Response Surface Methods (RSM) for Peak Process Performance at the Most Robust Operating Conditions

Mark J. Anderson (mark@statease.com) and Patrick J. Whitcomb

Summary

Response surface methods (RSM) provide superb statistical tools for design and analysis of experiments aimed at process optimization. At the final stages of process development, RSM illuminates the sweet spot where high yield of in-specification products can be achieved at lowest possible cost. It produces statistically-validated predictive models and, with the aid of specialized software, response surface maps that point the way to pinnacles of process performance.

This article starts with the basics on RSM before introducing two enhancements that focus on robust operating conditions:

- Modeling the process variance as a function of the input factors
- Propagation of error (POE) transmitted from input factor variation.

Putting all these tools together, the process engineer can find the find the flats – high plateaus for maximum yield and broad valleys that minimize defects.

Response Surface Methods

Response surface methods (RSM) are powerful optimization tools in the arsenal of statistical design of experiments (DOE). Before employing RSM, process engineers should take full advantage of a far simpler tool for DOE – two-level factorials, which can be very effective for screening the vital few factors (including interactions) from the trivial many that have no significant impact. (For details, see *DOE Simplified*.[1]) Assuming the potential for further financial gain, it's best to follow up screening studies by doing an in-depth investigation of the surviving factors via RSM. Then generate a response surface map and move the process to the optimum location.

This article provides a brief introduction to RSM. For a complete primer, read *RSM Simplified*.[2]

RSM at its most elementary level – one process factor

To illustrate the elements of response surface methods, we present a very simple study that involves only one factor – cure temperature – and its effect on the ultimate shear strength of a polymer. The data are loosely derived from a problem presented in a standard textbook on RSM.[3] Table 1 shows the experimental design in a convenient layout that sorts the "X" variable (input) by level. The actual run order for experiments like this should always be randomized to counteract any time-related effects due to ambient conditions, etc.

This RSM design on one factor, generated with the aid of statistical software developed for this purpose,[4] provides seven levels of temperature, with three of them replicated – the two extremes (#'s 1-2 and 11-12) – twice each, and the center point (5-8) – four times over. This provides a total of 5 measures, or degrees of freedom, for pure error. Note that repeated measures or re-sampling from a given run will provide more stable averaged results, but only a complete re-run, for example – recharging a vessel, bringing it up to temperature and so forth, will suit for measuring overall process/sample/test variation. In general, the minimum requirement for an RSM design is that each factor be tested at three levels over a continuous scale. Additional levels provide for a statistical test on lack of fit measured against the pure error obtained via replications of one or more design points.

Table 1: One-factor RSM design on a curing process

#	A:Cure Temp (deg F)	Ultimate Shear (psi)
1	280.0	711.2
2	280.0	739.9
3	286.0	847.9
4	292.0	849.0
5	297.5	806.9
6	297.5	828.9
7	297.5	776.0
8	297.5	844.0
9	303.0	663.5
10	309.0	513.0
11	315.0	218.9
12	315.0	243.0

There is no significant lack of fit in this case as one can infer by inspection of Figure 1 – the response surface for ultimate shear strength of material cured at varying temperatures. The dotted lines represent the 95 percent confidence band on the mean prediction for any given factor level.

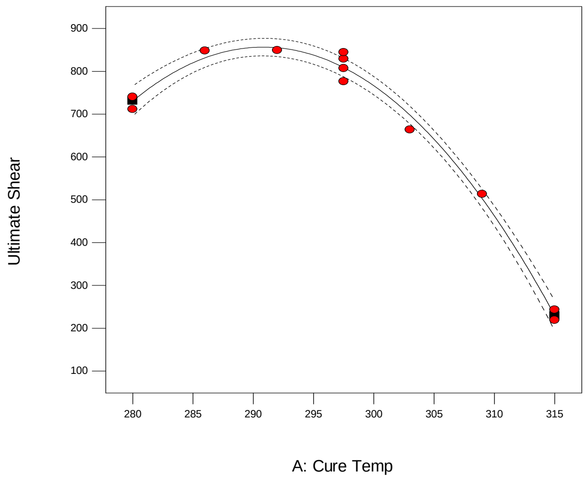


Figure 1. Response surface of ultimate shear versus cure temperature

This curve was created from the following second-order polynomial model, called a "quadratic," via least squares regression:

$$\hat{Y} = 808.77 - 250.45\ X - 328.58\ X^2$$

The experiment design (Table 1) provides sufficient input levels to fit a third-order (cubic) term – X^3. However, statistics show no significant improvement to the model's predictive capability, thus there will be no advantage to making it cubic – only complication. When modeling data, it is best to keep things as simple as possible by a statistical principle called "parsimony."

The 'hat' over the response (output) variable "Y" indicates that this is a predicted value. The coefficients are based on coded values of X (the input variable) scaled from –1 to +1 over the range tested (280 to 315 degrees F). Coded models, a standard practice for RSM, facilitate comparison of coefficients, which becomes more useful with multiple factors, as will be seen in the next example. It pays immediate dividends for predicting the ultimate shear strength at the center point value for cure temperature of 297.5 degrees F: Simply plug in zero for X, which leaves the model intercept of 808.77 as the expected outcome for ultimate shear in units of pounds per square inch (psi).

Of much greater interest for predictive purposes is the location of the maximum shear strength. For a single response-measure, the polynomial model lends itself to simple calculus. However, numerical search algorithms, such as simplex hill-climbing, work better in general and they can be done quickly with the aid of computers. In this case, the optimal cure temperature is found at 290.8 degrees F (-0.381 coded) at which the ultimate shear strength reaches its peak at 856.5 psi. To convey the uncertainty of a point estimate derived by modeling sample data from a particular experiment, it helps to provide its associated prediction interval (PI), in this case: 799 to 914 psi (p=0.05 or 95 percent). Note that the PI will always be wider than confidence interval (CI) on the mean prediction. As a practical matter for the engineer or scientist, reporting the PI will lessen any unrealistic expectations of confirming precisely the value predicted in a one-shot follow-up test.

Calculating propagation of error (POE) to find the flats

Propagation of error (POE) measures the variation transmitted from input factors to the response as a function of the shape of the surface. It facilitates finding the flats – stable spots to locate your process, for example a high plateau of yield. For example, in Figure 2 you can see how a constant 5 degree variation in cure temperature creates a very small response (ultimate shear) variation at the mid-range area "A," but when the set point is at the high end of the scale ("B"), the variation in ultimate shear becomes very large.

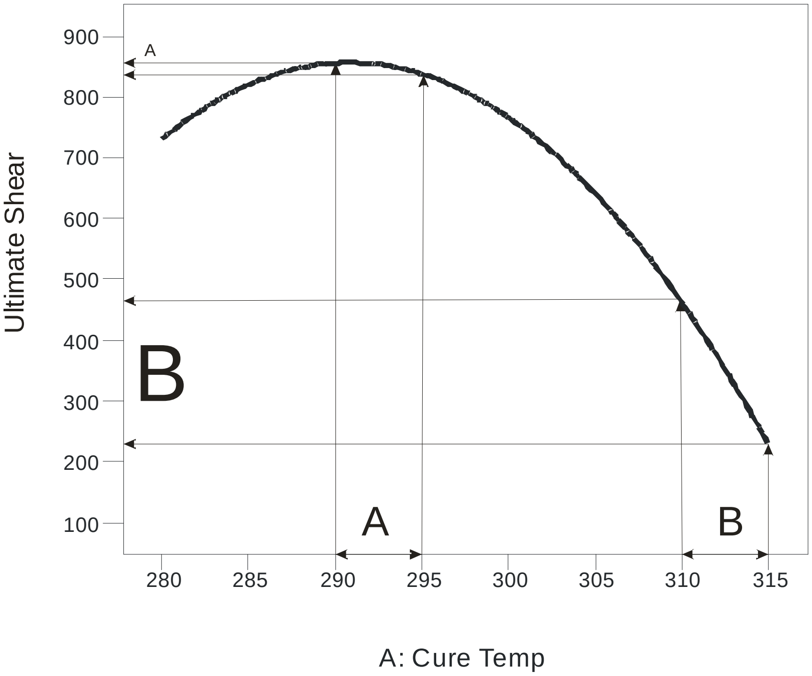


Figure 2. Variation transmitted via the response surface

The formula for POE, which involves the application of partial derivatives of the function (δf) with respect to the individual factors (X_i), is:

$$\mathrm{POE} = \sqrt{\sigma^2_{\hat{Y}}} = \sqrt{\sum_i \left(\frac{\partial f}{\partial X_i}\right)^2 \sigma^2_{X_i} + \sigma^2_{\mathrm{residual}}}$$

As a convenience to process engineers, this calculation produces an estimate of standard deviation in original units of their response measure. However, for statistical purposes, it's best to work in terms of variance – symbolized by σ^2, where the Greek letter sigma represents the standard deviation of the predicted response Y-hat, the input factors X and the unexplained residual (error); respectively in the equation.

As noted already, calculus comes into play in the POE equation with the partial derivative of the model function taken with respect to each of the individual inputs (X_i) expressed in <u>actual</u> units, which can be derived by reversing the –1/+1 coding. (Keep in mind that the standard deviations of the input factors (X) are expressed in actual units.)

These calculations become clearer by example. In this case, the actual equation for predicting ultimate shear strength is:

$\hat{Y} = -89892 + 624.06\ X - 1.0729\ X^2$

(Remember that actual units of experimental temperatures were in the hundreds of degrees, which become quite large when squared, hence the small coefficient for this term. This exemplifies why, as was discussed earlier, the coded equation serves better for interpretation.)

Assume for the curing process that temperature can be controlled only to within 2.5 degrees F of standard deviation. The residual standard deviation comes from an analysis of variance (ANOVA) done in conjunction with the fitting of the model – it is 23.72 psi.

$$\mathrm{POE} = \sqrt{(624.06 - 2 * 1.0729 * \mathrm{X}) * 2.5^2 + 23.72^2}$$

With some further number-crunching, this equation now serves to produce the picture shown in Figure 3 of the error transmitted via the surface from the variation in the model input – the temperature of this curing process. The minimum POE occurs around 290 degrees – where the shear strength peaks, which is very fortuitous!

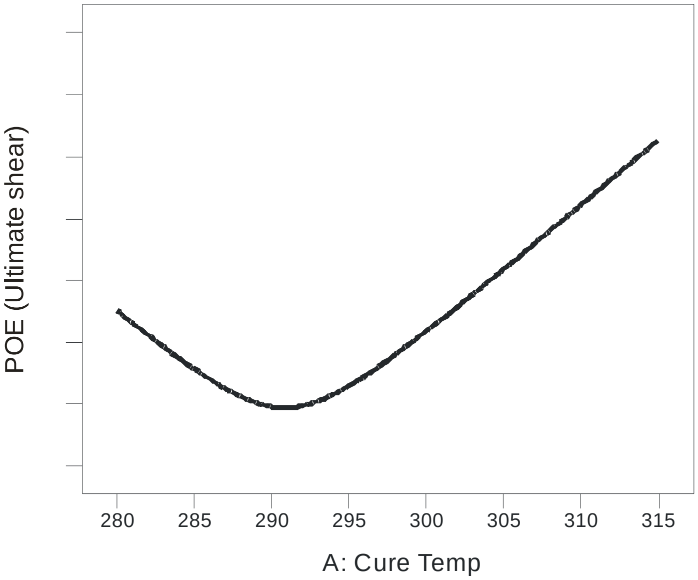


Figure 3. POE surface for cure-process

This simple example provides the basics of RSM enhance by application of POE. The next case adds another element helpful for robust process design – a second, “dual,” response: A measure of variation at each experimental setup (run).

RSM on several key factors affecting a semiconductor process

Semiconductor manufacturing engineers[5] desired a more robust result for resistivity (the response output “Y”) as a function of three key factors (the input “X”s) known to affect their single-wafer etching process:

A. Gas flow rate

B. Temperature
C. Pressure

Other variables, for example radio frequency (RF) power, could not be controlled very precisely. To measure the resulting variation over time, batches of wafers were collected over 11 different days from each of 17 runs in a central composite design (CCD). The process engineers hoped to hit a target resistivity of 350 ohm-cm with minimal variation.

The CCD is a popular template for RSM because it requires only a fraction of all the possible combinations from a full three-level factorial. Details on the CCD can be found in references 2 and 3. Figure 4 shows the CCD structure for three factors.

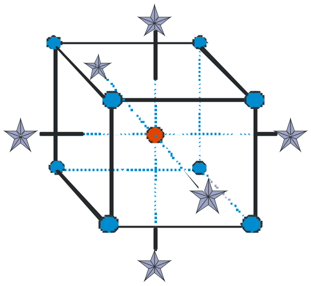


Figure 4. Central composite design on three factors

The star points project from the center point of the cubical two-level factorial. They are located a prescribed distance along the three main factor axes as shown in Table 2, which list factor levels in coded units (the experimenters kept the actual levels secret). For example, the star point projecting out to the right on Figure 4, identified by number 9 in Table 2, is located 1.68 units from the center (coded 0). To clarify what the implications of this design geometry for experiment, let’s say that the current setting of a factor is 100 and the factorial range will be plus or minus 10. Then the upper star point for the three-factor CCD would be set at 116.8 (and the lower star an equal interval below the center point at 100). These statistically-desirable distances increase as the number of factors goes up. However, the model-fit will be reliable only within the factorial ‘box.’

The CCD template calls for replication of the center point a number of times, ideally six for the best predictive properties in the middle region of experimentation. However, these experimenters ran only four center points – still not bad. The actual run order, including center points, should always be done at random. Otherwise the effects will become biased by time-related lurking variables such as the RF, thus confounding true cause-and-effect relationships.

Table 2: Design matrix for RSM on single-wafer etching process

Std	A	B	C	Resistivity	
#	Gas flow	Temp.	Press.	Mean	Std Dev
1	-1	-1	-1	263.99	107.42
2	1	-1	-1	389.94	96.12
3	-1	1	-1	205.84	66.92
4	1	1	-1	292.53	110.06
5	-1	-1	1	290.10	141.33
6	1	-1	1	302.32	147.24
7	-1	1	1	164.29	79.95
8	1	1	1	160.37	82.63
9	-1.68	0	0	211.04	57.15
10	1.68	0	0	272.08	53.42
11	0	-1.68	0	293.78	68.93
12	0	1.68	0	147.13	39.40
13	0	0	-1.68	418.55	221.96
14	0	0	1.68	273.06	193.89
15	0	0	0	268.38	64.29
16	0	0	0	236.46	81.86
17	0	0	0	250.02	73.98
18	0	0	0	315.56	99.11

Modeling both the mean and the process variance

By collecting repeated samples for each run, experimenters can model both the mean (average) and variance (or standard deviation). This enables the following tactics for process optimization:

- From the mean response find factor settings that meet the targeted response
- Use the statistics on variation to achieve operating conditions that are robust to uncontrolled (noise) variables.

Ideally, the responses measured during the course of any given run will be representative of the long-term process variability of the process. For example, the values for mean and standard deviation in Table 2 are derived from nearly a dozen daily batches over several weeks on the line. However, as few as three samples per experimental run can suffice for this dual response approach. In any case, no matter what the sample size (n), if the study conditions are not representative of true manufacturing conditions, this method may underestimate the overall variation.

To re-set the stage for this case, here is the experimenters' purpose statement "… Wafers produced on any given day (i.e., within the same batch) may be different than wafers produced on another day… Variation due to time is designed into the experimentation process by using test wafers chosen at random across several days… It may be possible to minimize...[this]…variation…by manipulating the…control variables."

An engineer from a major chip-maker told one of the authors (Mark) that variations from batch-to-batch can be a “huge” problem in semiconductor manufacturing.[6]

Least-squares regression of the Table 2 data produced these coded predictive models, for resistivity mean and standard deviation:

- Mean = $255.71 + 23.69A - 49.06B - 35.14C - 25.54AC - 16.57B^2 + 27.75C^2$ ($p<0.0001$, Adjusted R^2 0.84)
- Log_{10} Std Dev = $1.82 - 0.077B + 0.012C + 0.18C^2$ ($p<0.0001$, Adjusted R^2 0.76)

Both models are quadratic, i.e., second-order polynomials, and they are highly-significant statistically as indicated by their low “p” values and high adjusted r-squared values. The standard deviation has been transformed via a logarithm, which is standard practice for statistical reasons. To keep them simple, these models were reduced by backward regression at p of 0.10. Keep in mind that these predictive models are strictly empirical – constructed only to provide an adequate approximation of the true response surface.

> *All models are wrong, but some are useful.* - George Box (inventor of RSM)

Notice in the model for the mean that it includes squared terms for B and C, but not factor A. Thus, one can infer that the response surface will be less ‘curvy’ along the A dimension. The “perturbation” plot shown in Figure 5 illustrates this by the straight line for A. This plot originates from the center point of the experimental region and from there it measures response in each of the three dimensional axes.

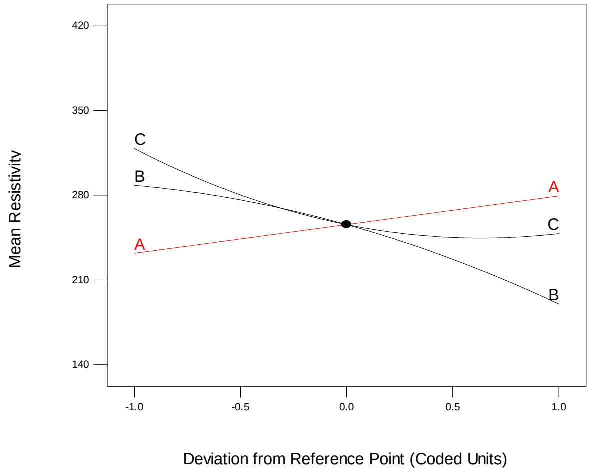


Figure 5. Perturbation plot

Figure 6 shows the contour plot for factors B and C with A set at its +1 (high factorial) level. (Recall that it’s best to stay within the ‘box’ of factorial settings in the CCD – do not extrapolate to the axial levels – 1.68 coded units in this case.)

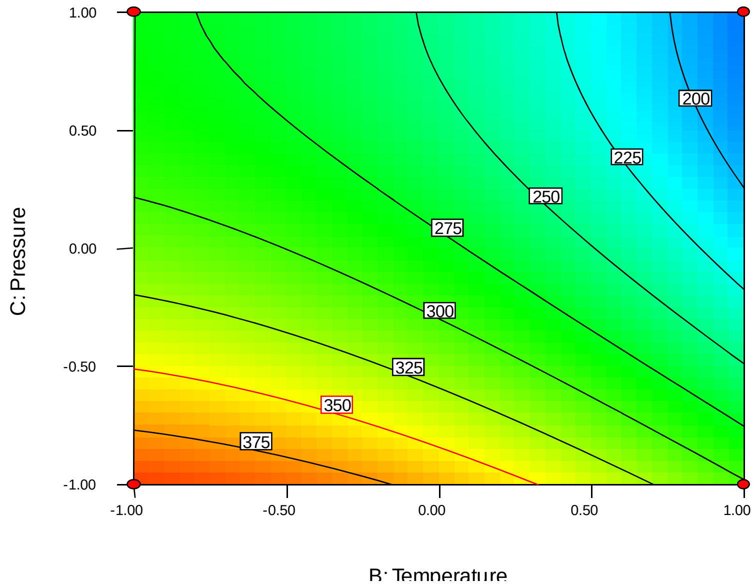


Figure 6. Contour plot of temperature vs pressure with gas flow (A) at +1 level

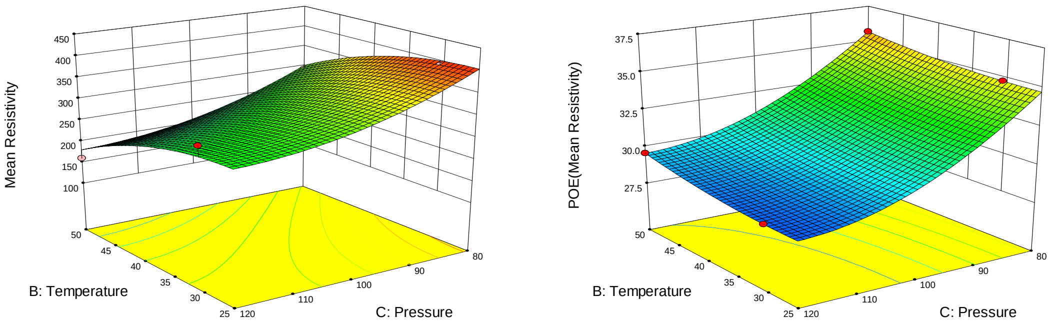
The contour for the targeted resistance of 350 cuts through a region where pressure is relatively low, but the range of possible temperatures is fairly broad. Before choosing a specific setup, the POE can be taken into account to minimize manufacturing variation caused by variability in the control-factor settings. However, at this stage the actual factor levels, –1 to +1, and their standard deviations (in parentheses) must be detailed. For illustrative purposes, assume these are:

A. 30 to 40 (1.0) sccm *[26.591 to 43.409]
B. 30 to 50 (1.0) deg C *[23.1821 to 56.8179]
C. 80 to 120 (3.0) mTorr *(66.3641 to 133.636)

[The ranges shown in asterisked brackets represent the axial star points that protrude outside the factorial box of the central composite design.] These factor settings produce the following actual predictive equation (rounded):

$$\text{Mean Resistivity} = -3.71 + 30.3\,A + 8.35\,B - 6.69\,C - 0.255\,AC - 0.166\,B^2 + 0.0694\,C^2$$

Specialized DOE software[4] performed the necessary calculations to produce the POE surface displayed by Figure 7b. For comparison's sake, the resistivity mean model graph is shown in Figure 7a. (Both graphs were generated with factor A fixed at its +1 level.) Now you can see how the POE finds the flats – the regions where process response remains most robust to factor variations.



Figures 7a & b. Surfaces of resistivity mean (left) and POE (right)

The last puzzle piece for determining where to set up the single-wafer etching process is the view of measured variation caused by batch-to-batch differences (Figure 8).

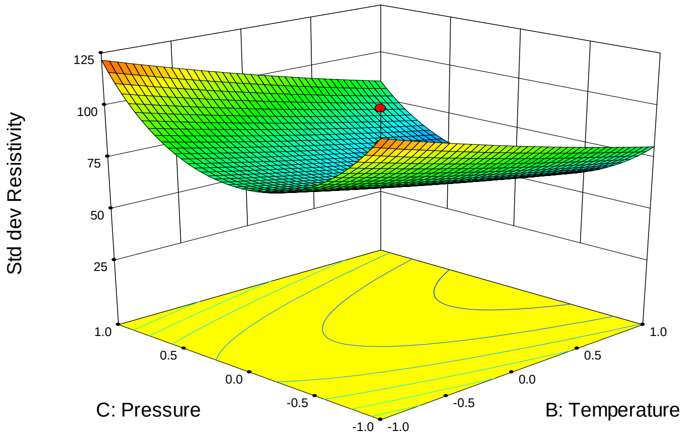


Figure 8. Response surface of resistivity standard deviation

The least variation occurs at relatively high temperature and mid-pressure. The gas flow causes little or no difference in standard deviation, which may be helpful for making the required tradeoffs – a compromise of meeting product specifications, while maintaining them from batch-to-batch in spite of control-factor variations. For example, if setting temperature and pressure for reduction of variation causes the resistivity to go off target, perhaps the gas flow can be adjusted to get the process outback back in specification.

Accomplishing the most desirable tradeoff of performance and robustness

To determine the most desirable combination of responses, RSM practitioners[7,8] typically establish this objective function:

$$D = \left(d_1 \times d_2 \times \ldots \times d_n\right)^{\frac{1}{n}}$$

In this equation the overall desirability, D, is computed by multiplying the individual desirabilities for each response, all of which are scaled from 0 to 1. Figure 9a shows how this is done for a targeted response such as resistance in this case. The goal of minimize, desired for POE and standard deviation, is pictured in Figure 9b.

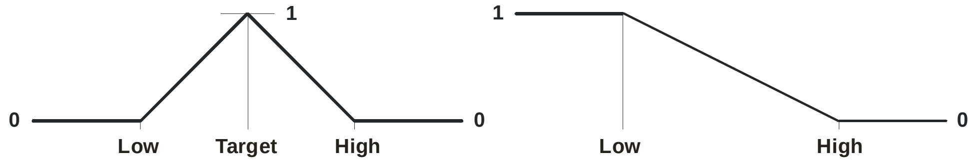


Figure 9a & b. Desirability scales for target (left) and minimization (right)

Figure 10 shows the results of a computer[4] search of the factorial region of the modeled process space for the most desirable setup based on goals of meeting the product specification of 350 (plus or minus 10), while simultaneously minimizing POE and batch-to-batch deviations.

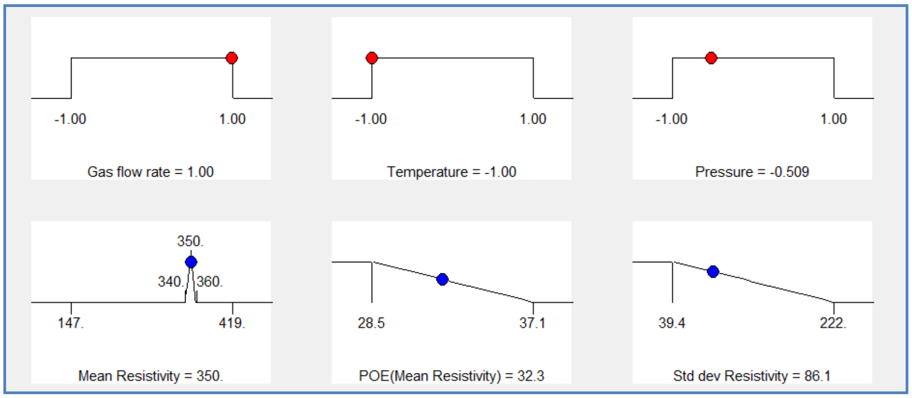


Figure 10. Most desirable process settings

The top row depicts the recommended factor settings that produce the predicted responses in the second row. (Notice, for example, that the resistivity hits the targeted spot for maximum desirability.) Figure 11 presents a view of the desirability surface.

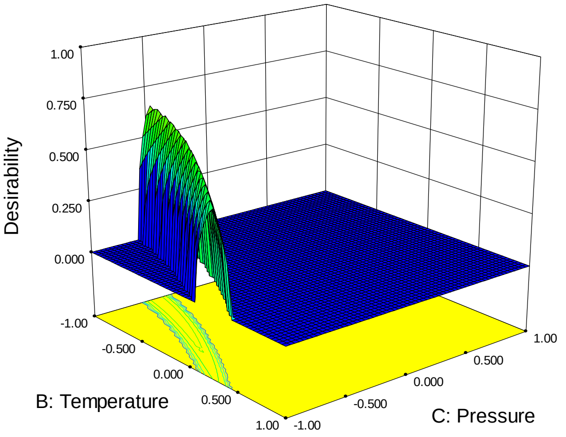


Figure 11. 3D view of desirable combinations of temperature vs pressure (gas flow set at +1 level)

The ideal setup coordinate (A,B,C) for meeting specification with least variation is (1,-1,-0.5). The authors of this original case study[5] recommended coordinate (1.18, -0.80, -0.57), which extrapolates factor A (gas flow) beyond the factorial region. We were more conservative. Nevertheless, the results do not differ appreciably. Follow-up runs are always recommended to put predictions to the test.

Conclusion

Response surface methods (RSM) provide statistically-validated predictive models that can then be manipulated for finding optimal process configurations. Variation transmitted to responses from poorly-controlled process factors can be accounted for by the mathematical technique of propagation of error (POE), which facilitates 'finding the flats' on the surfaces generated by RSM. The dual response approach to RSM captures the standard deviation of the output(s) as well as the average. It accounts for unknown sources of variation. Dual response plus POE provides a more useful model of overall response variation. The end-result of applying these statistical tools for design and analysis of experiments will be in-specification products that exhibit minimal variability – the ultimate objective of robust design.

The authors

Mark and Pat are principals of Stat-Ease, Inc. They both are chemical engineers by profession (State of Minnesota). Mark and Pat co-authored two books on that detail statistical tools for process experimentation: *DOE Simplified*[1]and *RSM Simplified*.[2] They've also collaborated on numerous articles on design of experiments (DOE), most of which can be found posted at www.statease.com.

References

1. Anderson, M. J., Whitcomb, P.J., *DOE Simplified – Practical Tools for Effective Experimentation*, *2nd Edition*, Productivity, Inc., New York NY, 2007.
2. Anderson, M. J., Whitcomb, P.J., *RSM Simplified – Optimizing Processes Using Response Surface Methods for Design of Experiments*, Productivity, Inc., New York NY, 2005.
3. Myers, R.H., Montgomery, D.C., *Response Surface Methodology*, John Wiley and Sons, Inc., New York, NY, 2002, problem 2.21.
4. Helseth, T. J., et al, Design-Expert® Version 7.1 software for Windows, Stat-Ease, Inc., Minneapolis ($995).
5. Robinson, Wulff and Montgomery, Robust Parameter Design Using Generalized Linear Mixed Models, *Journal of Quality Technology*, Vol. 38, No. 1, Jan. 2006, p. 70, Table 1.
6. Private conversation, SEMATECH Symposium on Manufacturing Effectiveness, International SEMATECH Manufacturing Initiative (ISMI), Austin, TX, October, 2007.
7. Derringer, George C., “A Balancing Act: Optimizing a Product's Properties,” *Quality Progress*, June 1994, pp51-58 (posted with publishers permission at www.statease.com/pubs/derringer.pdf).
8. “Multiple responses: The desirability approach,” section 5.5.3.2.2., NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/.