Mixture design

Mixture designs represent a special class of DoE based on response surface, in which the response is related to a mixture made up of different ingredients, e.g., an insecticide that blends four compounds, a food product including several components, a multi-solvent mobile phase for liquid chromatography.

In this case factors represent the proportions of ingredients; thus, they cannot be varied independently.

Typical assumptions of mixture designs are:

1) errors are normally (and independently) distributed with mean equal to zero and constant variance;
2) the real response surface is continuous over the entire domain of factors;
3) response depends only on the proportions of single ingredients, not on the amount of the mixture.

Given q components of a mixture, each with a X_j proportion, the sum of such proportions must equate a fixed total **T** (T = 100 if percentages are adopted):

$$X_1 + X_2 + \ldots + X_q = T,$$

Each component may eventually be limited by a lower, L_j, and an upper limit, U_j:

$$L_j \leq X_j \leq U_j$$

Since each component might be expressed with different units, the use of the so-called **pseudo-components** can be convenient.

A pseudo-component, indicated with x_j, is defined so that its values become 0 and 1 in correspondence of values L_j and U_j for the actual component:

$$x_j = \frac{X_j - L_j}{T - \sum_{j=1}^{q} L_j}$$

Note that the denominator of the x_j expression corresponds to the difference existing between the maximum and the minimum value of each component.
Indeed, the difference existing between the total T and the lower limits of all the other components but the j-th one, corresponds to the upper limit of the j-th component. According to the expression in the denominator, this is then subtracted of the L_j value itself.

As an example, let us consider a mixture of two components, whose total volume is **T** = 50 mL.

If L_1 = 20 mL, then U_2 = 30; if L_2 = 0, then U_1 = 50.

The table of components and pseudo-components is the following:

Components		Pseudo-components	
X_1	X_2	x_1	x_2
50	0	1.0	0.0
35	15	0.5	0.5
20	30	0.0	1.0

Indeed, for the first experiment:

$x_1 = (50 - 20) / [50 - (20 + 0)] = 1.0$ and $x_2 = (0 - 0) / [50 - (20 + 0)] = 0.0$

For the second experiment:

$x_1 = (35 - 20) / [50 - (20 + 0)] = 15/30 = 0.5$ and $x_2 = (15 - 0) / [50 - (20 + 0)] = 15/30 = 0.5$

For the third experiment:

$x_1 = (20 - 20) / [50 - (20 + 0)] = 0.0$ and $x_2 = (30 - 0) / [50 - (20 + 0)] = 30/30 = 1.0$

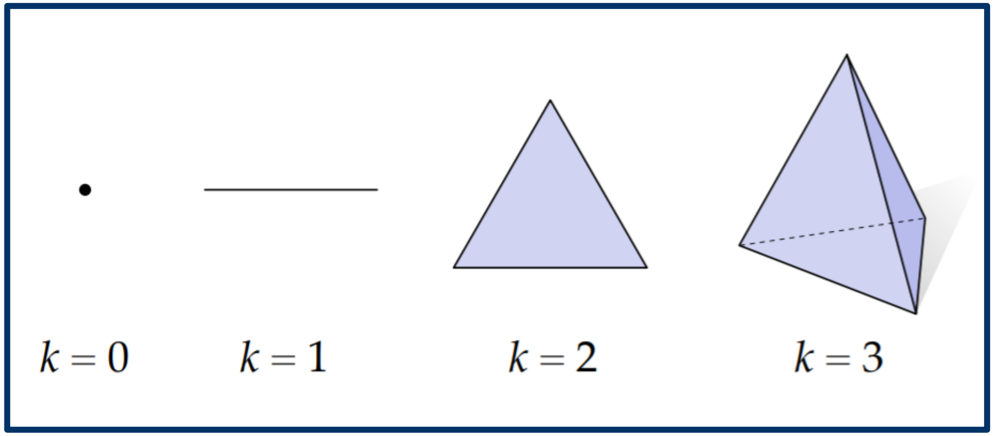
By definition, the sum of pseudo-components is always equal to 1:

$$\sum_{i=1}^{q} x_i = x_1 + x_2 + \ldots + x_q = 1.0$$

From a geometrical point of view this equation represents a simplex with q-1 dimensions.

As a general definition, a k-dimensional simplex is a k-dimensional polytope corresponding to the convex envelope of k+1 vertices.

For k = 0 the simplex is a point; for k = 1 it is a segment; for k = 2 it is a triangle; for k = 3 it is a tetrahedron:



In a mixture design all experimental points lie on or inside the simplex.

Types of mixture design

1. Simplex-lattice – a design consisting in a number of experiments uniformly spaced on a (q-1) dimensional simplex, where q represents the number of components in the mixture. It can be not available if some constraints are present.

2. Simplex-centroid – a design consisting in $(2^q - 1)$ experiments, performed with all components taken alone, all binary, tertiary, etc., mixtures up to the design centroid. It can be not available in the presence of some constraints.

3. Extreme vertices – a design consisting in an experiment for each vertex (always available).

Mathematical models for mixture designs

Mathematical models for mixture designs take into account the fundamental mixture constraint. They are usually of three types:

1. First-degree

2. Second-degree or quadratic models

3. Third-degree or cubic models

First-degree models

In this case the basic assumption is that changes in the response depend only on the proportions of single components in the mixture.

For a three-component mixture the model can be thus written, in terms of pseudo-components, as:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$$

However, the constraint between pseudo-components has to be considered:

$$x_1 + x_2 + x_3 = 1$$

This equation can be integrated in the model:

$$y = a_0 (x_1 + x_2 + x_3) + a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\Downarrow$$

$$y = (a_0 + a_1) x_1 + (a_0 + a_2) x_2 + (a_0 + a_3) x_3$$

If the following new parameters are introduced:

$$b_1 = a_1 + a_0 \qquad b_2 = a_2 + a_0 \qquad b_3 = a_3 + a_0$$

The model can be written as follows:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3$$

The consequence of the constraint existing between factors is the absence of a constant term in the model.

Second-degree models

The second-degree (or quadratic) mathematical model contains first degree terms, crossed terms and squared terms.

Considering that there is no constant, due to the constraint between factors, the model can be written as follows, in the case of two components:

$$y = a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + a_{11} x_1^2 + a_{22} x_2^2$$

Since: $x_1 + x_2 = 1$ which can be written in terms of x_1: $x_1 = 1 - x_2$

the following equations can be easily obtained: $x_1^2 = x_1 (1 - x_2) = x_1 - x_1 x_2$

A similar equation can be written for x_2^2, thus square terms are, in fact, equal to a first-degree term and a crossed term.

The model can then be written as: $y = b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2$

with: $b_1 = a_1 + a_{11}$ $b_2 = a_2 + a_{22}$ $b_{12} = a_{12} - a_{11} - a_{22}$

By analogy, for three components the quadratic model becomes:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3$$

Third-degree models

The complete third-degree model, also named complete cubic model, in the case of three components can be obtained starting from the following polynomial equation:

$$y = b'_0 + b'_1 x_1 + b'_2 x_2 + b'_3 x_3 + b'_{11} x_1^2 + b'_{22} x_2^2 + b'_{33} x_3^2 + b'_{12} x_1 x_2$$
$$+ b'_{13} x_1 x_3 + b'_{23} x_2 x_3 + b'_{112} x_1^2 x_2 + b'_{122} x_1 x_2^2 + b'_{113} x_1^2 x_3$$
$$+ b'_{133} x_1 x_3^2 + b'_{223} x_2^2 x_3 + b'_{233} x_2 x_3^2 + b'_{123} x_1 x_2 x_3$$
$$+ b'_{111} x_1^3 + b'_{222} x_2^3 + b'_{333} x_3^3$$

However, due to constraints existing between mixture components (e.g., $x_1 = 1 - x_2 - x_3$), the equation can be simplified as follows:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + g_{12} x_1 x_2 (x_1 - x_2) +$$
$$g_{13} x_1 x_3 (x_1 - x_3) + g_{23} x_2 x_3 (x_2 - x_3) + b_{123} x_1 x_2 x_3$$

Actually, a simplified cubic mathematical model, also named special or restricted cubic model, containing first-degree terms, crossed terms and a supplementary term corresponding to the product of the three components is commonly adopted:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3$$

Simplex-Lattice Designs

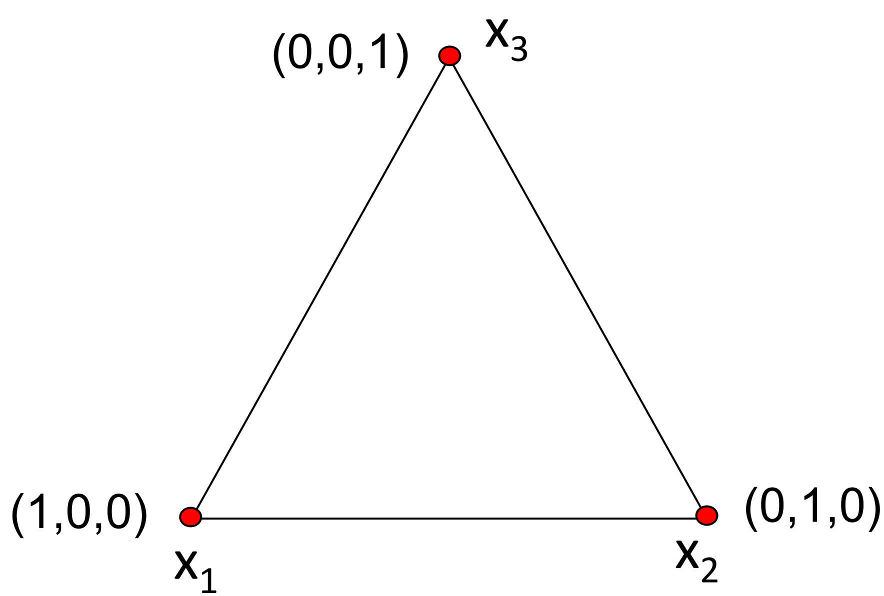
Based on the general requirement (experiments spaced uniformly on a (q-1)-dimensional simplex), the number of experiments for a simplex-lattice design depends on the model order, since for a model of order m the proportions assumed by each component are:

$$x_j = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 1$$

For m = 1 (linear model) and q = 3 each component can only assume values 0 and 1, thus 3 experiments are required:

Pure components

$(x_1=1,\ x_2=0,\ x_3=0)$
$(x_1=0,\ x_2=1,\ x_3=0)$
$(x_1=0,\ x_2=0,\ x_3=1)$



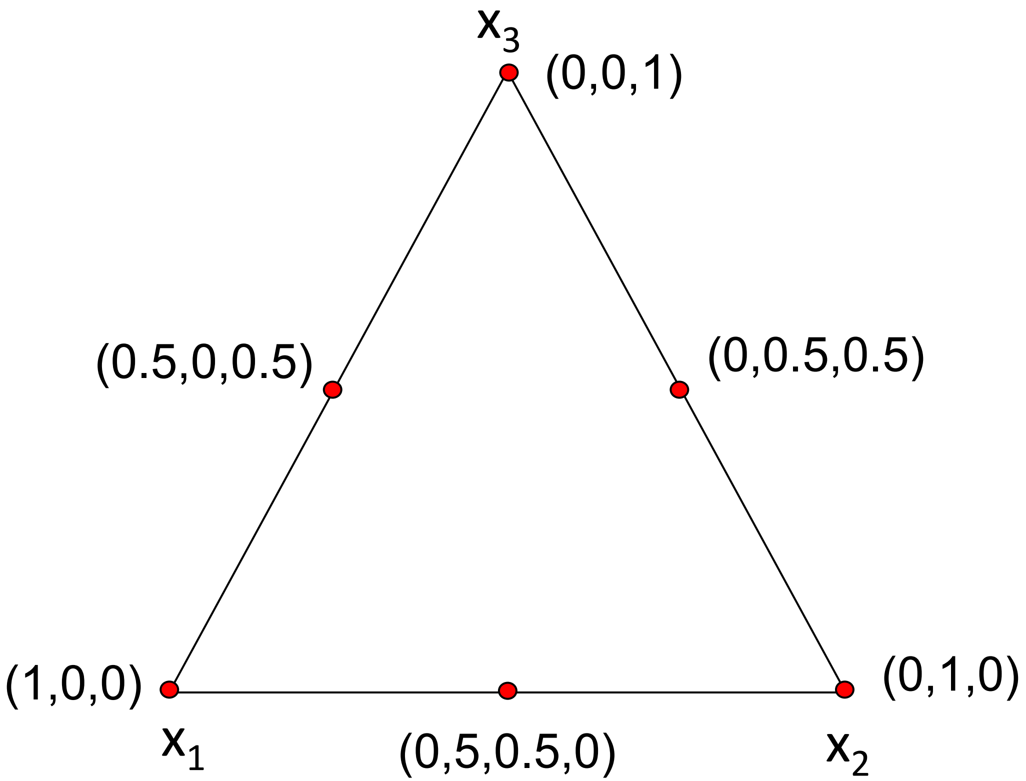
For m = 2 (quadratic model) and q = 3, each component can assume values 0, 0.5 and 1, thus 6 experiments are required:

Pure components

$(x_1=1,\ x_2=0,\ x_3=0)$
$(x_1=0,\ x_2=1,\ x_3=0)$
$(x_1=0,\ x_2=0,\ x_3=1)$

Binary mixtures

$(x_1=.5,\ x_2=.5,\ x_3=.0)$
$(x_1=.5,\ x_2=.0,\ x_3=.5)$
$(x_1=.0,\ x_2=.5,\ x_3=.5)$



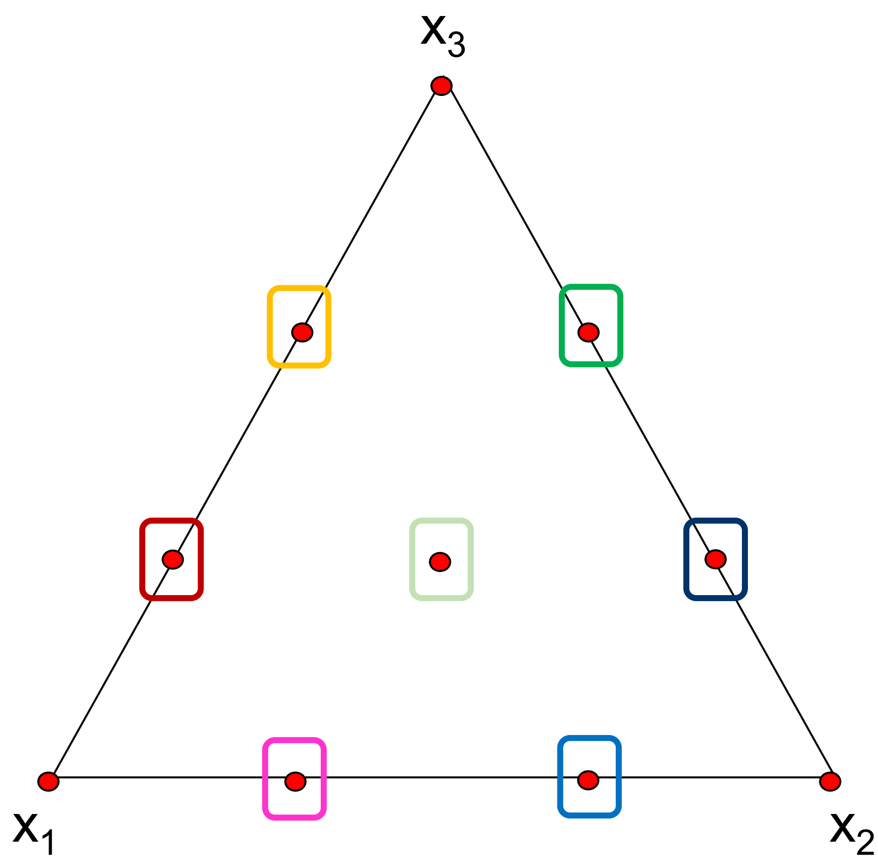
For m = 3 (cubic model) and q = 3, each component can assume values 0, 1/3, 2/3 and 1, thus 10 experiments are required:

Pure components

$(x_1=1,\ x_2=0,\ x_3=0)$
$(x_1=0,\ x_2=1,\ x_3=0)$
$(x_1=0,\ x_2=0,\ x_3=1)$

Binary/ternary mixtures

$(x_1=.667,\ x_2=.333,\ x_3=.000)$
$(x_1=.667,\ x_2=.000,\ x_3=.333)$
$(x_1=.000,\ x_2=.667,\ x_3=.333)$
$(x_1=.333,\ x_2=.667,\ x_3=.000)$
$(x_1=.333,\ x_2=.000,\ x_3=.667)$
$(x_1=.000,\ x_2=.333,\ x_3=.667)$
$(x_1=.333,\ x_2=.333,\ x_3=.333)$



Simplex-Centroid Designs

Simplex-Centroid Designs include all possible combinations of components:

q permutations of (1, 0, 0, …,0)

q permutations of (1/2, 1/2, 0, …., 0)

……………………………………………………

centroid (1/q, 1/q,….,1/q)

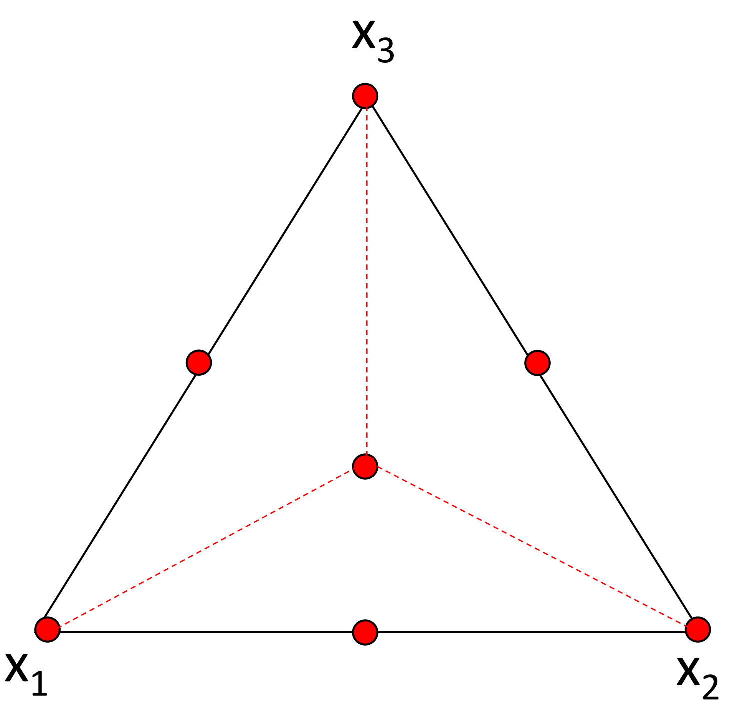
As an example, for q = 3, the combinations are:

(1, 0, 0); (0, 1, 0); (0, 0, 1)

(1/2, 1/2, 0); (1/2, 0, 1/2); (0, 1/2, 1/2)

(1/3, 1/3, 1/3)

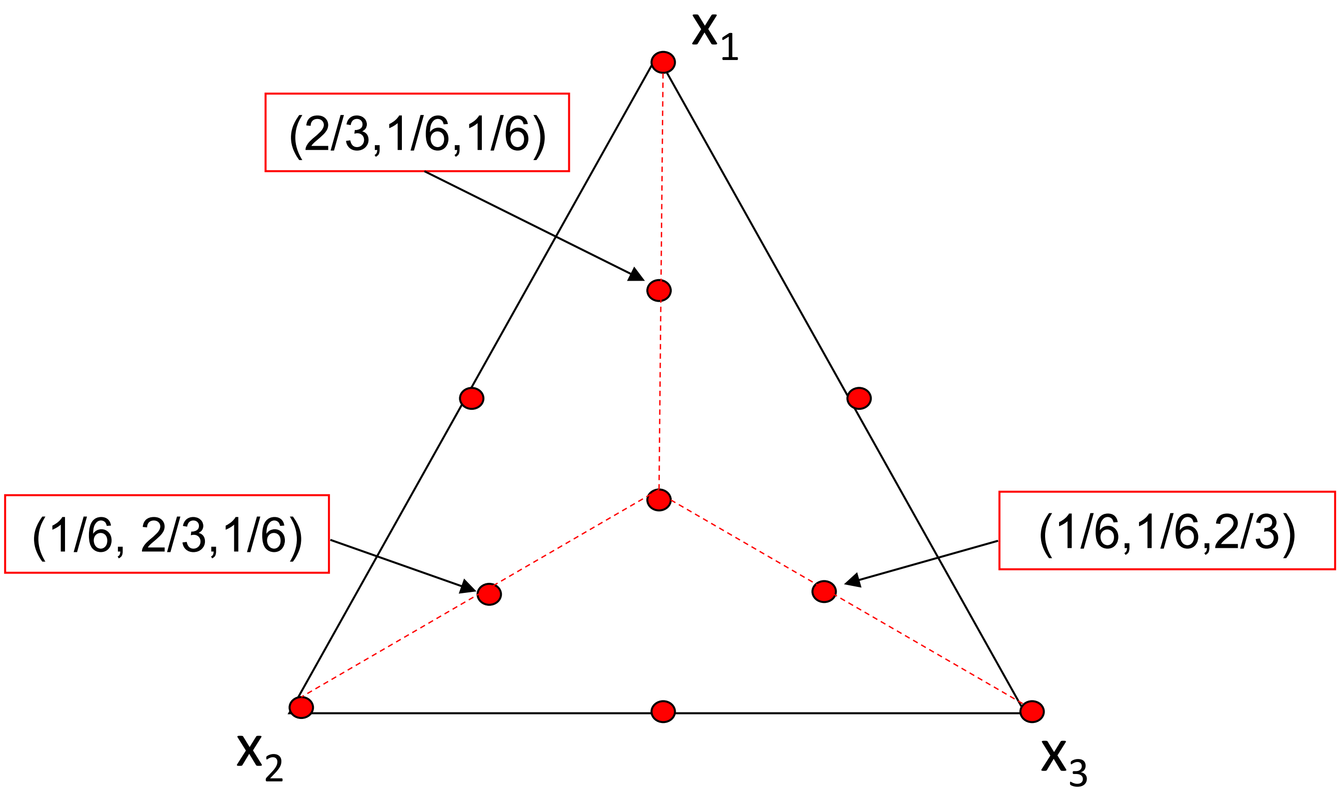
The total number of experiments is $(2^q – 1) = 7$



Additional experiments

Further experiments (augmented points) can be performed in a mixture design to increase the degrees of freedom for the evaluation of lack of fit and for model significance analysis.

Typical augmented points are related to combinations placed at the centre of segments having the simplex centre and one of its vertices as extremities:



Steps to complete an optimization based on mixture design

1. Definition of problem and objectives
2. Selection of the mixture components and relative proportions
3. Identification of the response
4. Choice of the most appropriate model to fit data and of the most appropriate design to achieve a good fit and, at the same time, to evaluate the eventual lack of fit
5. Execution of experiments
6. Data analysis (e.g., by ANOVA)
7. Formulation of conclusions and recommendations.

Example of Mixture Design: formulation of a rocket propellant

The formulation of a rocket propellant, i.e., a mixture of three components: fuel, oxidizer and binder, with the aim of achieving the most satisfactory burn rate, can be considered a typical example of mixture design.

Considering that 10% of the propellant consists in an inert component, the main constraint for the three components is:

fuel + oxidizer + binder = 90%

In addition, each of the three components has a lower limit:

30% ≤ fuel
20% ≤ oxidizer
20% ≤ binder

Consequently, the remaining 20% of the mixture can be any combination of the three components.

The elaboration of a Simplex-Centroid design based on the Statgraphics software is described in the next slides.

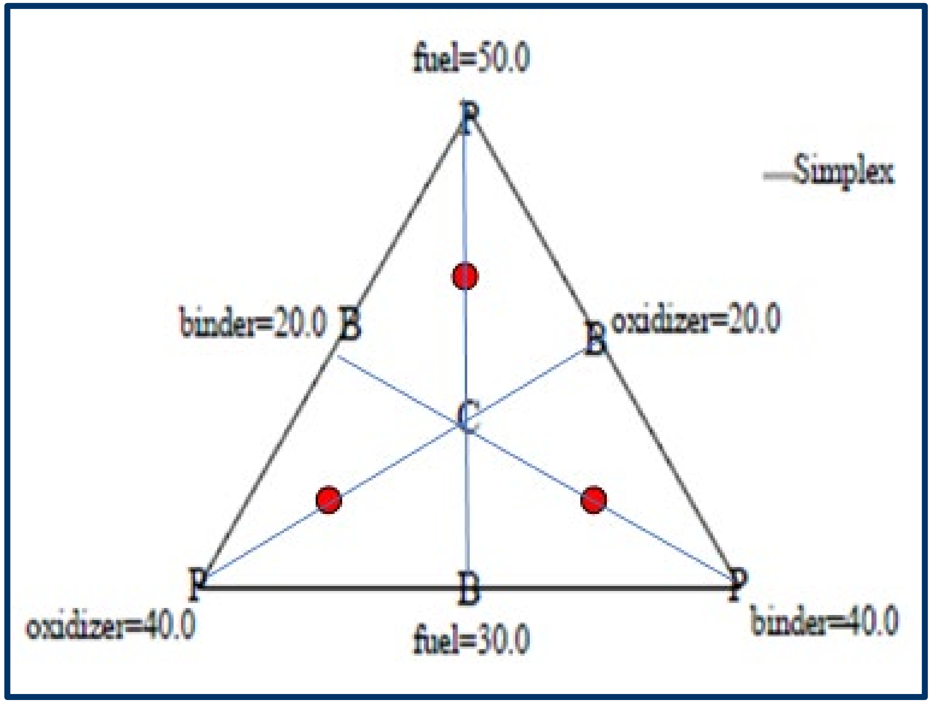
Experiments typical of the Simplex-Centroid design for three component were integrated by 3 points (augmented design); 10 runs were thus performed.

In the figure letter P indicates primary blends, i.e., mixtures in which a specific component is at its maximum value and the other at their minimum; B indicates binary blends and C the centroid. Additional experiments are indicated by red points.

Components	*Low*	*High*	*Units*
fuel	30.0	50.0	percent
oxidizer	20.0	40.0	percent
binder	20.0	40.0	percent

Mixture total = 90.0 percent

Responses	*Units*
burn rate	cm per second



5 further experiments, one at each vertex and 2 in the centroid, were performed for the evaluation of pure error and for the lack of fit test.

A summary of experimental conditions and responses is reported in the following table, whereby different types of experiments are emphasised by different colors:

		x_1	x_2	x_3	response
	run	fuel (percent)	oxidizer (percent)	binder (percent)	burn rate (cm per second)
Vertices (P)	1	50.0	20.0	20.0	32.5
	2	30.0	40.0	20.0	54.5
	3	30.0	20.0	40.0	64.0
Binary blends (B)	4	40.0	30.0	20.0	44.0
	5	40.0	20.0	30.0	63.2
	6	30.0	30.0	30.0	94.0
Centroid	7	36.6667	26.6667	26.6667	112.5
Additional runs	8	43.3333	23.3333	23.3333	67.1
	9	33.3333	33.3333	23.3333	73.0
	10	33.3333	23.3333	33.3333	87.5
Replicates	11	50.0	20.0	20.0	37.9
	12	30.0	40.0	20.0	32.5
	13	30.0	20.0	40.0	78.5
	14	36.6667	26.6667	26.6667	98.5
	15	36.6667	26.6667	26.6667	103.6

Notably, the mixture composition at the centroid depends on the compositions at the vertices and at the half points of opposite sides (i.e., at the ends of triangle medians)

As an example, the percentages at these points for fuel are 50 and 30%. Since the centroid is located at 2/3 of the median length, the fuel percentage at the centroid is 50 - 2/3 × (50 -30) = 36.6667.

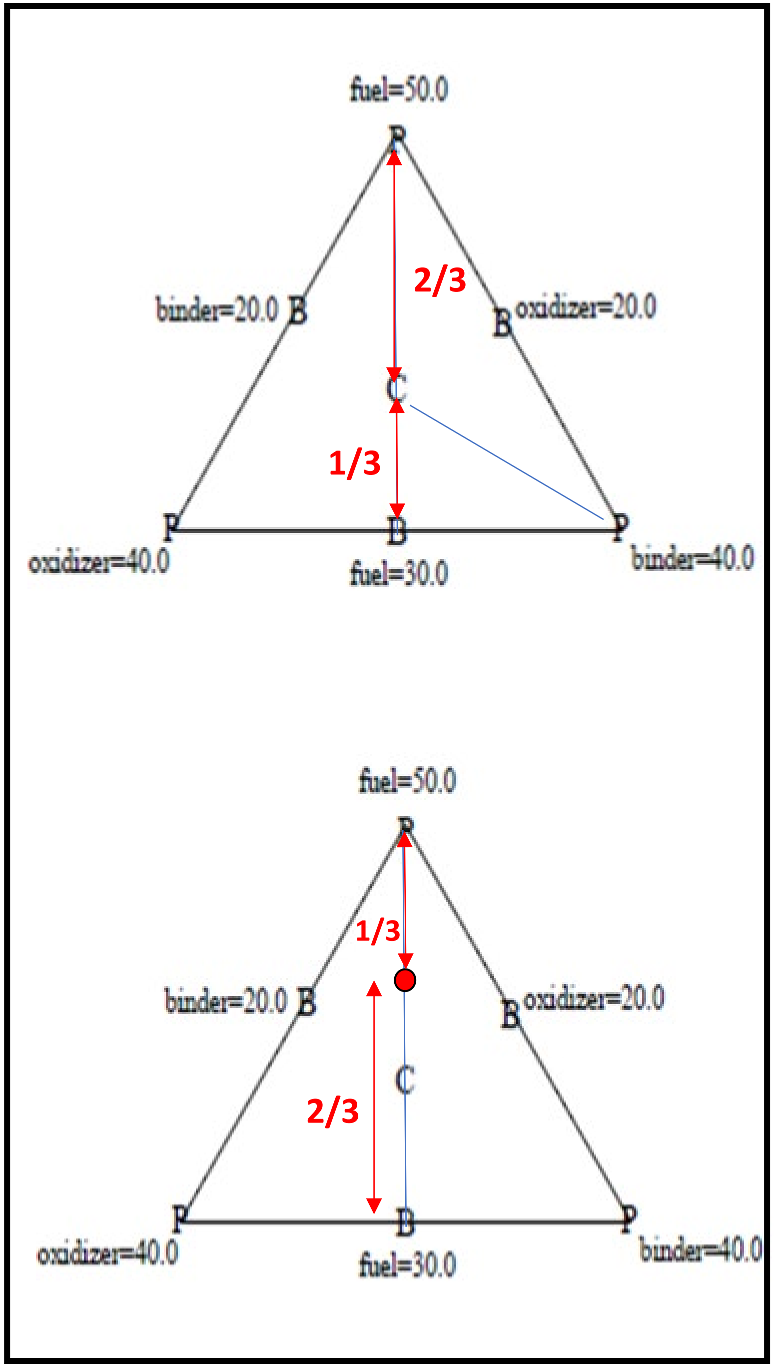
By analogy, the percentages of oxidizer and binder at the centroid are:

40 - 2/3 × (40 -20) = 26.6667

Additional points are located at 1/3 of each median length, starting from the corresponding vertex, thus at the additional point close to the fuel vertex, shown in the figure, fuel percentage is:

50 - 1/3 × (50 -30) = 43.3333

Oxidizer and binder have the same percentages at that point, i.e., (90 – 43.3333) / 2 = 23.3333.



A special cubic model was adopted in this case:

$$y = b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3$$

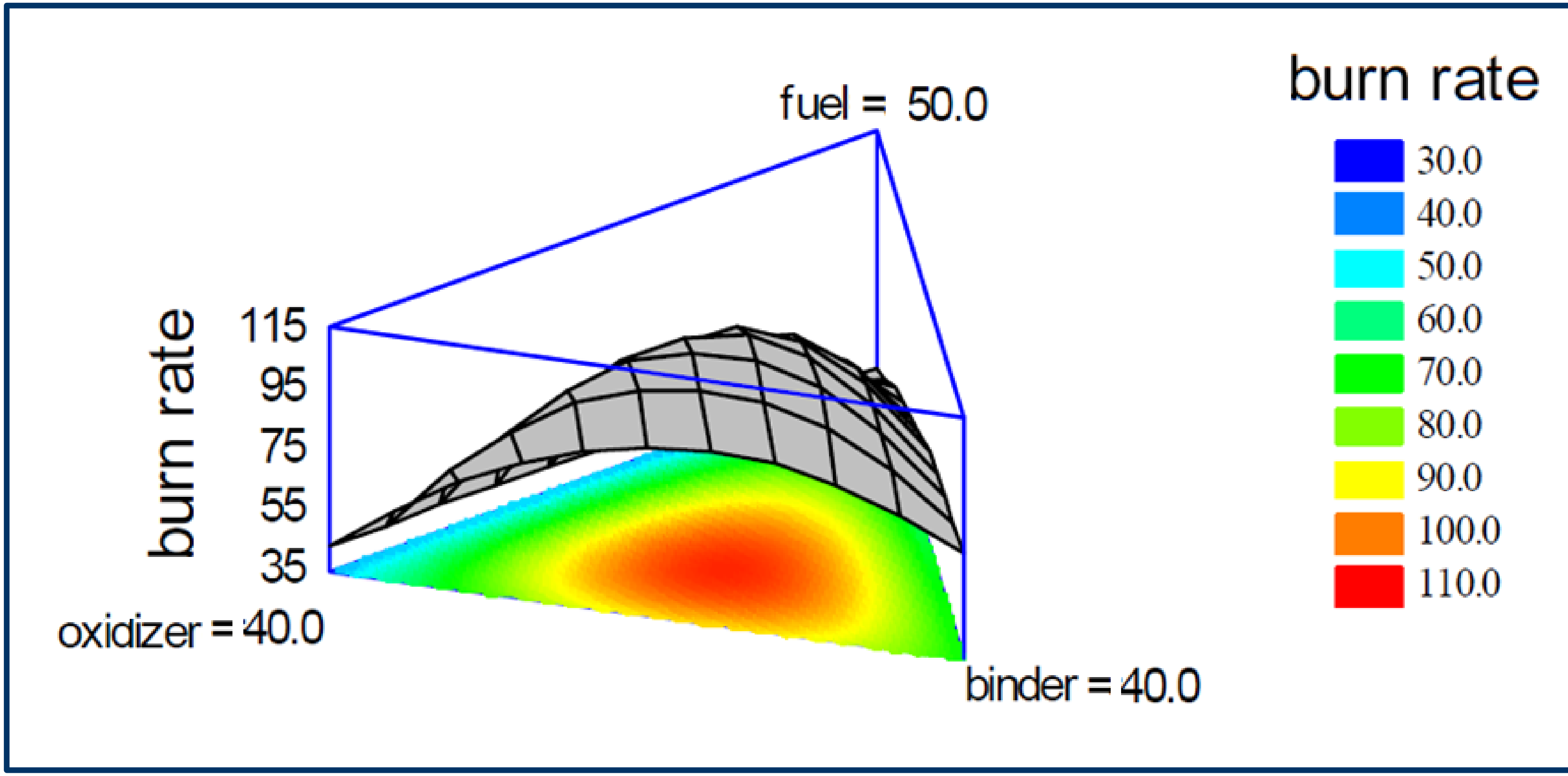
The coefficients for the special cubic model are summarized in the following table:

		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
A:fuel	35.4946	6.07193		
B:oxidizer	42.7756	6.07193		
C:binder	70.3613	6.07193		
AB	16.0213	38.2911	0.418408	0.6867
AC	36.3356	38.2911	0.948931	0.3704
BC	136.821	38.2911	3.57319	0.0073
ABC	854.962	229.174	3.73063	0.0058

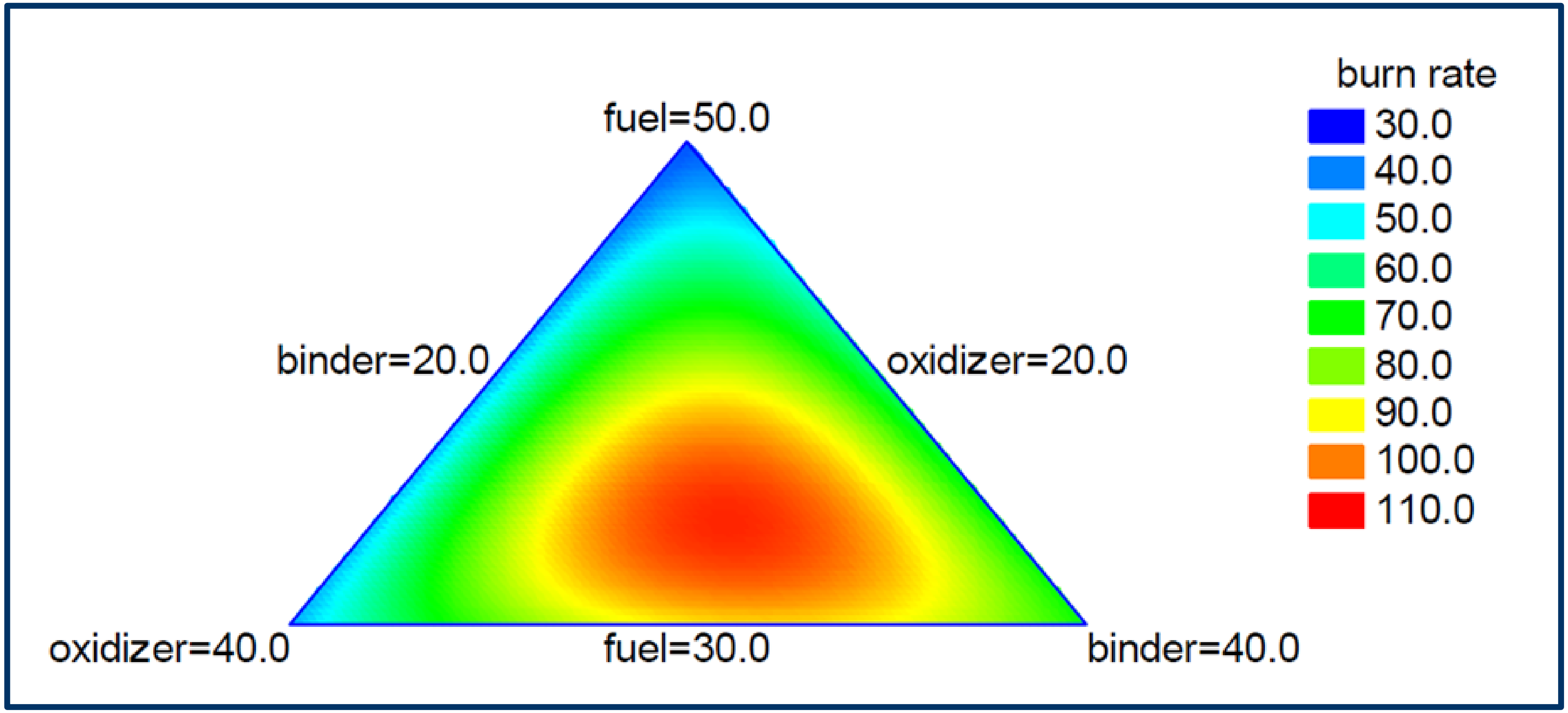
The model equation is:

burn rate = 35.4946*fuel + 42.7756*oxidizer + 70.3613*binder + 16.0213*fuel*oxidizer +
36.3356*fuel*binder + 136.821*oxidizer*binder + 854.962*fuel*oxidizer*binder

The estimated response surface, represented as a 3D-plot together with a 2D color plot, is reported in the following figure:



A detailed view of the 2D color plot is:



In the table on the right the propellant composition leading to the optimal burn rate, as assessed from the model, is reported:

Optimum value = 106.65

Factor	*Low*	*High*	*Optimum*
fuel	30.0	50.0	34.2503
oxidizer	20.0	40.0	26.8729
binder	20.0	40.0	28.8768

In the following table observed and fitted values, the corresponding residuals and the studentized residuals are reported:

Row	Observed Value	Fitted Value	Residual	Studentized Residual
1	32.5	35.4946	-2.99462	-0.451862
2	54.5	42.7756	11.7244	2.31885
3	64.0	70.3613	-6.36129	-1.01317
4	44.0	43.1404	0.859581	0.307934
5	63.2	62.0119	1.18814	0.428287
6	94.0	90.7738	3.2262	1.27419
7	112.5	102.23	10.2702	1.47548
8	67.1	67.9691	-0.869131	-0.103167
9	73.0	79.9833	-6.98335	-0.872148
10	87.5	95.4691	-7.96905	-1.01201
11	37.9	35.4946	2.40538	0.361088
12	32.5	42.7756	-10.2756	-1.87238
13	78.5	70.3613	8.13871	1.36139
14	98.5	102.23	-3.72984	-0.475496
15	103.6	102.23	1.37016	0.172283

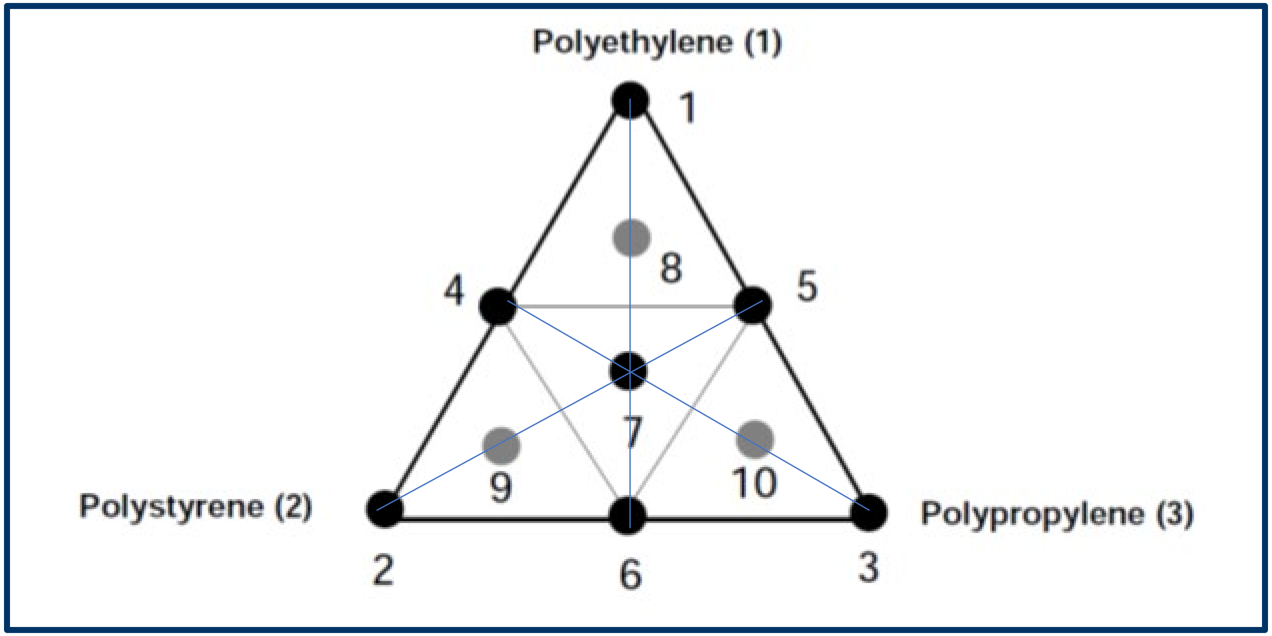
Note that studentized residuals correspond to residuals divided by an estimate of the standard deviation.

A further Example of Mixture Design: modulating the elongation of a thread through optimization of a mixture of three polymers

A mixture of polymers (polyethylene, polystyrene and polypropylene), used to fabricate a synthetic fibre, was optimized to obtain the optimal elongation of the thread.

In this case the proportions of each polymer could vary from 0 to 100%, thus the study domain was the complete equilateral triangle.

An augmented simplex-centroid design was adopted by the experimenter, with additional points, reported in grey in the figure (points 8, 9 and 10), used as control points, i.e., to verify the predictive power of the model:



The obtained results were:

Trial	Polyethylene (1)	Polystyrene (2)	Polypropylene (3)	Responses
1	1	0	0	32
2	0	1	0	25
3	0	0	1	42
4	1/2	1/2	0	38
5	1/2	0	1/2	39
6	0	1/2	1/2	30.5
7	1/3	1/3	1/3	37
8	2/3	1/6	1/6	37
9	1/6	2/3	1/6	32
10	1/6	1/6	2/3	38

The coefficients obtained for the special cubic model were:

Coefficient	Value
b_1	32
b_2	25
b_3	42
b_{12}	38
b_{13}	8
b_{23}	–12
b_{123}	6

The model equation was, then:

$$\hat{y} = 32\,x_1 + 25\,x_2 + 42\,x_3 + 38\,x_1 x_2 + 8\,x_1 x_3 - 12\,x_2 x_3 + 6\,x_1 x_2 x_3$$

The comparison between observed and predicted responses at control points is described by the table on the right:

Trial	Observed Response	Predicted Response
8	37	37.38
9	32	32.22
10	38	38.22

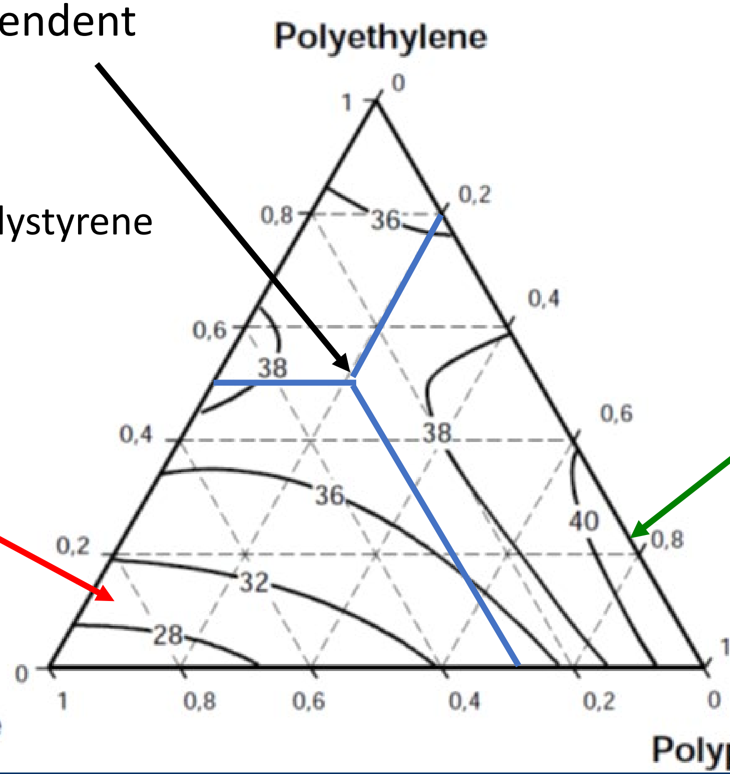
The contour plot resulting from the model provided useful indications on the process:

Thread elongation independent of composition

20% polypropylene, 30 % polystyrene 50% polyethylene

Low value of thread elongation

high polystyrene %



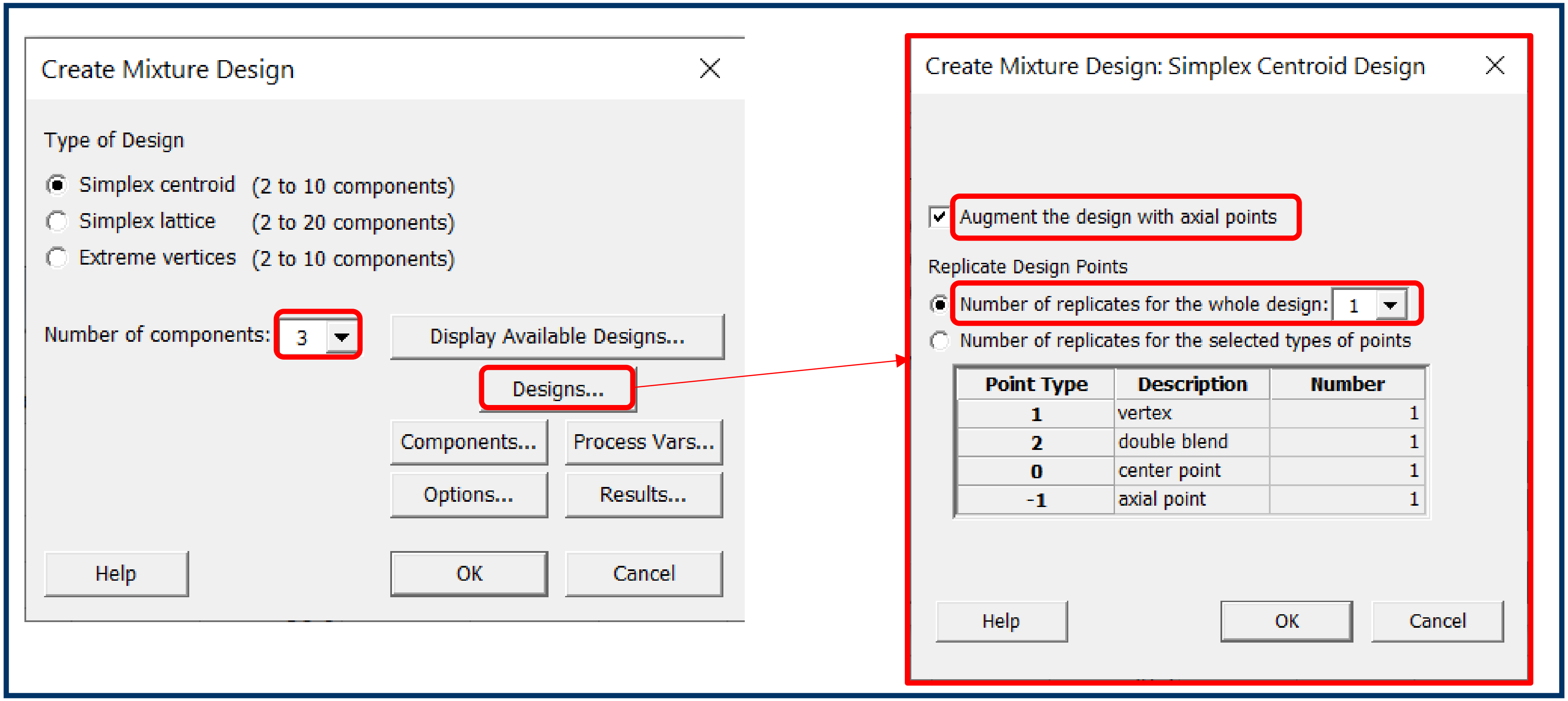
High value of thread elongation

high polypropylene %

Application of Minitab 18 to Mixture Design

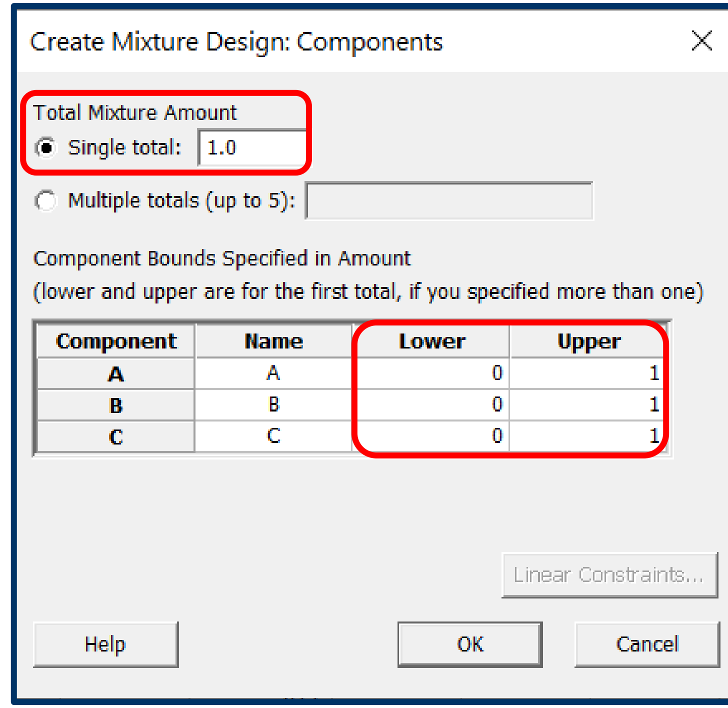
The same dataset considered for the optimization of the polymer mixture has been applied to show the application of Minitab 18 to Mixture Design.

The main menu for the setting of Mixture Designs in Minitab 18 can be accessed using the Stat > DOE > Mixture > Create Mixture Design… pathway.

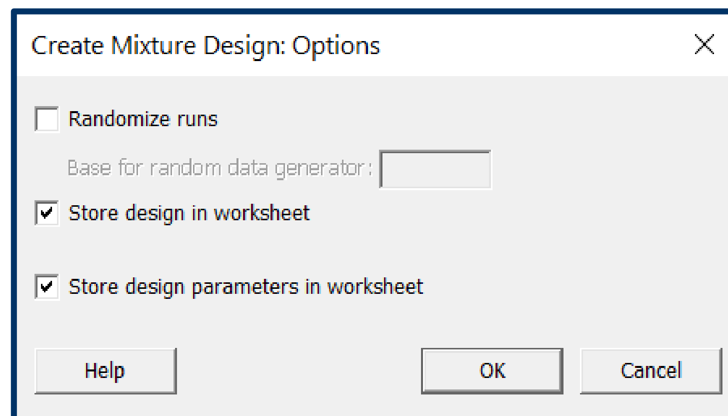


In this case the Simplex centroid design for 3 components is selected and the presence of augmented data (axial points) and the absence of replicates is specified in the Designs… window.

In the Components window the total mixture amount (1, i.e., 100%,, in the present case) is specified, along with lower and upper values of components (0 and 1, respectively, in the present case).



In the Options window the analyst can specify if randomization of runs is required and if the design and its parameters have to be stored in the worksheet.



As a consequence, all settings for runs to be performed (10 runs in this case) are automatically reported in the Worksheet:

Worksheet 1 ***

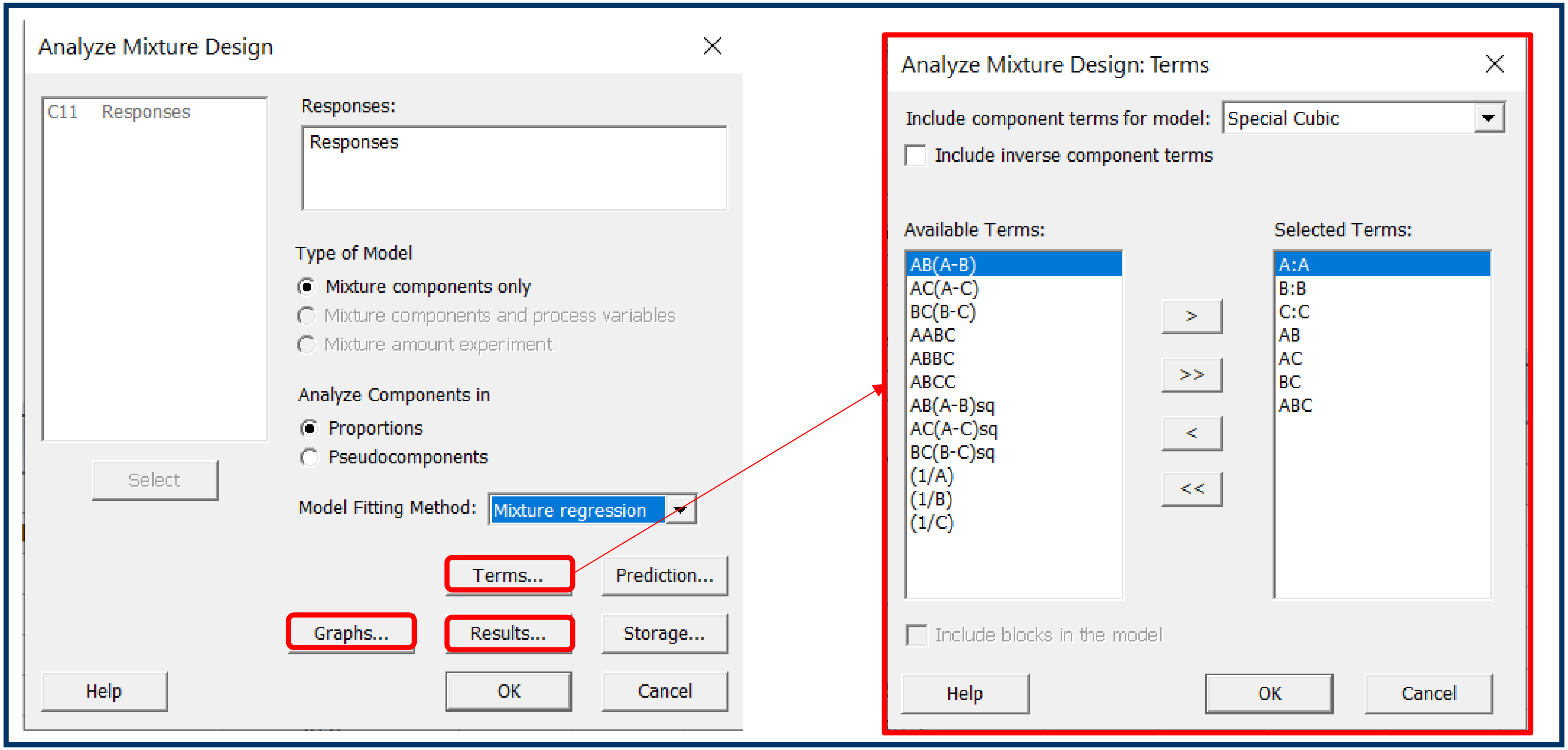
↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
	StdOrder	RunOrder	PtType	Blocks	A	B	C	Totals	Lower	Upper	Responses
1	1	1	1	1	1.00000	0.00000	0.00000	1	0	1	32.0
2	2	2	1	1	0.00000	1.00000	0.00000		0	1	25.0
3	3	3	1	1	0.00000	0.00000	1.00000		0	1	42.0
4	4	4	2	1	0.50000	0.50000	0.00000				38.0
5	5	5	2	1	0.50000	0.00000	0.50000				39.0
6	6	6	2	1	0.00000	0.50000	0.50000				30.5
7	7	7	0	1	0.33333	0.33333	0.33333				37.0
8	8	8	-1	1	0.66667	0.16667	0.16667				37.0
9	9	9	-1	1	0.16667	0.66667	0.16667				32.0
10	10	10	-1	1	0.16667	0.16667	0.66667				38.0

Note that numbers reported in the PtType (Point Type) column correspond to those shown in the table reported inside the Create Mixture Design: Simplex Centroid Design window shown before.

Point Type	Description	Number
1	vertex	1
2	double blend	1
0	center point	1
-1	axial point	1

Once experiments have been made, responses can be entered in the appropriate column of the worksheet and then model calculations can be started by accessing the Stat > DOE > Mixture > Analyze Mixture Design... pathway.

In the specific case, components are analysed in proportions (their percentages), not as pseudocomponents.
The type of model, a special cubic one, in the present case, can be specified in the Terms… window:



Different types of Graphs and Results to be shown at the end of calculations can be selected in the corresponding windows.

A table including all regression coefficients and the ANOVA table can be included among the Results and are shown in the Session window at the end of calculations.

Notably, a slight discrepancy with results shown before can be observed for the coefficient related to the A*B*C product in the regression model.

Actually, as shown by SE coefficient and by the P-Value, much higher than 0.05, the contribution of that term to the model is not statistically significant.

Estimated Regression Coefficients for Responses (component proportions)

Term	Coef	SE Coef	T-Value	P-Value	VIF
A	31.945	0.207	*	*	1.97
B	24.991	0.207	*	*	1.97
C	41.991	0.207	*	*	1.97
A*B	37.87	1.04	36.33	0.000	2.38
A*C	7.87	1.04	7.55	0.005	2.38
B*C	-12.04	1.04	-11.55	0.001	2.38
A*B*C	1.59	6.87	0.23	0.832	2.47

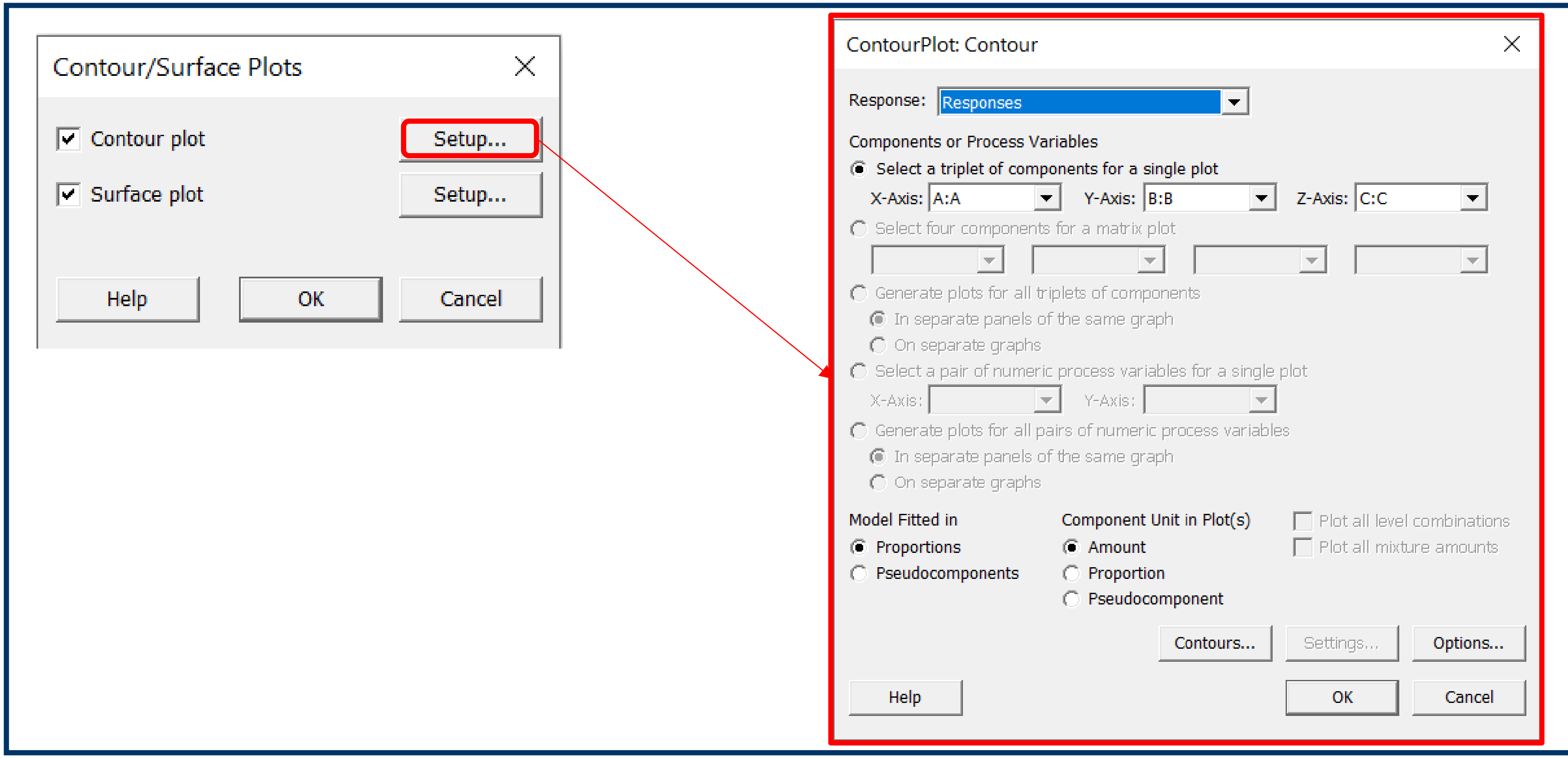
Analysis of Variance for Responses (component proportions)

Source	DF	Seq SS	Adj SS	Adj MS	F-Value	P-Value
Regression	6	229.087	229.087	38.1812	831.83	0.000
Linear	2	145.528	160.702	80.3508	1750.55	0.000
Quadratic	3	83.557	76.212	25.4041	553.46	0.000
A*B	1	73.068	60.571	60.5707	1319.62	0.000
A*C	1	3.274	2.617	2.6168	57.01	0.005
B*C	1	7.215	6.119	6.1193	133.32	0.001
Special Cubic	1	0.002	0.002	0.0025	0.05	0.832
A*B*C	1	0.002	0.002	0.0025	0.05	0.832
Residual Error	3	0.138	0.138	0.0459		
Total	9	229.225				

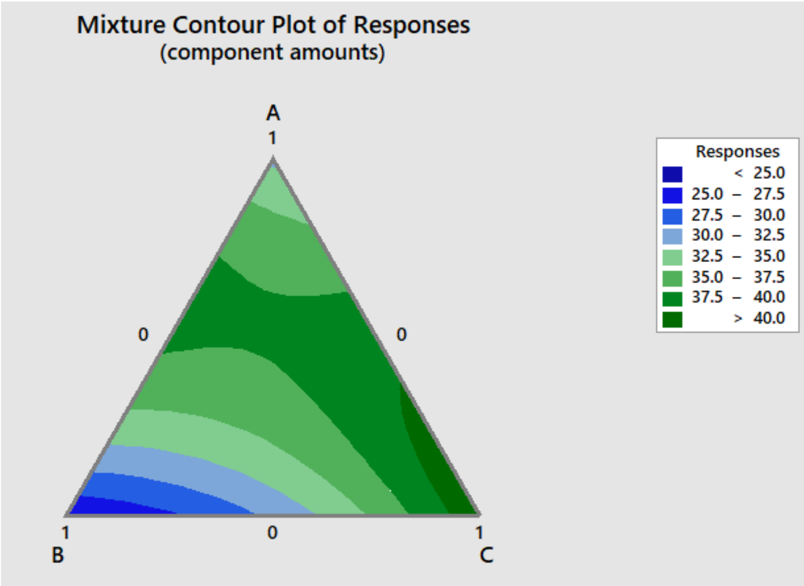
Not surprisingly, the same outcome can be inferred from the ANOVA table.

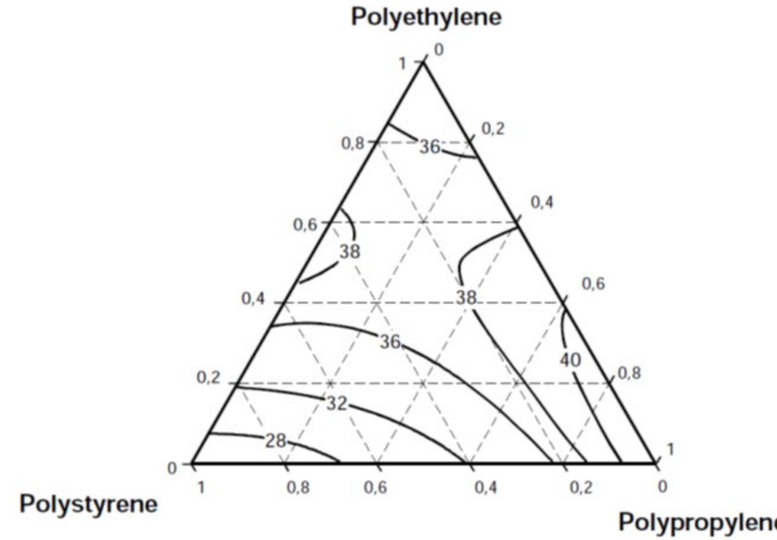
Note also that Variance Inflation Factors (VIF) were all higher than 1. This is reasonable, in a model based on mixture design, since a correlation is expected between variables in this case.

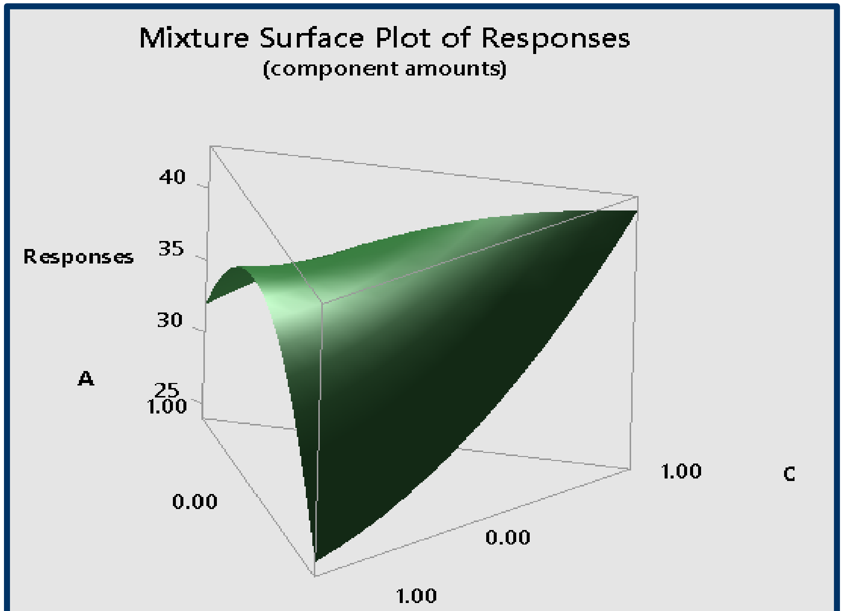
As a last step, contour and/or surface plots resulting from the calculations can be displayed by accessing the Stat > DOE > Mixture > Contour/Surface Plots… pathway.
Different graphical settings can be selected from the Setup… windows referred to the two types of plots. The one referred to the ContourPlot is reported as an example:



As apparent, the resulting contour plot is virtually identical to the one reported before:







The surface plot (that can be rotated in any direction) provides a more direct representation of the variation of response with the polymer mixture composition, emphasizing the increase in response related to the increase in the percentage of polypropylene (component C).