Design of Experiments (DoE): How to Handle Hard-to-Change Factors Using a Split Plot

This methodology facilitates multifactor testing. However, it comes at a price: a loss in power for detecting effects.

Mark J. Anderson, Stat-Ease, Inc.

The traditional approach to experimentation — often referred to as the "scientific method" — requires changing only one factor at a time (OFAT). Unfortunately, the relatively simplistic OFAT approach falls flat when users are faced with factor-interactions, for example, the combined impact of time and temperature on an exothermic reaction. Because interactions abound in the chemical process industries (CPI) operations, the multifactor test matrices provided by the design of experiments (DoE) approach appeal greatly to chemical engineers. However, carrying out DoE correctly requires that runs be randomized "whenever possible" [*1*] to counteract the bias that may be introduced by time-related trends, such as aging of feedstocks, decay of catalysts and the like.

But what if complete randomization proves to be so inconvenient that it becomes impossible to run a statistically designed experiment? In this case, a specialized form of design called "split plot" becomes attractive, because of its ability to effectively group hard-to-change (HTC) factors [*2, 3*]. A split plot accommodates both HTC factors (for instance, the cavities being tested for their effects on a molding process), and those factors that are easy to change (ETC; such as the pressure applied to the part being formed).

Split-plot designs originated in the field of agriculture, where experimenters applied one treatment to a large area of land, called a *whole plot*, and other treatments to smaller areas of land within the whole plot, called *subplots*. For example, Figure 1 shows two alternative experiments [*4*] that were carried out on six varieties of sugar beets (Number 1 through 6) that were sown either early (*E*) or late (*L*) in the growing season:

- A completely randomized design in one field shown on the top row versus
- The bottom row where the whole plot (single field) has been split into two subplots (in this case, sown early versus late)

E5	L1	L4	E2	E6	E3	L3	E1	L6	L5	E4	L2
E4	E1	E6	E5	E3	E2	L2	L3	L6	L5	L1	L4

Figure 1. Comparing a completely randomized experiment (top row) versus one that is divided into split plots (bottom row) [*4*]

The split-plot layout made it far sweeter (pun intended) for the sugar beet to sow the seeds because of the grouping, it being far easier to plant subplots early versus late, rather than doing it in random locations

Case in point: The use of a split plot in an industrial experiment

DoE pioneer George Box developed a clever experiment that led to the discovery of a highly corrosion-resistant coating for steel bars [*5*]. Four different coatings were tested (which is easy to do) at three different furnace temperatures (which is hard to change), each of which was run twice to provide statistical power. The design that Box pioneered (a split plot) for this experiment is shown in Table 1 (results for relative corrosion resistance—the higher the better—are shown in parentheses). Note the bars being placed at random by position.

TABLE 1. Using a Split-Plot Design to Increase the Corrosion-resistance of Steel Bars

Group	*Heat (°C) (Whole plots)*	*Positions (Subplots)*			
1	360	C2 (73)	C3 (83)	C1 (67)	C4 (89)
2	370	C1 (65)	C3 (87)	C4 (86)	C2 (91)
3	380	C3 (147)	C1 (155)	C2 (127)	C4 (212)
4	380	C4 (153)	C3 (90)	C2 (100)	C1 (108)
5	370	C4 (150)	C1 (140)	C3 (121)	C2 (142)
6	360	C1 (33)	C4 (54)	C2 (8)	C3 (46)

Observe in this experiment-design layout how Box made it even easier, in addition to grouping by temperature (i.e., “heats”), by increasing the furnace temperature run-by-run and then decreasing it gradually. This had to be done out of necessity due to the difficulties of heating and cooling a large mass of metal. The saving grace, however, is that, although shortcuts like this undermine the resulting statistics when they do not account for the restrictions in randomization, the effect estimates remain true. Thus, the final results can still be assessed on the basis of subject matter knowledge (in terms of whether they indicate important findings). Nevertheless, if at all possible, it will always be better to randomize levels in the whole plots and, furthermore, reset them (i.e., turn the dial away and then back to the same value) when they have the same value, for example, between Groups 3 and 4 in this design.

In this case as often happens, the resetting of an HTC factor (temperature) created so much noise in this process that in a randomized design it would have overwhelmed the ability to detect the effect of coating. The application of a split plot overcomes this variability by grouping the heats (i.e., oven batches), in essence, filtering out the temperature differences. Figure 2 tells the story.

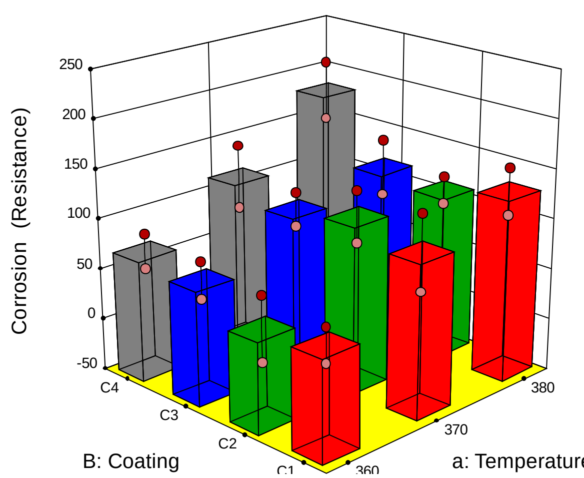

Figure 2. In this effects graph, 3D bars show the impact of temperature (a) versus coating (B) on corrosion resistance—the higher the better

The best corrosion resistance occurred for coating C4 at the highest temperature (see the tallest grey tower, located at the back corner of Figure 2). This finding — the result of the two-factor interaction *aB* between temperature (*a*) and coating (*B*), achieved significance at $p < 0.05$ (that is, a level exceeding 95% confidence). The main effect of coating (*B*) also emerged as statistically significant. If this experiment had been run in a completely randomized way, the *p*-values (to put it very simply, the probability on a 0 to 1 scale of results being caused by chance) for the coating effect and the coating-temperature interaction would have been roughly 0.4 and 0.85, respectively — that is, not sufficient to be considered statistically significant. In his work, Box concludes by suggesting that metallurgists try even higher temperatures with the C4 coating while simultaneously working at better controlling the furnace temperature. Furthermore, he urges the experimenters to work at understanding better the physiochemical mechanisms causing the corrosion of the steel. This really was the genius of George Box — his matchmaking of empirical modeling tools with his subject matter expertise.

Caveats

Split plots essentially combine two experiment designs into one. As a result, they produce both split-plot and whole-plot random errors. For example, the corrosion-resistance design discussed above introduces whole-plot error with each furnace re-set, due to potential variation that can result from, for instance, operator error in dialing in the temperature, inaccurate calibration, changes in ambient conditions and so forth [*6*]. Meanwhile, split-plot errors arise from bad measurements, variation in the distribution of heat within the furnace, differences in the thickness of the steel-bar coatings and more.

This split-error structure creates complications in computing proper *p*-values for the effects, particularly when departing from a full-factorial, balanced and replicated experiment, such as the corrosion-resistance case. If you really must use this route, be prepared for your DoE software applying specialized statistical tools that differ from standard analysis.

Furthermore, you cannot expect that doing an experiment more conveniently will not come at a cost — nothing good comes free. The price you pay for taking advantage of split plots is the loss of power to pin down some effects on those factors that are grouped, that is, not completely randomized [*7*].

For example, consider the experiment in Table 2 that tests five factors at two levels, an example taken from a workshop on DoE [*8*]. The factors are:

1. *a* =Temperature, °C
2. *B* = Catalyst, wt.%
3. *C* = Agitation, rev/min
4. D = Feed rate, L/min
5. E = Atmosphere (blanketed by nitrogen or left open to the air)

Note that the first factor is designated by the lower case letter *a* to distinguish it as being hard to change (HTC) versus the others (*B* through *E*) which are characterized as being easy to change (ETC).

The response of interest is molecular yield, for which a difference of 5% at the least is desired to be detected (this is considered the “signal”), and its normal variation is 2% (this is considered the “noise”). (The actual results are not provided as this is not relevant to the issues being discussed, which are matters of design.) Observe in Table 2 how the experiment design groups the runs by temperature (“a”) — an HTC factor. This is characteristic of a split-plot design as opposed to a standard DoE that is completely randomized. The other four factors — those that are ETC — are randomized within each group.

TABLE 2. A Split-plot Experiment Design Aimed atImproving Chemical Yield

Group	Run	*a* (Temperature, °C)	*B* (Catalyst, %)	*C* (Agititation rate, rpm)	*D* (Feed rate, L/min)	*E* (Atmosphere, nitrogen versus air)
1	**1**	10	1	120	140	Nit
1	**2**	10	2	120	140	Air
1	**3**	10	2	100	180	Air
1	**4**	10	1	100	180	Nit
2	**5**	10	1	120	180	Air
2	**6**	10	2	120	180	Nit
2	**7**	10	1	100	140	Air
2	**8**	10	2	100	140	Nit
3	**9**	15	2	120	140	Nit
3	**10**	15	1	100	180	Air
3	**11**	15	2	120	180	Air
3	**12**	15	1	100	140	Nit
4	**13**	15	2	100	180	Nit
4	**14**	15	1	120	180	Nit
4	**15**	15	1	120	140	Air
4	**16**	15	2	100	140	Air

Based on the user-specified signal-to-noise ratio of 2-to-1, and an HTC/ETC variance ratio of 1 (a standard assumption), the DoE-dedicated software [*9*] that was used to build this design computed the power results that are shown in Table 3.

TABLE 3. Comparison of Power for Detecting Main Factor Effects in Split-Plot versus Randomized Experiments

	a	*B*	*C*	*D*	*E*
Split-Plot	26%	99%	99%	99%	99%
Random	89%	89%	89%	89%	89%

Not surprisingly, the power for the main effect of *a* (the HTC factor of temperature) drops to a fraction of what it would have been in a completely randomized design — far below the generally accepted level of 80%. On the other hand, the power for the ETC factor effects goes up a bit due to being ‘protected’ from the impact of changing temperature by the grouping. Also, it turns out that the power for interactions of the HTC with ETCs (that is, the interactions between temperature and catalyst, *aB*), also comes in higher, for the same reason. Thanks to this bonus, a split-plot design such as this one is a viable alternative to a fully randomized design when a factor such as temperature cannot be easily or quickly changed without creating a big upset in the reaction.

Closing thoughts

Keep the power loss on HTC factor(s) in mind before settling for a split-plot design. Perhaps grouping HTCs for convenience may not be worth this cost — you would be better off taking the trouble to randomize the whole design. However, for many processes, running any experiment may become impossible if it requires certain factors, such as temperature, to be re-set and equilibrated for each test run; the time and expense to do this becomes prohibitive. These are situations for which a split plot can come to the rescue.

References

1. Kleppmann, Wilhelm, Design of Experiments (DoE): Optimizing Products and Processes Efficiently, *Chem. Eng.*, November 2014, 50–57.
2. Anderson, M.J. and Whitcomb, P.J., "DOE Simplified: Practical Tools for Effective Experimentation," 3rd ed., Productivity Press, Publishing, New York, N.Y., 2015, Chapter 11.
3. Anderson, M.J. and Whitcomb, P.J., "RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments," 2nd Ed., Productivity Press, 2016, Chapter 12.
4. Cox, D. R., "Planning of Experiments", John Wiley & Sons, 1958.
5. Box, G.E.P., Split Plot Experiments, *Quality Engineering*, 8(3), 1996, 515–520.
6. Jones, B., and Nachtsheim, C., Split-Plot Designs: What, Why, and How, *Journal of Quality Technology*, 41 (4), October 2009, 340–361.
7. Anderson, M.J. and Whitcomb, P.J., Employing Power to 'Right-Size' Design of Experiments, *ITEA Journal*, 35, March 2014, 40–44.
8. Experiment Design Made Easy, Stat-Ease, Inc., Minneapolis. www.statease.com/training/workshops/class-edme.html
9. Design-Expert® software, Version 10, Stat-Ease, Inc., Minneapolis, 2016. www.statease.com/software/dx10-trial.html

Author

Mark J. Anderson, PE, CQE, MBA, is a principal and general manager of Stat-Ease, Inc., 2021 East Hennepin Avenue, Suite 480, Minneapolis, MN 55082, mark@statease.com, 612-746-2032. Prior to joining the firm, he spearheaded an award-winning quality-improvement program for an international manufacturer, generating millions of dollars in profit. Anderson has a diverse array of experience in process development, quality assurance, marketing, purchasing and general management. He is also the co-author of two books (Ref. 2 and 3) and has published numerous articles on design of experiments (DOE). He is also a guest lecturer at the University of Minnesota's Chemical Engineering & Materials Science Department, and Ohio State University's Fisher College of Business. Anderson is a certified Professional Engineer (PE) in Minnesota, and a Certified Quality Engineer (CQE). He holds a B.S.Ch.E. (high distinction) from the University of Minnesota, and an MBA from the University of Minnesota, where he was inducted into Beta Gamma Sigma (the honorary academic society).