Three Principal Components

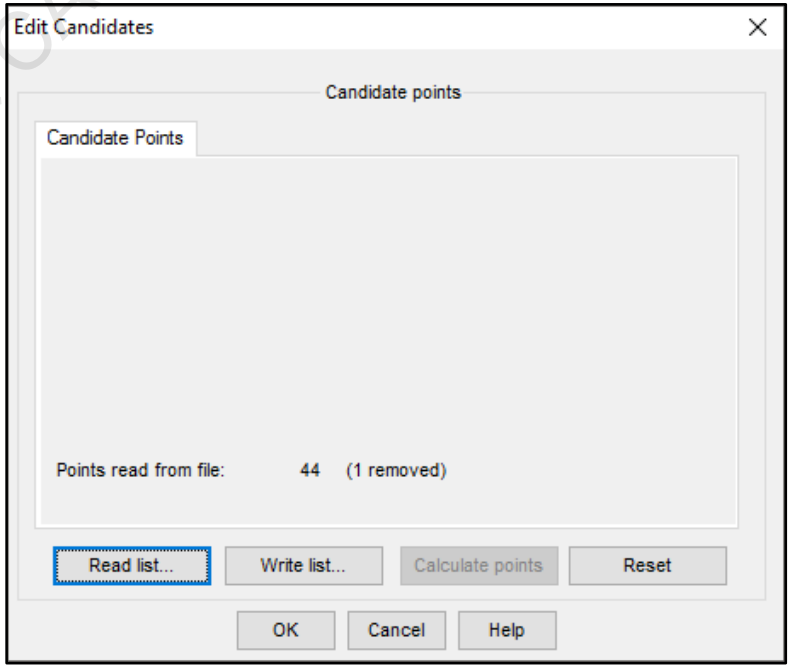
Build an Optimal Design *(page 1 of 3)*



Choose Design Points *(from 44 solvents)*

Build an Optimal Design *(page 2 of 3)*



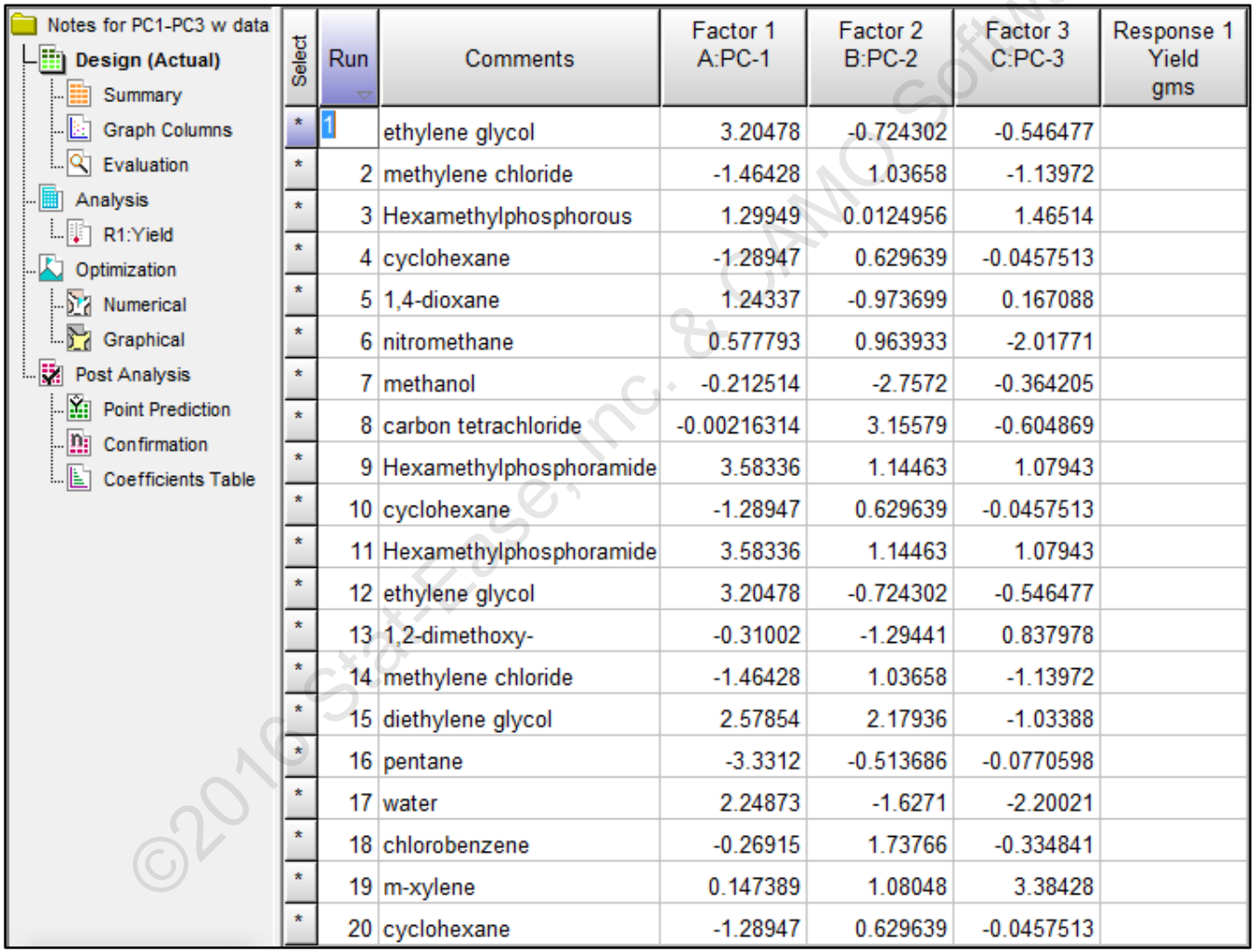


Stat-Ease statistics made easy

CAMO Bring data to life

First Three Principal Components

Build an Optimal Design *(page 3 of 3)*



Select	Run	Comments	Factor 1 A:PC-1	Factor 2 B:PC-2	Factor 3 C:PC-3	Response 1 Yield gms
*	1	ethylene glycol	3.20478	-0.724302	-0.546477	
*	2	methylene chloride	-1.46428	1.03658	-1.13972	
*	3	Hexamethylphosphorous	1.29949	0.0124956	1.46514	
*	4	cyclohexane	-1.28947	0.629639	-0.0457513	
*	5	1,4-dioxane	1.24337	-0.973699	0.167088	
*	6	nitromethane	0.577793	0.963933	-2.01771	
*	7	methanol	-0.212514	-2.7572	-0.364205	
*	8	carbon tetrachloride	-0.00216314	3.15579	-0.604869	
*	9	Hexamethylphosphoramide	3.58336	1.14463	1.07943	
*	10	cyclohexane	-1.28947	0.629639	-0.0457513	
*	11	Hexamethylphosphoramide	3.58336	1.14463	1.07943	
*	12	ethylene glycol	3.20478	-0.724302	-0.546477	
*	13	1,2-dimethoxy-	-0.31002	-1.29441	0.837978	
*	14	methylene chloride	-1.46428	1.03658	-1.13972	
*	15	diethylene glycol	2.57854	2.17936	-1.03388	
*	16	pentane	-3.3312	-0.513686	-0.0770598	
*	17	water	2.24873	-1.6271	-2.20021	
*	18	chlorobenzene	-0.26915	1.73766	-0.334841	
*	19	m-xylene	0.147389	1.08048	3.38428	
*	20	cyclohexane	-1.28947	0.629639	-0.0457513	

Simulate Response, Analyze and Maximize

Select	Run	Comments	Factor 1 A:PC-1	Factor 2 B:PC-2	Factor 3 C:PC-3	Response 1 Yield gms
*	1	ethylene glycol	3.20478	-0.724302	-0.546477	22.32
*	2	methylene chloride	-1.46428	1.03658	-1.13972	31.4
*	3	Hexamethylphosphorous	1.29949	0.0124956	1.46514	57.07
*	4	cyclohexane	-1.28947	0.629639	-0.0457513	38.1
*	5	1,4-dioxane	1.24337	-0.973699	0.167088	47.02
*	6	nitromethane	0.577793	0.963933	-2.01771	22.29
*	7	methanol	-0.212514	-2.7572	-0.364205	16.27
*	8	carbon tetrachloride	-0.00216314	3.15579	-0.604869	36.72
*	9	Hexamethylphosphoramide	3.58336	1.14463	1.07943	31.54
*	10	cyclohexane	-1.28947	0.629639	-0.0457513	33.95
*	11	Hexamethylphosphoramide	3.58336	1.14463	1.07943	33.82
*	12	ethylene glycol	3.20478	-0.724302	-0.546477	21.56
*	13	1,2-dimethoxy-	-0.31002	-1.29441	0.837978	42.21
*	14	methylene chloride	-1.46428	1.03658	-1.13972	30.11
*	15	diethylene glycol	2.57854	2.17936	-1.03388	29.22
*	16	pentane	-3.3312	-0.513686	-0.0770598	9.96
*	17	water	2.24873	-1.6271	-2.20021	0.23
*	18	chlorobenzene	-0.26915	1.73766	-0.334841	39.63
*	19	m-xylene	0.147389	1.08048	3.38428	64.01
*	20	cyclohexane	-1.28947	0.629639	-0.0457513	44.03

- PC1-PC3 w data.dxpx

Simulate Response, Analyze and Maximize

Transform | Fit Summary | Model | ANOVA

Response 1 Yield

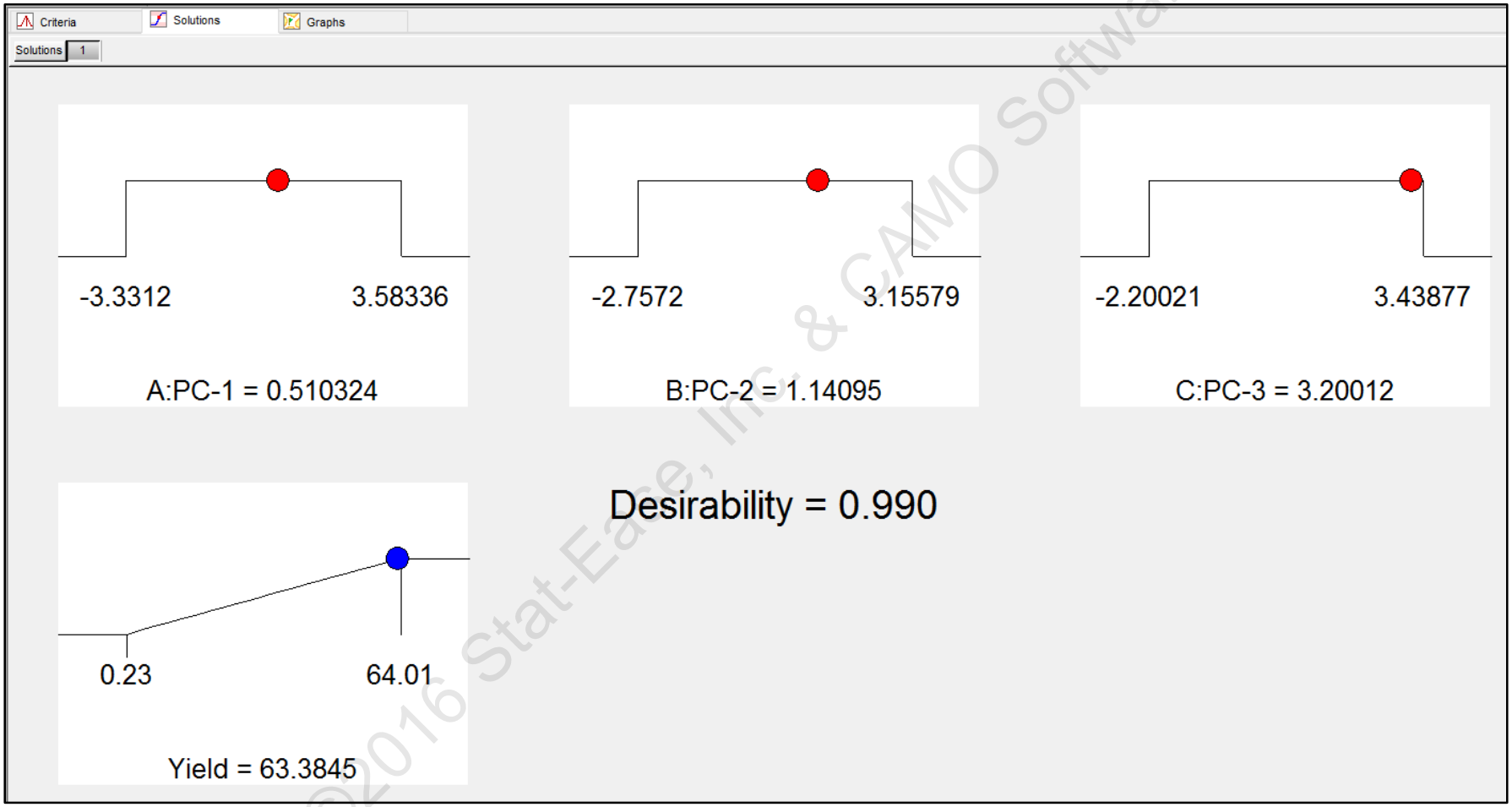
ANOVA for Response Surface Reduced Quadratic model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	4121.51	6	686.92	50.79	< 0.0001
A-PC-1	*172.70*	*1*	*172.70*	*12.77*	*0.0034*
B-PC-2	*404.07*	*1*	*404.07*	*29.87*	*0.0001*
C-PC-3	*1385.73*	*1*	*1385.73*	*102.45*	*< 0.0001*
A²	*1241.58*	*1*	*1241.58*	*91.79*	*< 0.0001*
B²	*307.74*	*1*	*307.74*	*22.75*	*0.0004*
C²	*199.83*	*1*	*199.83*	*14.77*	*0.0020*
Residual	175.83	13	13.53		
Lack of Fit	*120.78*	*8*	*15.10*	*1.37*	*0.3789*
Pure Error	*55.05*	*5*	*11.01*		
Cor Total	4297.34	19			

Std. Dev.	3.68	R-Squared	0.9591
Mean	32.57	Adj R-Squarec	0.9402
C.V. %	11.29	Pred R-Square	0.8932

Simulate Response, Analyze and Maximize

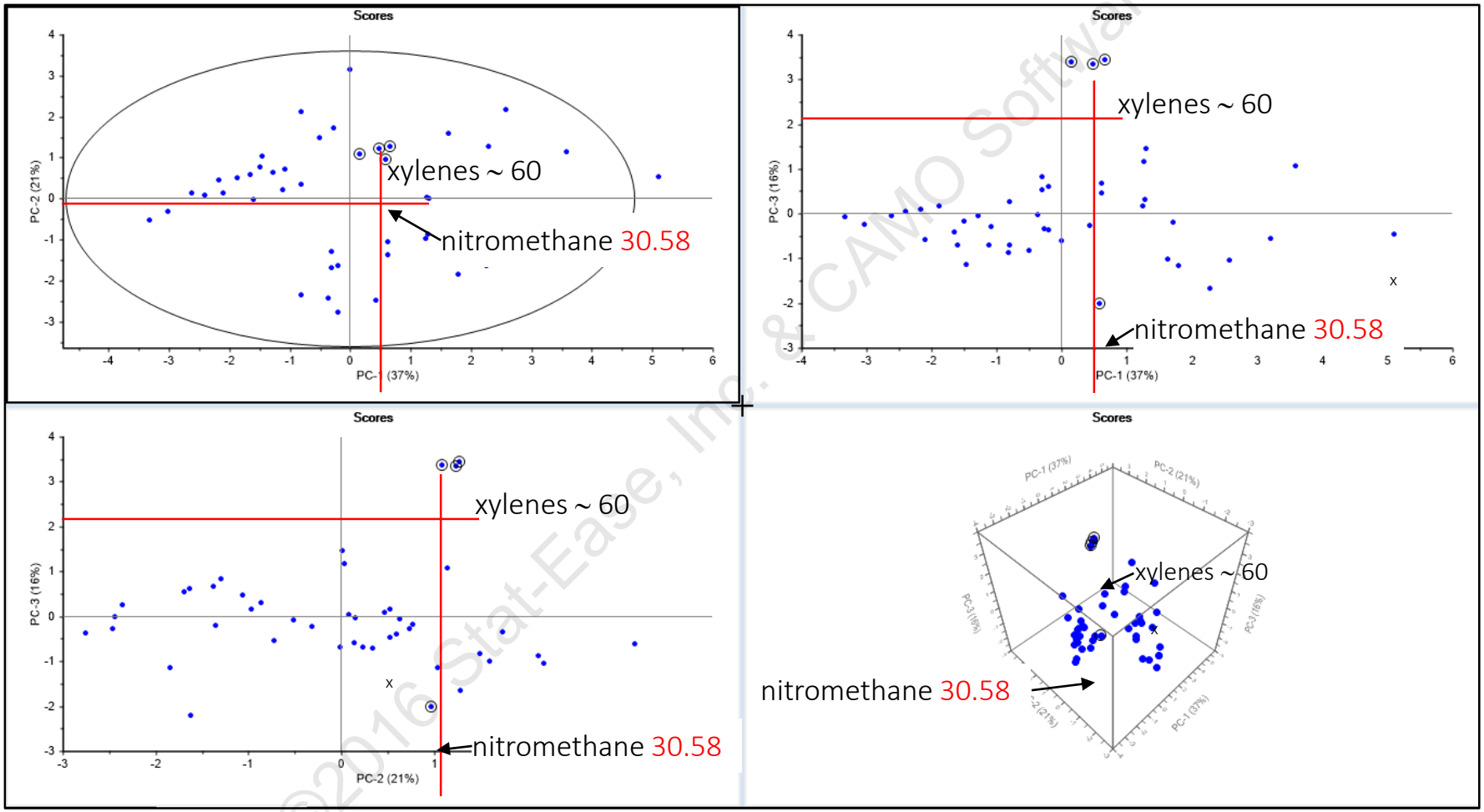


Stat-Ease statistics made easy

CAMO Bring data to life

Points near Optimum

PC-1 = 0.5, PC-2 = 1.1 and PC-3 = 3.2



Stat-Ease statistics made easy

CAMO Bring data to life

THANK YOU!

Frank Westad

CAMO Software

fw@camo.com

Pat Whitcomb

Stat-Ease

pat@statease.com

The presentation and recording will be posted on
http://www.camo.com/training/webinars-seminars.html