Designing a Better Paper Helicopter

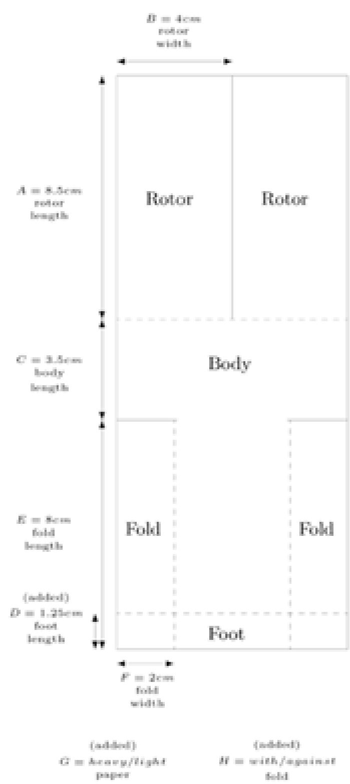
USING RESPONSE SURFACE METHODOLOGY

Webinar: Designing a Better Paper Helicopter: Using Response Surface Methodology
Author Erik Barry Erhardt discusses the article in this webinar (60 minutes running time).

https://www.amstat.org/publications/stats/index.cfm?fuseaction=paperhelicopter

by Erik Barry Erhardt

*S*uppose a group of your friends are having a contest to design a paper helicopter that remains aloft the longest when dropped from a certain height. To be fair, everyone starts with the helicopter pattern given in Figure 1, which is an easy pattern to modify and replicate. Armed with knowledge about response surface methodology and a desire to strive for excellence, you could have an advantage. Let us go through the steps together, and I will show you how my classmate, Hantao Mai, and I designed “a better paper helicopter” to become two-time paper helicopter champions at Worcester Polytechnic Institute. After you see what we did, you can try to do even better.

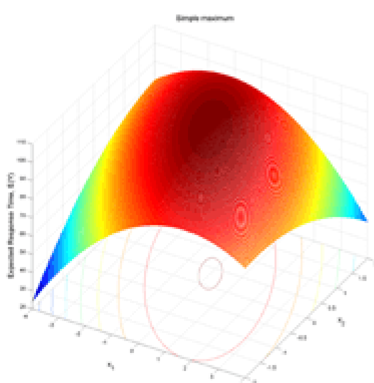


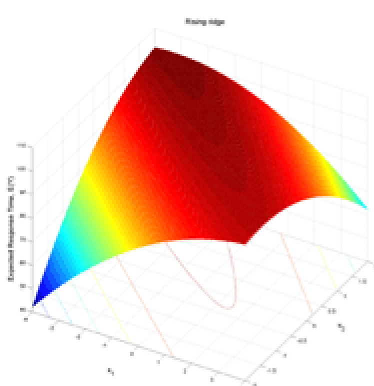
[+] **Figure 1.** Initial helicopter pattern. Cut along the solid lines and fold along the dotted lines. The foot fold, ***D***, paper weight, ***G***, and fold direction, H, were not included as part of the initial pattern, but were added after brainstorming about potential factors that might influence flight time.

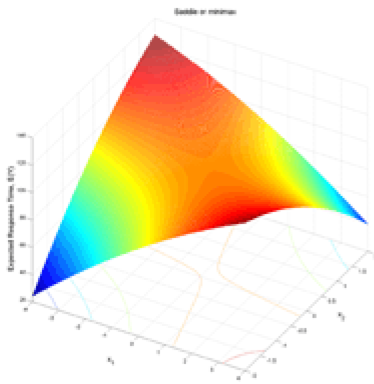
Response Surface Methodology

Response surface methodology (RSM) is a collection of statistical and mathematical techniques to explore efficiently the performance of a system to find ways to improve it. A good reference on RSM is *Response Surface Methodology* by Raymond Myers and Douglas Montgomery. A response surface can be envisioned as a curved surface representing how the system's output performance (a dependent variable) is affected by specified input factors (independent variables). Examples of response surfaces in three-dimensional space are shown in Figure 2.

George Box is the statistician credited with first proposing the ideas behind RSM more than 50 years ago. The major tools in RSM are design of experiments, multiple regression, and optimization. (These three phrases are the outline for the rest of the article. That is why we highlighted them. They need to stand out in some way so the reader can use them as sign posts while reading the article. They provide structure.)







[+] **Figure 2.** Examples of response surfaces in three dimensions: a maximum, a rising ridge, and a "saddle" with a maximum in one direction and a minimum in a transecting direction.

In the paper helicopter project, our goal was to use these tools to find the combination of design factors that would maximize flight time. We designed a set of experiments that allowed us to discover enough about the shape of the response surface to design the winning paper helicopter—twice!

Design of Experiments

Brainstorming

An important question to start with is, “What factors should we test?” A brainstorming session can reveal many factors that might influence the response. We brainstormed about helicopter design using our knowledge about the aerodynamics of flight, and tried to think of everything about the design of a paper helicopter that we could control. Then, we made a few helicopters, dropped them, and made some modifications—a fun way to start a project.

Factorial Designs

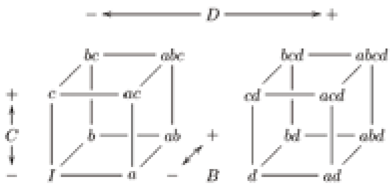
In designing experiments, the workhorses of experimental designs are factorial designs. These are efficient designs for assessing the effects of several factors on a response. A 2^k design is a factorial design for k factors, each tested at two levels, coded as +1 for “high” and –1 for “low.” The low and high levels are selected to span the range of possible values for each factor. A graphical representation of a 2^4 design is shown in Figure 3. The cube shows the orthogonal relationship of the experimental points.

Center points are experimental runs with all factors at level 0 midway between –1 and +1. Adding center points to the experiment is a logical way of checking curvature and obtaining an estimate of the error variance.

Pareto Principle and Factor Screening

The Pareto Principle says there are a “vital few” important factors and a “trivial many” less important factors.

A screening experiment can distinguish the trivial variables from the vital few important factors influencing the response. The dimensionality of the experimental region often is greatly reduced, exponentially decreasing the number of experimental runs required.



[+] Figure 3. A 2^4 factorial design consists of four factors, each taking two levels (+1, -1), represented here with a letter for +1 and without a letter for 1. For example, point *abd* in the figure corresponds to factors ***A***, ***B***, and ***D*** set to high level +1 and ***C*** to low level -1.

Fractional Factorial Designs

A full factorial design contains all possible combinations of the k factors at the tested levels. When the number of factors is large, a 2^k design will have a large number of runs. For example, $k = 8$ requires $2^8 = 256$ runs. Fractional factorial designs can be used for screening to find the factors that contribute most to the response, if a full factorial design requires a prohibitive number of runs. Fractional factorial designs are designated as 2^{k-p}, with $p = 1$ for a design with one-half the full factorial runs, $p = 2$ for one-fourth the full factorial runs, etc.

The price paid for the efficiency of a 2^{k-p} design is that ambiguity is introduced through confounding of the effects. In a useful screening design, the main effects will not be confounded with each other, although there can be confounding with the interaction effects. Because the main effects are not confounded with each other, their relative importance can be distinguished among each other, but not from certain interactions.

		Coded Values	
Factors	-1	0	+1
A = Rotor Length	5.5	8.5	11.5
B = Rotor Width	3.0	4.0	5.0
C = Body Length	1.5	3.5	5.5
D = Foot Length	0.0	1.25	2.5
E = Fold Length	5.0	8.0	11.0
F = Fold Width	1.5	2.0	2.5
G = Paper Wght.	light	(none)	heavy

H = Fold Dir.	against	(none)	with

Table 1. Factors potentially influencing paper helicopter flight time and the factor levels for the screening experiment in actual units for each coded value. Length and width were measured in centimeters.

Our Screening Experiment

The goal during a screening experiment is to identify those factors with the most influence on the response, and then eliminate the remaining factors from further analysis. To determine the vital few factors that contribute to flight time, we performed a screening experiment using a 2^{8-4} fractional factorial design with the initial helicopter pattern as our center point.

We selected eight factors possibly influencing the flight time of the helicopter: the original five factors plus three that we added (i.e., Dfoot length, G paper weight, H and fold direction). Paper weight levels were phonebook white pages paper light (–1) and standard copy paper heavy (+1). Fold direction levels indicated the fold direction was against (opposite) or with the direction of rotation. Table 1 shows the coded and uncoded values for the eight factors we investigated.

We produced 16 helicopters at the design points listed in Table 2. Then, we dropped them in random order from 20 feet and recorded the flight time of each. All flights were conducted indoors in still air.

We used SAS macro EFFECTS from *Applied Statistics for Engineers and Scientists* by J. D. Petruccelli, B. Nandram, and M. Chen to compute the factor effect estimates, as shown in Table 3.

Factors and Coded Levels

Run	Order	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	Flight Time
1	12	-1	-1	-1	-1	-1	-1	-1	-1	11.80
2	7	-1	-1	-1	1	1	1	1	-1	8.29
3	11	-1	-1	1	-1	1	1	-1	1	9.00
4	15	-1	-1	1	1	-1	-1	1	1	7.21
5	1	-1	1	-1	-1	1	-1	1	1	6.65
6	4	-1	1	-1	1	-1	1	-1	1	10.26
7	16	-1	1	1	-1	-1	1	1	-1	7.98
8	8	-1	1	1	1	1	-1	-1	-1	8.06
9	3	1	-1	-1	-1	-1	1	1	1	9.20

10	10	1	-1	-1	1	1	-1	-1	1	19.35
11	9	1	-1	1	-1	1	-1	1	-1	12.08
12	5	1	-1	1	1	-1	1	-1	-1	20.50
13	6	1	1	-1	-1	1	1	-1	-1	13.58
14	13	1	1	-1	1	-1	-1	1	-1	7.47
15	14	1	1	1	-1	-1	-1	-1	1	9.79
16	2	1	1	1	1	1	1	1	1	9.20

Table 2: Screening experiment with coded factor levels and center points

With only a single measurement for each combination of factor levels in the screening design, we could not estimate the error, as all the degrees-of-freedom would be used for estimating the factor effects. A Pareto Chart of the absolute value of the effects (see Figure 4) shows that the main effects ***A***, ***B***, and ***G*** should be considered the vital few, as they are more than twice as large as the next largest main effect (i.e., ***D***).

We were left with three main effects, but, by inspection, we could see that ***G*** = –1 (the light paper) always gave a higher response. Also, we fixed ***H*** = –1 (fold against), as we observed that this setting gave more stable helicopters. Further testing suggested favorable levels of body length ***C*** = 2, foot length ***D*** = 2 fold length ***E*** = 6, and fold width ***F*** = 2 in their uncoded units of centimeters. So, we had only two factors to consider (***A***, rotor length, and ***B***, rotor width) to start our model-building process.

Factor	Estimate	
A	3.9900	*
B	-3.0550	*
C	-0.3475	
D	1.2825	
E	0.2500	
F	0.7000	
G	-4.2825	*
H	-1.1375	
AB	-2.2175	
AC	0.8400	
AD	1.6850	
BC	-0.3850	
BD	-2.0350	
CD	0.2475	
ABCD	1.5625	

Table 3. Factor effects estimate with SAS macro EFFECTS. Factors *A*, *B*, and *G* are indicated as the vital few.

Multiple Regression

Regression Modeling

In building a regression model, we assume the response of flight time, *y*, depends upon *k* controllable factors, such as rotor length and width, $\xi_1,\ldots,\xi_k$, in $y = f(\xi_1,\ldots,\xi_k) + \varepsilon$, where *f* is an unknown function and ε is random error with expected value $E(\varepsilon) = 0$ and common variance $Var(\varepsilon) = \sigma^2$. The response surface is the expected flight time for varying factor values defined by $E(y) = f(\xi_1,\ldots,\xi_k)$. For convenience, the uncoded variables, $\xi_1,\ldots,\xi_k$, in their original units, are transformed to the coded variables, $x_1,\ldots,x_k$ with zero mean $\mu_i = 0$. The response surface is in terms of the coded variables $E(y) = f(x_1,\ldots,x_k)$.

Because we do not know the form of *f*, we need to approximate it. For this purpose, second-order polynomials are useful. A second-order polynomial has the form:

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_{ii}^2$$

The first-order terms describe a plane, while the cross-products terms (also called interaction terms) and the second-order terms describe curvature.

With only two factors, it is practical to use a 2^2 replicated design. New measurements were taken for all helicopters, as shown in Table 4.

Our Initial Model

We applied muliple regression to the results in Table 4 to determine our initial model to estimate the response surface as:

$$\begin{aligned}\hat{y} &= b_0 + b_1 x_1 + b_2 x_2 \\ &= 11.1163 + 1.2881 x_1 - 1.5081 x_2\end{aligned}$$

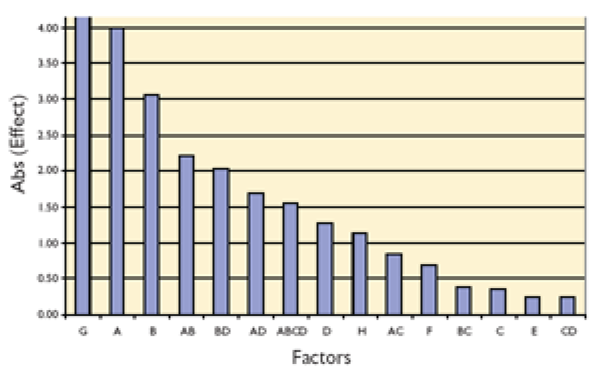


Figure 4. Pareto Chart of the absolute value of the effects, in order from largest to smallest. The paper weight, rotor length, and rotor width stand out as the important factors.

Optimization

Steepest Ascent

What is the shortest path to the top of a hill? Always take the steepest possible step. The goal is to move the process variables, *x*, to a new region of improved response by following the path of steepest ascent. The procedure is to compute the path of the steepest ascent using the first-order model and conduct experiments along this path until a local maximum is reached. The direction and magnitude of steepest ascent is the gradient. For the response surface $E(y)=f(x_1,\ldots,x_k)$, the gradient is $\nabla f=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\ldots,\frac{\partial f}{\partial x_k}\right)$, a vector of slopes.

In particular, for the first-order model $y=b_0+b_1x_1+b_2x_2+\ldots+b_kx_k$, the resulting estimated direction of steepest ascent is $\left(b_1,b_2,\ldots,b_k\right)'$.

Obs	*A*	*B*	Time
1	-1	-1	10.24
2	-1	-1	9.11
3	1	-1	16.52
4	1	-1	16.99
5	-1	1	10.20
6	-1	1	9.26
7	1	1	10.02
8	1	1	9.94
9	-1	-1	11.31
10	-1	-1	10.94
11	1	-1	12.58
12	1	-1	13.86
13	-1	1	8.20
14	-1	1	9.92
15	1	1	9.95
16	1	1	9.93

Table 4. The results of a 2^2 design with replication to determine the initial estimate of the

response surface.

Therefore, a step change proportional to the regression coefficient for each factor will be the path to follow in subsequent experiments to "climb the hill" to the optimum. For our climb, we decided each step would be 1.0cm for factor *A* (rotor length) and correspondingly -0.39 cm for factor *B* (rotor width). Five steps along the path of steepest ascent, starting at the base, are shown in Table 5. An approximate maximum was found at Step 3.

Central Composite Designs

The basic experimental designs to estimate second-order response surface models are central composite designs. A two-dimensional central composite design is shown in Figure 5. These designs consist of a factorial design for estimating first-order and two-factor interactions, 2*k* axial (or "star") points at $(\pm a, 0, \ldots, 0), (0, \pm a, 0, \ldots, 0), \ldots, (0, \ldots, 0, \pm a)$ for estimating the second-order terms, and replicated center points to estimate the second-order terms and to provide an estimate of error. The value of *a* is set at $\sqrt{k}$ for rotability, so the accuracy of prediction with a quadratic equation will not depend on direction.

Finding an Optimum

A near-optimum location was found and a second-order model was used to describe the curvature near the optimum to determine optimal conditions.

Near the optimum, the response surface is curved and can be approximated by a second-order model. A maximum or minimum will occur at a stationary point, where the partial derivatives are all 0.

$$\frac{\partial \hat{y}}{\partial x_i} = 0, \text{ for } i = 1, \ldots, k.$$

Determining the shape of the surface is easy if *k* is 1 or 2 by plotting the surface or looking at contour plots. If there are many *x* variables, then visualizing the surface is difficult, but information about curvature can be understood analytically. Canonical analysis takes a square matrix of the β_{ij} coefficients and calculates eigenvalues $\lambda_1, \ldots, \lambda_k$ and eigenvectors. The eigenvalues tell us the direction and magnitude of curvature, with the sign indicating downward (- = concave) or upward (+ = convex) curvature and magnitude indicating fairly flat (near 0) to steeply peaked (far from 0). Wanting a maximum in our experiment, we hoped all the λs would be negative.

Our Final Model

Using regression analysis, the estimated second-order model is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 + b_4 x_1^2 + b_5 x_2^2$$
$$= 16.713 - 0.702 x_1 + 0.735 x_2 - 0.510 x_1 x_2 - 1.311 x_1^2 - 1.554 x_2^2.$$

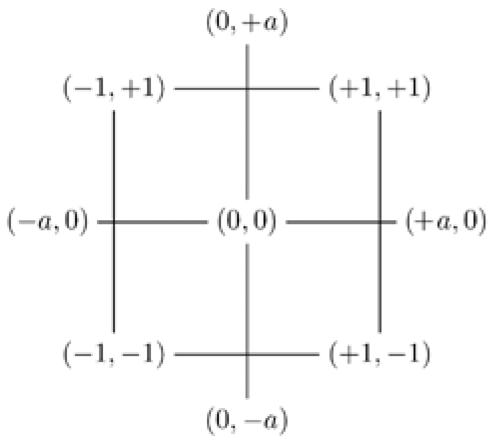


Figure 5. A two-dimensional central composite design.

	Factors			
Step	***A***	***B***	***F***	**Time**
1	8.50	4.00	2.0	12.99
2	9.50	3.61	2.0	15.22
3	10.50	3.22	2.0	16.34
4	11.50	2.83	1.5	18.78
5	13.50	2.05	1.2	7.24

Table 5. Steepest ascent coordinates and results in uncoded units (centimeters). Time measured in seconds. Although the fold width (***F***) was not a factor, it is shown in the table because it had to vary with rotor width (***B***). ***F*** had to decrease as ***B*** decreased.

Factor	$-\sqrt{2}$	**-1**	**0**	**1**	$\sqrt{2}$
A	10.08	10.50	11.50	12.50	12.91
B	2.28	2.44	2.83	3.22	3.38

Table 6. Central composite design factor levels in coded and uncoded units (centimeters)

Model Validation

To validate our model, we checked the significance of the model terms, performed a lack-of-fit test, and checked our statistical assumptions.

Using t-tests for the regression coefficients, Table 8 indicates that all factors except the interaction term were significant at the 0.05 level. To test lack-of-fit, the pure error in the process can be estimated using the center points. The lack-of-fit test suggests the model adequately describes the data because the p-value =0.3737 in Table 9 is large; we do not reject the null hypothesis (H_0) of fit (i.e., no lack of fit).

We checked that the statistical requirements of independent and identically distributed residuals were satisfied. The residuals plot indicated the residuals were random with constant variance. Additionally, we tested the residuals for normality. The test had a large p-value ($p = 0.761$), indicating the normality assumption should not be rejected. Therefore, our second-order model appeared adequate.

We used SAS's RSREG (Response Surface Regression) procedure to predict the stationary point at (a= -0.32, b=0.29) with the predicted response $\hat{y}$=16.9. The eigenvalues (λ_a, λ_b)= (-1.15, -1.71) indicated a maximum—yes!

Run	Order	*A*	*B*	Time
1	7	-1	-1	13.65
2	3	1	-1	13.74
3	11	-1	1	15.48
4	5	1	1	13.53
5	9	0	0	17.38
6	2	0	0	16.35
7	1	0	0	16.41
8	10	+1.414	0	12.51
9	4	-1.414	0	15.17
10	6	0	+1.414	14.86
11	8	0	-1.414	11.85

Table 7. Experimental results from the central composits design. Time econds. The order of runs was randomized to minimize the risk of the sequence of experimentation affecting the response.

Our Optimum Design

The estimated response surface is plotted in Figure 6, with the 90% confidence region for the location of the maximum emphasized at the top. (See the reference on Page X by E. D. Castillo and S. Cahya about computing confidence regions on a response surface stationary point.) For our model, $R^2 = .92$, which indicates the model should be able to explain about 92% of the variation in the response over the range of factor values we tested.

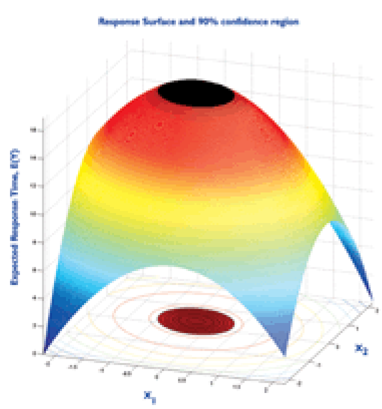
The final optimal helicopter design is shown in Figure 7.

Parameter	df	Estimate	Std Err	*T*	*p*-value
Intercept	I	16.713	0.408	40.98	0.0001
$x_1 = a$	I	-0.702	0.250	-2.81	0.0374
$x_2 = b$	I	0.735	0.250	2.94	0.0322
$x_{12} = axb$	I	-0.510	0.353	-1.44	**0.2083**
$x_1^2 = a \times a$	I	-1.311	0.297	-4.41	0.0070
$x_2^2 = b \times b$	I	-1.554	0.297	-5.23	0.0034

Table 8. Second-order response surface parameters estimates are all significant, except for the interaction term, indicating the interaction term could be dropped from the model.

Confirmatory Experiment

Last, we conducted a confirmatory experiment at the optimum location to verify the second-order prediction. The results of our confirmatory experiment consisting of six new helicopters at the predicted optimal setting are shown in Table 10. The mean response is 17.81 seconds with standard deviation 1.67 seconds, with a 95% confidence interval on the mean response of (16.1, 19.6). Our model's predicted response, $\hat{y}$=16.9 seconds, falls within this interval, thus confirming the usefulness of our response model for the factor levels we tested and under the environmental conditions in our experiments.



[+] **Figure 6.** Estimated response surface with 90% coincidence region for the coordinates of the maximum response.

There You Have It

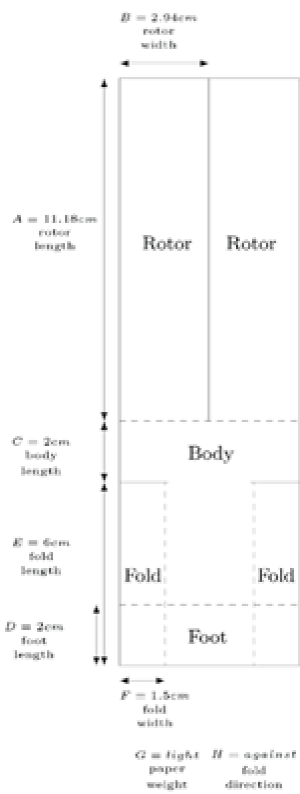
So, there you have it. That is what we did. Now, see if you can do even better. Good luck, and please let us know what you find out about using response surface methodology to design a better paper helicopter. Have fun!

Residual	df	SS	MS	*F*	*p*-value
Lack of Fit	3	1.826	0.609	1.82	**0.3737**
Pure Error	2	0.668	0.334		
Total Error	5	2.495	0.499		

Table 9. Lack-of-fit test indicates the model is adequate (i.e., it does not "lack fit").

Flight Time (sec)	15.5	16.4	19.7	19.4	18.6	17.3

Table 10. Results from the confirmatory experiment corroborate the model's predicted response.



[+] **Figure 7.** The optimal helicopter design is longer and narrower and has a shorter base than the initiallicopter design.

Design of Experiments
VOCABULARY

Effect: The mean difference in response due to changing a factor's level, such as the mean difference in flight time of paper helicopters when rotor length is changed from short to long.

Experimental Design: A systematic procedure for changing factor levels to measure and compare the responses, such as changing rotor length and rotor width on paper helicopters.

Factor: An independent variable deliberately changed in an experiment to study its effect on the response, such as the length of a paper helicopter's rotors.

Factorial Design: An experiment that measures the responses from all possible combinations of two or more factors at fixed levels, such as the four combinations of two rotor lengths and two rotor widths for paper helicopters.

Interaction: When the response due to one factor is influenced by another factor, such as the influence of rotor width on rotor length for paper helicopters.

Response: The performance of an experimental unit, such as the flight time of a paper helicopter.

Response Surface: A curved surface representing how a system's output performance is affected by specified input factors, such as how paper helicopter flight time is affected by rotor length and rotor width.