Factorial experimental designs, including fractionally replicated designs, are widely used to screen factors that might affect a quality characteristic.

After the screening step, which is part of process characterization, the next step is usually process optimization (or at least process improvement).
Response Surface Methods

Consider a response variable $Y$, such as the yield of a chemical process, that is affected by the levels of certain factors, such as reaction temperature ($x_1$) and reaction time ($x_2$).

The expected value of $Y$ can be thought of as a function of $x_1$ and $x_2$:

$$E(Y) = f(x_1, x_2).$$

Sometimes we may know enough about the chemistry and physics of the process to specify $f(\cdot, \cdot)$, but often it is largely unknown, except that it should change smoothly as $x_1$ and $x_2$ change.
Every smooth function can be approximated locally by low-order polynomials:

- **first-order** approximation:

  \[ E(Y) \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2; \]

- **second-order** approximation:

  \[ E(Y) \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1,2} x_1 x_2 + \beta_{1,1} x_1^2 + \beta_{2,2} x_2^2. \]
The “response surface method” consists of:

- Choosing initial settings of the important factors;
- Designing and carrying out an experiment in the neighborhood of the initial settings;
- Deciding on and estimating an appropriate approximation;
- Using that approximation to improve the factor settings.
Example: plasma etching process

Recall that two factors (gap, \( x_1 \), and power, \( x_4 \)) were found to be important.

The levels for gap were \( 1.0 \pm 0.2 \) cm, and for power were \( 300 \pm 25 \) W, both coded as \( \pm 1 \).

Fit the first-order model:

\[
\text{plasma} \leftarrow \text{read.csv("Data/Table-13-15.csv")}
\]
\[
\text{plasmaLm} \leftarrow \text{lm(Rate} \sim \text{A + D, plasma)}
\]
\[
\text{summary(plasmaLm)}
\]

The fitted equation, in coded variables, is

\[
\hat{y} = 776.06 - 50.81x_1 + 153.06x_4.
\]
Examine the response surface in the neighborhood of the design \( (x_1 \text{ and } x_4 \text{ between } -1 \text{ and } +1) \):

```r
aGrid <- seq(from = -1, to = 1, length = 40);
dGrid <- aGrid;
yHat <- predict(plasmaLm, expand.grid(A = aGrid, D = dGrid));
yHat <- matrix(yHat, length(aGrid), length(dGrid));
contour(aGrid, dGrid, yHat, xlab = "A", ylab = "D")
# alternatively, a perspective plot:
persp(aGrid, dGrid, yHat)
```

The predicted etch rate increases steadily as \( x_1 \) decreases and \( x_4 \) increases, and will continue to do so outside this neighborhood.
Steepest ascent

To improve the etch rate the most for a given step length, follow the path of steepest ascent.

That is, the changes in $x_1$ and $x_4$ should be proportional to their coefficients in the fitted equation, $-50.81 : 153.06$, or approximately $-1 : 3$.

Experiments were carried out by incrementing $x_4$ by $+1$ from the center point ($x_1 = x_4 = 0$), and decrementing $x_1$ by $-0.33$.

Three steps resulted in a satisfactory increase in etch rate, with $x_1 = -1$ (gap = 0.8 cm) and $x_4 = 3$ (power = 375 W).
Second-order response surface

A new experiment was carried out centered at these settings. The unreplicated $2^2$ design was supplemented by:

- Replicated center points, at $(0, 0)$;
- Unreplicated axial points, at $(\pm \alpha, 0)$ and $(0, \pm \alpha)$, with $\alpha = \sqrt{2}$.

This is a **central composite design** (CCD); the choice of $\alpha = \sqrt{2}$ makes it a **rotatable design**.

This design allows estimation of the second-order model:

```r
plasmaNew <- read.csv("Data/Table-14-01.csv")
plasmaNewLm <- lm(Rate ~ A * D + I(A^2) + I(D^2), plasmaNew);
summary(plasmaNewLm)
```
Neither $x_1^2$ nor $x_4^2$ has a significant coefficient, so a simpler reduced model was fitted, omitting them:

```r
plasmaNewLmReduced <- lm(Rate ~ A * D, plasmaNew);
summary(plasmaNewLmReduced)
```

Use the `anova()` method to compare the models:

```r
anova(plasmaNewLmReduced, plasmaNewLm)
```

The last line allows testing $H_0: \beta_{1,1} = \beta_{2,2} = 0$; the small $F$-ratio and large $P$-value mean that we do not reject $H_0$. 
Examine the response surface in the neighborhood of the design ($x_1$ and $x_4$ between $-\alpha$ and $+\alpha$):

```
aGrid <- seq(from = -sqrt(2), to = sqrt(2), length = 40);
dGrid <- aGrid;
yHat <- predict(plasmaNewLmReduced, expand.grid(A = aGrid, D = dGrid));
yHat <- matrix(yHat, length(aGrid), length(dGrid));
contour(aGrid, dGrid, yHat, xlab = "A", ylab = "D")
```
The data set contains a second response, Uniformity:

```r
plasmaNewLmUnif <- lm(Uniformity ~ A * D + I(A^2) + I(D^2), plasmaNew);
summary(plasmaNewLmUnif)
```

All coefficients are significant, so no reduced model is considered:

```r
uHat <- predict(plasmaNewLmUnif, expand.grid(A = aGrid, D = dGrid));
uHat <- matrix(uHat, length(aGrid), length(dGrid));
contour(aGrid, dGrid, uHat, xlab = "A", ylab = "D")
```
The ultimate goal was to find settings where the etch rate is between 1100 and 1150, and the uniformity is at most 80:

```r
contour(aGrid, dGrid, yHat, levels = c(1100, 1150));
contour(aGrid, dGrid, uHat, levels = 80, col = "blue", add = TRUE)
```

In case it’s not clear where the conditions are met:

```r
image(aGrid, dGrid,
     ifelse (yHat >= 1100 & yHat <= 1150 & uHat < 80, TRUE, NA),
     col = hsv(0.33, alpha = 0.5), add = TRUE)
```

Settings within the acceptable region would be chosen based on other criteria, such as equipment reliability, etc.