Quality Engineering

ISSN: 0898-2112 (Print) 1532-4222 (Online) Journal homepage: http://www.tandfonline.com/loi/lqen20

Quality quandaries: Understanding aspects influencing different types of multiple response optimization

Christine M. Anderson-Cook, Yongtao Cao & Lu Lu

To cite this article: Christine M. Anderson-Cook, Yongtao Cao & Lu Lu (2016): Quality quandaries: Understanding aspects influencing different types of multiple response optimization, Quality Engineering, DOI: 10.1080/08982112.2016.1226339

To link to this article: http://dx.doi.org/10.1080/08982112.2016.1226339

Accepted author version posted online: 26 Aug 2016.
Published online: 26 Aug 2016.

QUALITY ENGINEERING
2016, VOL. 0, NO. 0, 1–13
http://dx.doi.org/10.1080/08982112.2016.1226339

Taylor & Francis
Taylor & Francis Group

Quality quandaries: Understanding aspects influencing different types of multiple response optimization

Christine M. Anderson-Cook[a], Yongtao Cao[b], and Lu Lu[c]

[a]Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, New Mexico; [b]Indiana University of Pennsylvania, Indiana, Pennsylvania; [c]University of South Florida, Tampa, Florida

KEY POINTS

Optimizing with several responses can benefit from an objective approach of eliminating non-contenders, understanding trade-offs between competing responses, and then identifying a final choice that matches optimization priorities. To offer insights that can help guide thoughtful decisions, we explore and summarize different patterns of solution sets and their trade-offs for different types of optimization with responses that are to be maximized and/or to achieve a target.

Introduction

Many optimization scenarios involve a common objective of selecting an optimal location in the input space (sometimes called the design space or X-space) that simultaneously balances good performance for several competing responses characterized using response surface methodology. Such problems are called multiple response optimizations (MROs) and require practitioners to initially run a designed experiment to allow modeling of each response to characterize the relationship between the response and the inputs which were manipulated during the experiment. Once this characterization is completed, the overall goal is to identify where input levels can be set to achieve desired performance for the responses.

A single-response optimization typically involves either maximizing (or minimizing, which can be transformed into a maximization of the negative of the response) or achieving a target value. Even though a targeting problem can also be turned into maximizing the negative absolute distance between the response and the target value, Anderson-Cook, Cao, and Lu (2016) shows that these two types of optimization often have quite different behaviors. In this article, we investigate how combinations of different types of single—objective optimization can suggest promising choices for MRO problems. Specifically, we consider three scenarios with two responses: maximizing versus maximizing (MM), maximizing vs. targeting (MT) and targeting vs. targeting (TT). Since it is rarely the case that all responses can be simultaneously optimized with the same input settings, trade-off and compromises are needed. How the practitioner values good performance for each response contributes a subjective component to the decision-making that is not present in single response optimization. An essential part of the problem is to explore different alternatives and understand what options are available and how they can satisfy different priorities.

A Pareto front (Lu, Anderson-Cook, and Robinson 2011) is an excellent way of eliminating non-contending solutions and identifying rational options that represent best choices for some combination of priorities of the different responses. The Pareto set (PS) is the set of all solutions (here, locations in the input space) that are not dominated by other solutions, and the Pareto front (PF) is the set of response values associated with the PS. Finding a solution with a balance of good performance for the different responses can be achieved by looking at a PF constructed from a grid of locations in the input space (Chapman, Lu, and Anderson-Cook 2014a, 2014b). In this article, we explore some of the different patterns and sets of

CONTACT Christine M. Anderson-Cook candcook@lanl.gov Statistical Sciences Group, Los Alamos National Laboratory, P.O. Box 1663, MS F660, Los Alamos, NM 87545.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lqen.

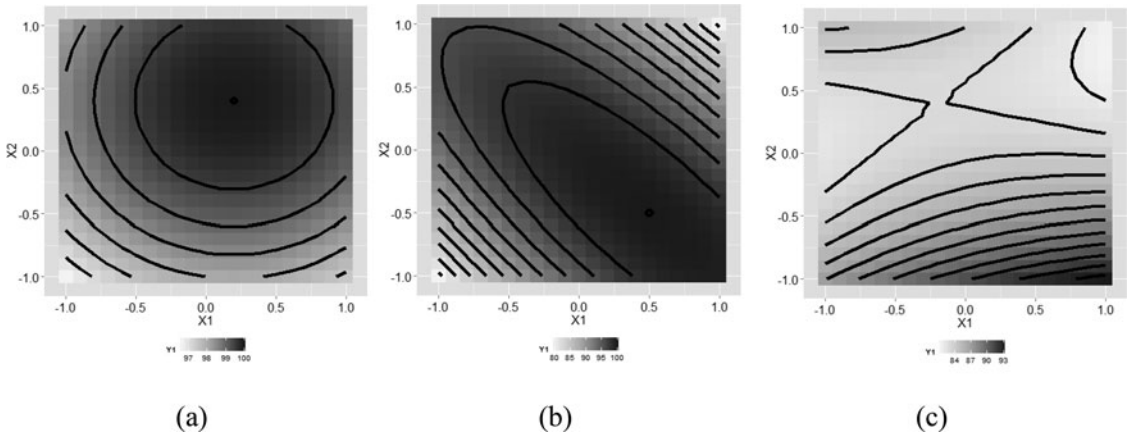

Figure 1. Different response surfaces: (a) stationary point at $(X_1, X_2) = (0.2, 0.4)$ as maximum with equal-sized eigenvalues; (b) stationary point at $(0.5, -0.5)$ as maximum with unequal eigenvalues with larger magnitude eigenvalue (faster rate of change) directed at 60° counter clockwise from horizontal; and (c) saddlepoint stationary point at $(-0.2, 0.4)$ with positive eigenvalue at 105° counter clockwise from horizontal. The faster rate of change for the positive eigenvalue means that it has larger absolute value.

solutions for the three types of multiple response optimizations (MM, MT, TT). As we will see, the sets of solutions are dependent on the nature of each response as well as the trade-off between the responses.

Consider a simple scenario, where two responses (Y_1 and Y_2) are both characterized by second order response surface models based on the two input variables (X_1 and X_2) of the form

$$Y_i = \beta_{0i} + \beta_{1i}X_1 + \beta_{2i}X_2 + \beta_{11i}X_1^2 + \beta_{22i}X_2^2 + \beta_{12i}X_1X_2 + \varepsilon, \; i = 1, 2.$$

The estimated equation for each response can be expressed by replacing the model parameters with their maximum likelihood or least squares estimates to give

$$\hat{y}_i = b_{0i} + b_{1i}X_1 + b_{2i}X_2 + b_{11i}X_1^2 + b_{22i}X_2^2 + b_{12i}X_1X_2.$$

To make the characteristics of the surface more apparent, we convert this functional form to a canonical version (see Myers, Montgomery, and Anderson-Cook 2016, 277–282), which can be expressed as

$$\hat{y}_i = \hat{y}_{si} + \lambda_{1i}w_{1i}^2 + \lambda_{2i}w_{2i}^2,$$

where $\hat{y}_i$ is the estimated response for response i, $\hat{y}_{si}$ is its estimated response at the stationary point, the λ_is are the eigenvalues of a symmetric matrix of the estimated coefficients for the second-order model and w_is are the corresponding eigenvectors centered at the location of the stationary point. The advantage of using the canonical form is that the nature and characteristics of the surface can be readily identified. For example, if both eigenvalues are negative, then the second order surface is a "mountain." If both eigenvalues are positive, the surface is a "valley." Having one positive and one negative eigenvalue creates a "saddlepoint" surface. The relative size of the eigenvalues or their ratio, $\frac{\lambda_1}{\lambda_2}$, describes the relative rates of change of the surfaces in different directions. If this ratio is positive but smaller than one, then the surface has either a global maximum or minimum with larger elongation of the surface in the w_1 canonical direction (corresponding to slower change in the surface), while a ratio larger than one implies more elongation in the w_2 direction. If the ratio is equal to one, then the contour plot has circles (such as in Figure 1a) rather than ellipses. If the λs have the same sign but one of them is close to zero, then the response surface is a rising (positive ratio) or a falling (negative ratio) ridge. Figure 1 illustrates several of the different possible surfaces.

Many different combinations of surface types, locations of the stationary point, and ratios exist, so we initially focus on one class of response surfaces to generalize patterns in multiple response optimization. Later we address some other characteristics for other scenarios. In addition, three R Shiny apps are available at https://ycao.shinyapps.io/CAL_xx/ (where xx = MM, MT, or TT for the three scenarios), to experiment with other scenarios, beyond what is possible to discuss in this article. The Appendix provides some information about the options available in the apps. We begin by considering a case where both surfaces are "mountains" (λ_1 and $\lambda_2 < 0$), and the maximum of the surface (the stationary point) is located inside the design region. This is a natural starting point for our discussions, since if the goal is maximizing at least one response for the MM and MT scenarios, we anticipate that the experimenters have identified the region that includes

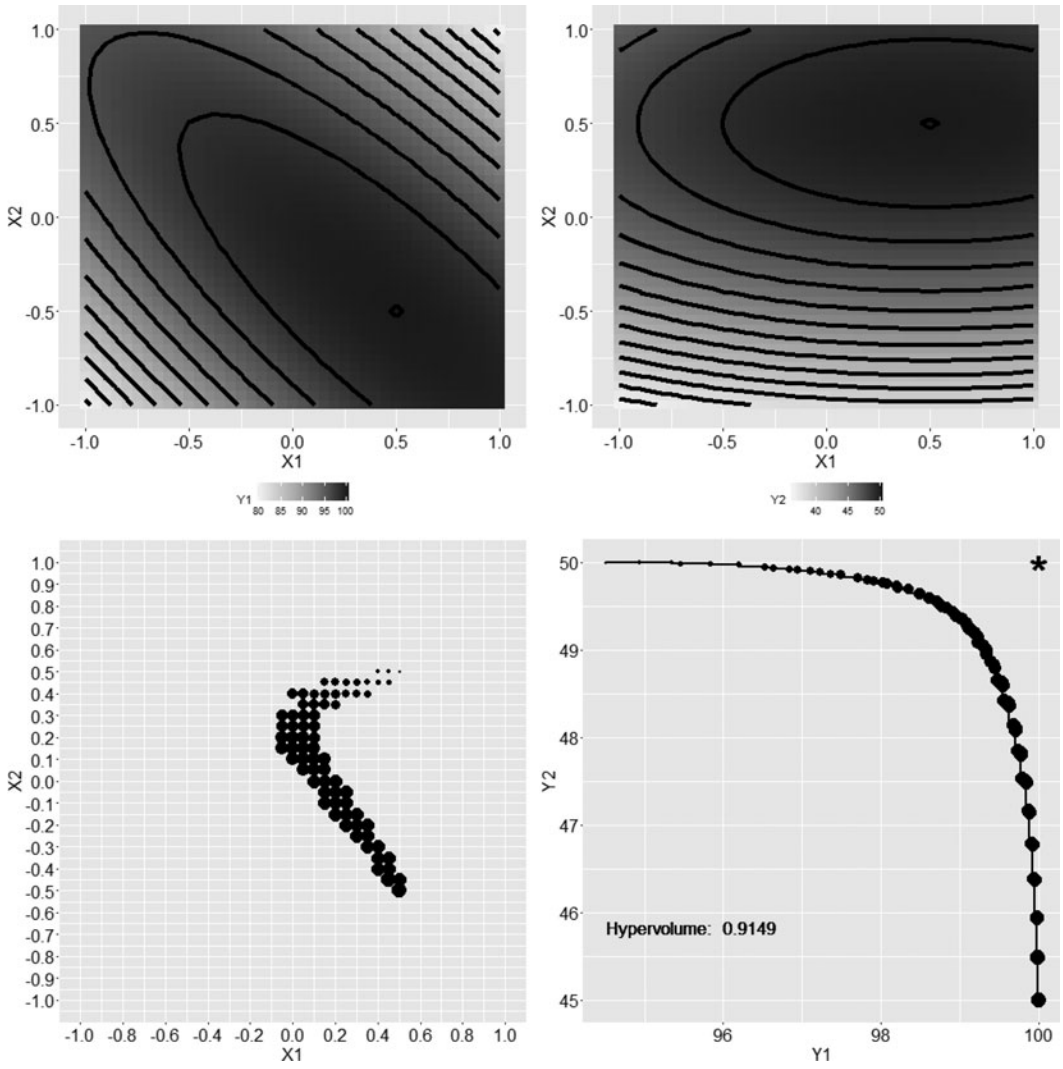

Figure 2. Summary of response surfaces and their associated Pareto set and Pareto front values for a MM optimization problem. The top row shows a contour plot of each response surface for Y_1 and Y_2. The bottom left plot shows locations in the input space of the Pareto set, while the bottom right plot shows the Pareto front. The size of dots in the bottom sub-figures is proportional to the value of Y_1 and inversely proportional to Y_2. In the bottom right figure, the "*" indicates the Utopia point, with the most desirable response values.

the location of the maximum using expert knowledge or using sequential experimentation techniques as with the path of steepest ascent (Myers, Montgomery, and Anderson-Cook 2016, Chapter 5).

Maximizing-maximizing (MM) optimization

MM is a common focus in the response surface literature, and thus a good starting point for our exploration. With this scenario, we explore the impact of changing the value of the responses at their stationary points ($\hat{y}_{si}$), the location in the input space of the stationary points, the orientation of response surface contour ellipses, and the ratios of the λ_is. First, we begin with an initial scenario to gain some understanding about finding the Pareto front of solutions and how these results can be displayed.

Figure 2 shows summary contour plots of the estimated responses, the Pareto set identified as well as the Pareto front. In this scenario, Y_1 is estimated to achieve a maximum of 100 at $(X_1, X_2) = (0.5, -0.5)$, while Y_2 achieves its maximum of 50 at $(X_1, X_2) = (0.5, 0.5)$. The values of Y_1 drop most slowly as we move in a northwest or southeast direction away from the maximum, while Y_2 changes most slowly moving directly east or west. Since we cannot simultaneously achieve a setting in the input space that leads to $Y_1 = 100$ and $Y_2 = 50$, we need to identify solutions that do as well as possible for both responses simultaneously. The point $(Y_1, Y_2) = (100, 50)$ is called the *Utopia point* in the Pareto front literature and represents the (usually unattainable) ideal against which other solutions can be compared. In the bottom right subfigure of Figure 2, it is denoted with an asterisk (*).

In this case we restrict ourselves to a gridded set of points in the input space, where the resolution of the grid is based on what is practically possible when we implement the final settings. The bottom left subplot of Figure 2 shows the solutions in the PS, which might be optimal choices for different prioritizations of the two responses. For example, the choice $(0.5, -0.5)$ is in the PS, and this would be the best if we valued only good performance for Y_1 and did not emphasize Y_2 at all (it only achieves a value of 45 for Y_2). Similarly, the solution $(0.5, 0.5)$ would be the choice that maximizes Y_2, without considering Y_1's value (approximately 94.5). Between those two are locations in the PS that correspond to different amounts of prioritization of the responses. Larger dots correspond to bigger Y_1 values and show the connection between different locations in the input space and their values of the responses in the bottom right subplot of Figure 2. Clearly, since we are only considering non-dominated points, improving Y_1 results in less desirable (smaller) values of Y_2. Hence, the smallest dot corresponds to the worst value of Y_1 on the PF, but the best value of Y_2.

In this example, the PS locations in the input space form a "horseshoe" shape region, which resembles the path of traveling from one maximum location to the other mainly along the direction allowing the slowest descent hence the most tradeoff between the two responses. The locations at the "elbow" of the "horseshoe" shape represent values that prioritize good performance on both responses nearly equally. Note that if we were to select a solution near the elbow, it is possible to achieve Y_1 and Y_2 values that are both relatively close to their global maxima, say $(Y_1, Y_2) = (99.1, 49.5)$. This serves to highlight the benefits of examining the PF, since it easily allows identification of potential solutions and easy exploration of candidates to see the trade-offs between them. Lu and Anderson-Cook (2012) and Lu, Chapman, and Anderson-Cook (2013) describe some graphical methods for selecting between solutions depending on how much the different responses are prioritized. The amount of trade-off required between values of the responses is inherent to the problem. The maxima for each of the responses and the value of the other response at that location define the ranges from which the practitioner will need to choose. In this case, there are a number of solutions that are quite close to the Utopia point, and hence it should be possible to achieve excellent (if not optimal) performance on both responses.

Also shown in the bottom right subfigure of Figure 2 is the hypervolume value (Lu and Anderson-Cook 2013; Cao, Smucker, and Robinson 2015) that summarizes the amount of trade-off between the two criteria. Note that the points on the PF are connected by a step function, indicating that with the discretization of the input space to possible settings, there are no available solutions between adjacent points. The hypervolume (HV), or "normalized hypervolume" (Poles, Fu, and Rigoni 2009; Bader 2010) is calculated by scaling the criteria between 1 and 0, with the best value (here the maximum for each response) being assigned a value of 1, and the worst value on the PF being assigned 0. In this case with just two responses, the HV is a fraction of the area of the square $[0,1]^2$ that has attainable solutions. Namely, it excludes the region between the PF and the Utopia point. If we are able to get quite close to the Utopia point with some solutions, then the excluded region is small, and the HV value will be close to 1. If there are more severe trade-offs between the responses, then the excluded portion will increase, and the HV will be smaller. There are two special cases: if the PF consists of just a single point, which is equal to the Utopia point, this is the ideal (but rare) case with no trade-off between criteria, and this leads to a HV value of 1. If there are just two points on the PF, then this represents maximum trade-off between the two responses, with either the maximum of one paired with the minimum of the other on the PF, or vice versa. This leads to a HV value of 0. In this example, the HV value is 0.9149 which indicates that there is not a too severe trade-off between responses, and that it is possible to nearly attain maximum values for both responses simultaneously.

We now consider variations of this MM optimization to gain more understanding about how changes to the different features of the response surfaces impact the resulting Pareto front characteristics. There are a number of different features of each of the response surfaces that might change between different scenarios. First, if the same shaped surfaces are chosen for each response, but they are just *shifted up or down*, (say with response Y_1 now having a maximum of 90 or 110), then the locations in Figure 2c remains unchanged with the same locations in the input space being considered. Figure 2d would have a PF that looks the same, except that the scales would be changed to reflect the

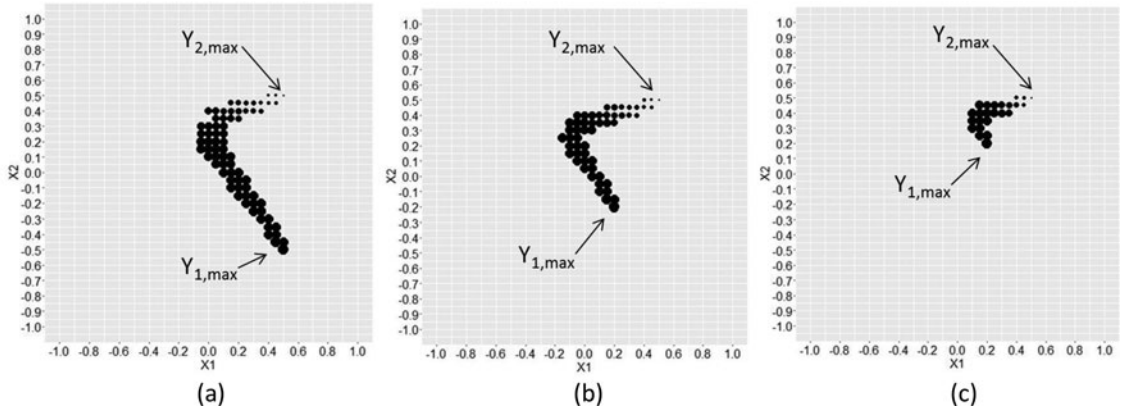

Figure 3. Pareto front set for different locations of the maximum for Y_1, with maxima at (a) (0.5, −0.5), (b) (0.2, −0.2), and (c) (0.2, 0.2). For all plots, the maximum for Y_2 is located at (0.5, 0.5).

shifted surfaces. The hypervolume value remains the same for any shift in the responses, since the scaling of the shifted response will adjust with it.

A second change to consider is if the locations of the two response surfaces *maxima are moved closer or farther apart*. For example, in Figure 2 the maxima are located at (0.5, −0.5) for Y_1 and (0.5, 0.5) for Y_2. If these optima move closer together, then the number of points in the PS gets smaller (corresponding to a shorter path between the maxima) and the range of values on the PF gets smaller, until there is just a single solution (the Utopia point) if the location of the two maxima are at the same point. If the solutions move further apart, then the PS includes more locations and the range of the PF values gets larger for each response. Figure 3 shows a sample of the PS locations for three scenarios as the distance separating the locations of the maxima are varied in the input space. As the locations of the maxima change, the PF plots that correspond to the bottom right of Figure 2 experience changes in the range of values on the PF for each response. If the shapes of the surfaces stay the same, but the maxima are moved further apart, the range of values included on the PF gets larger.

The hypervolume values for the three scenarios shown in Figure 3a–c are 0.915, 0.912, and 0.877, respectively, as the maxima are moved closer together. This is due to the use of a step function between adjacent points on the PF (Lu and Anderson-Cook 2013), as there are no options available between identified solutions. Hence, for the three PFs in similar shapes, the PF with more points (corresponding to maxima being further apart) has a larger HV based on having more intermediate choices (corresponding to a smaller area between the Utopia point and the front). However, note that this "normalized hypervolume" is calculated based on scaling the solutions on the PF between the "best" and "worst" values for each response on the PF. Note, when the PF has a narrow range of solutions with near-optimal response values, the HV value could be small due to the limited number of solutions identified on the PF for a given discretization of the input space. Therefore, it is desirable to check the range of response values first before using HV for interpreting trade-off relationship. If the range of values on the PF for a given response is small, then less emphasis on the HV is recommended since all points on the front may be deemed acceptable.

Third, the *orientation of the response surface ellipses* also impacts the PF and PS locations in the input space. In the example shown in Figure 2, the major (direction of the slowest change in the response) and minor (direction of the fastest change in the response) axes of the ellipses are oriented differently for each of Y_1 and Y_2. This results in the horseshoe shape for the PS locations in the input space. If the orientation of the two response surfaces match (both major axes share a common angle, regardless of where the locations of the maximum value lies in the input space), then the PS locations shown in the bottom left of Figure 2, would be located closely along the straight line connecting the locations of the two maxima. The PF locations will fall exactly along a straight line connecting the maxima locations if the orientation of either the major or the minor axis aligns with this straight line (a direct path between the maxima). Maximum curvature of the PF occurs when the major axes are 45° apart. At the other extreme, if the orientations of the surfaces are 90° apart, the PS locations may have some curvature, but not as much as for intermediate values. Figure 4

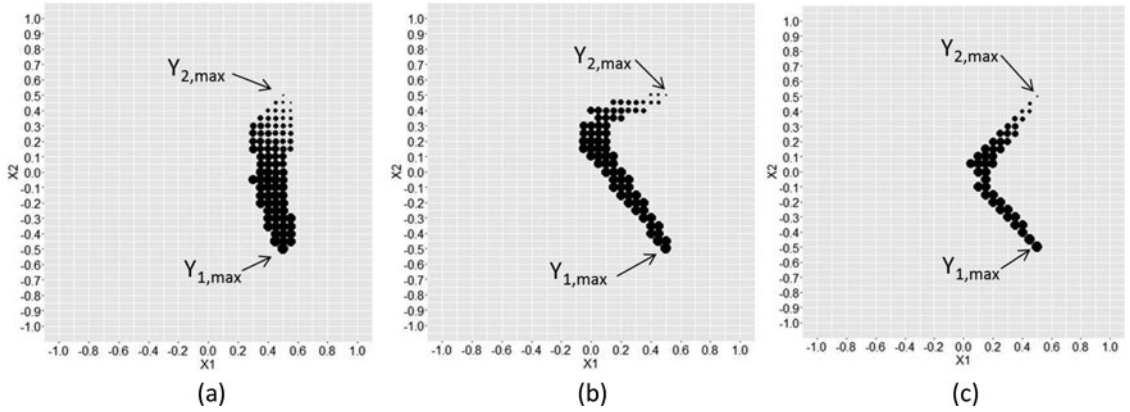

Figure 4. Pareto front set for different orientations of the responses. The maxima for the responses are at (0.5, −0.5) for Y_1 and (0.5, 0.5) for Y_2. (a) major axes for two responses parallel; (b) major axes differing by 45°; and (c) major axes perpendicular.

illustrates how the shape of the PS locations in the input space changes for different orientations of the surfaces. Note that the PS locations in the input space follow the orientation that allows for the slowest change in each surface, i.e., the locations stay close to the major axes for both surfaces to change values as slowly as possible. The HV values for the PFs in Figures 4a–c are 0.830, 0.914, and 0.943, respectively. The more curve shaped PF is associated with the less trade-off (higher HV) due to the existence of more intermediate solutions than the more straight line-shaped PF given the same end points of the PF determined by the same maxima locations.

Another aspect of the response surfaces that impacts the PS locations and PF is the *shape of the contours for each response*. The contours in Figure 2a and b are moderately elongated ellipses. If contours for one or both of the surfaces became circular (with all directions yielding the same rate of descent from the maxima), then the PS locations in the input space fall exactly on a straight line (connecting the two maxima). Figure 5 shows a case where Y_1 has circular contours, while Y_2 remains unchanged. This case is actually a special case of the discussion above about the orientation of the response surfaces, where the circular contours means that there are no differences in orientation between the two surfaces. Figure 6 shows an alternative case when the contours of Y_1 have been elongated to form a near ridge surface, which yields a very curved set of PS locations in the input space. Note, the hypervolume for the PF for the straight line PS is 0.816 which is considerably smaller than the hypervolume of the PF for the more curved PS as 0.930. This can again be partially explained by the richness of choices with the larger PS along the very curved path between the maxima allowing more intermediate solutions.

In the examples shown above, we have changed one feature at a time from the original example in Figure 2. If several features of the surfaces are changed, we can think of assembling the new PS locations and PF values

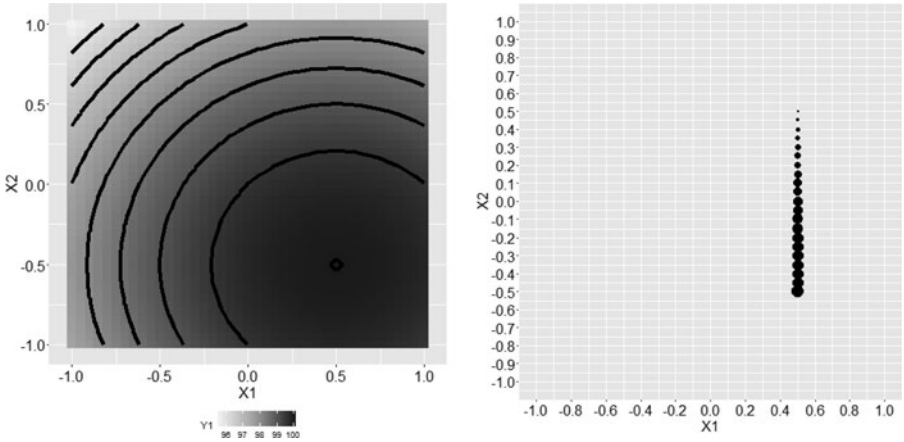

Figure 5. Example of shape of Pareto set locations when one response has circular contours. The maxima for the responses are at (0.5, −0.5) for Y_1 and (0.5, 0.5) for Y_2. (a) Y_1 contour plot (Y_2 contour plot same as Figure 2b), (b) Pareto set in input space.

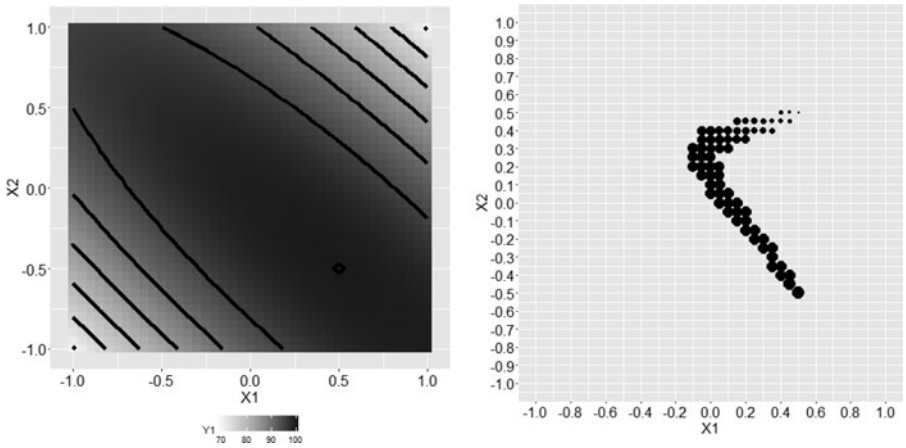

Figure 6. Example of shape of Pareto set locations when Y_1 is a near ridge surface. The maxima for the responses are at (0.5, −0.5) for Y_1 and (0.5, 0.5) for Y_2. (a) Y_1 contour plot (Y_2 contour plot same as Figure 2b), (b) Pareto set in input space.

with a thought experiment that walks through a progression of adaptations to change from this illustrated example to the scenario of interest.

Finally, we consider what would occur if the underlying scenario of having a "mountain" (both eigenvalues are negative) is changed. If the stationary points are located inside the design region and one or both of the surfaces has at least one positive eigenvalue, then the desirable large values for that response will likely occur at one or more of the edges of the selected input space. Another scenario that could result in the PS being pushed to the edge of the region is if the peak of the mountain is located outside of the input space. In this case, the PS typically has locations near the edges of the input space, and can have disjoint regions. This scenario is similar to some cases that we consider in greater detail in the coming sections.

Maximizing-targeting (MT) optimization

In this section we consider another two response optimization scenario with the goal of maximizing Y_1 while achieving a target for Y_2. We again restrict our examination to cases where both of the response surfaces have a maximum stationary point located in the design region. We return to the first scenario with the same response surfaces shown in Figure 2, but now consider setting a target value for response Y_2 at $T = 47$ (recall that the overall maximum for Y_2 is 50). Figure 7 shows the results for this optimization. Note that now the top right figure has been changed to show the absolute difference from the target value, with the ideal values of zero (white contour), matching where the contour for values of 47 are in Figure 2. The Pareto set in the bottom left corner of Figure 7 has changed to reflect the different criteria for the second response, and now has some disjoint groups of locations that connect the maximum of Y_1 with the target contour for Y_2.

Note that with disjoint locations in the input space, there is some additional complexity associated with choosing an optimal operating location, as more diverse options generally exist. The Pareto front plot in the bottom right corner of Figure 7 now has the Utopia point (*) located in the bottom right corner (maximizing Y_1 and minimizing $|Y_2 - T|$). The range of values attainable on the Pareto front is $Y_1 \in \{96.4, 100\}$ and $|Y_2 - T| \in \{0, 2\}$.

The hypervolume calculation shown in Figure 7 reflects changes in the optimization scenario. In this case, we are again examining the attainable fraction of the $[0,1]^2$ square for the scaled values of our responses. Visually, we can still interpret the HV value for this case as one minus the fraction of the region located between the Utopia point and the PF in Figure 7. We again scale the best possible values to 1 for each response, and the worst value on the Pareto front to 0. This time the worst value for Y_2 corresponds to the largest distance from the target, and we select a distance of 0 from the target to be scaled to the value of 1. For this example, there are a number of solutions that have response values quite close to the ideal values, and hence we have a HV that is very close to 1. As later examples in this section show, this can have an important impact on the values of the HV if the target is not achievable in the input space.

We now explore the effect on the PS and PF when various aspects of the response surfaces and our optimization criteria are manipulated. We first examine the impact of changing the target value (relative to the

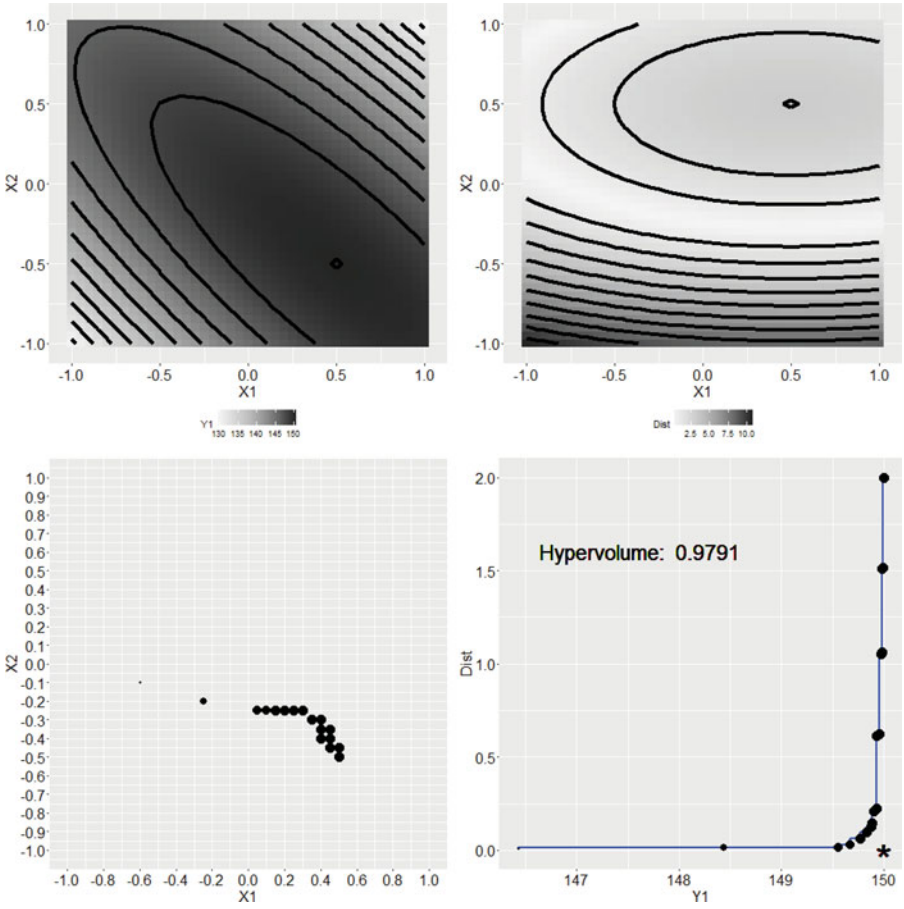

Figure 7. Summary of response surfaces and their associated Pareto set and Pareto front values for a MT optimization problem. Response surfaces for the two responses correspond to Figure 2, and the target value for Y_2 is 47. The top row shows a contour plot of each criterion being optimized for Y_1 and Y_2 with the dark-to-light color indicating large-to-small values of Y_1 (to be maximized) and $|Y_2 - T|$ (to be minimized). The bottom left plot shows locations in the input space of the Pareto set, while the bottom right plot shows the Pareto front. The size of dots in the bottom sub-figures is proportional to the value of Y_1. Note that the Utopia point (*) for this scenario is located in the bottom right corner of the bottom right figure.

range of the observed responses), and then we also generalize patterns in the PS and PF across other variations for the MT scenario.

For the response surfaces shown in Figure 2, we now consider changing the target value relative to the range of Y_2 in the input space. Figure 8 shows the $|Y_2 - T|$ contour (top row), the PS (middle row), and PF plots (bottom row) for several scenarios. When the target value falls below the smallest observed Y_2 (Figure 8a), then this is essentially equivalent to a maximizing-minimizing problem, with the PS locations connecting portions of the edge of the input space where the smallest Y_2 values are located with the best values of Y_1. It should be noted that in this case, achieving the target value is not possible, and the range of observed $|Y_2 - T|$ values does not contain 0. Note that this has a substantial impact on the HV value (here HV = 0.4745), since the PF is located far away from the Utopia point in the Y_2 distance dimension. This small value is a reminder that possible solutions require considerable sacrifice towards attaining the target value, since the smallest Y_2 value in the input space is approximately 37 and the target value is 30.

For target values above the maximum Y_2(Figure 8c), this corresponds to the MM problems described in the previous section, and again means that we are not able to achieve the target value. The PS is a continuous region and matches patterns described previously. In the PF plot for this scenario, we see that since the maximum value of Y_2 in the design region is 50 and the target is set to 52, we are not able to find a location that achieves the target. Hence, this penalty for a less desirable set of solutions is reflected by a relatively small HV value of 0.6635. For the cases described in Figures 8a and c, the Pareto set will remain the same for any choice of target beyond the range of attainable values, but the

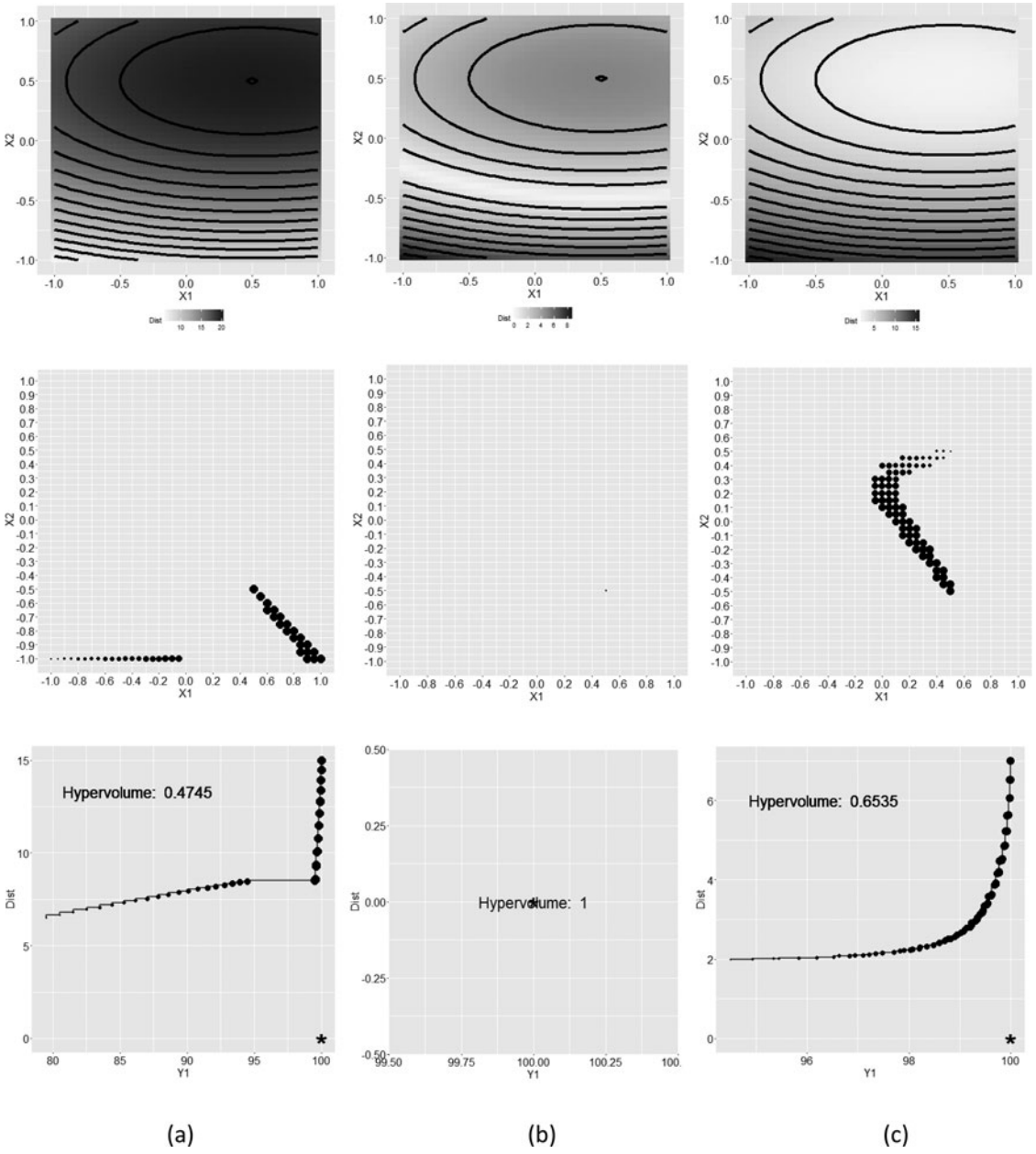

Figure 8. Changes in $|Y_2 - T|$ contour (top row), the Pareto set (middle row) and the Pareto front (bottom row) plots for target values of (a) 30 (PS same for any value below the smallest predicted Y_2); (b) 45; and (c) 52 (PS same for any value above maximum Y_2). Note that there can be multiple regions in the input space that contribute to the PS, and the PF solution values do not always connect smoothly.

HV will change depending on how far away the estimated values of Y_2 is from the target.

There are interesting patterns for the PS when the target falls within the predicted range of estimated Y_2 values. In these cases the PS locations seek to connect some portion of the $|Y_2 - T| = 0$ ellipse with the maximum of Y_1. In Figure 8b, we see an example of the rare special case of the target value (45) for Y_2 and the maximum value of Y_1 being obtainable at the same location in the input space. This allows the Utopia point to be achieved. This requires no trade-off between the responses since we are able to simultaneously optimize both, and hence the HV value is 1.

There are many different forms that the PS can take when we are considering the MT scenario. A key distinction between the MM case and MT case is that if the target is attainable in the input space, it is generally associated with a contour instead of a single location. (An exception is if the target and the maximum (or minimum) in the region have the same value.) This means that the PS location chosen for the optimum $|Y_2 - T|$ (at one end of the PF) is based on choosing the

largest Y_1 across the locations on the best contour for $|Y_2 - T|$. The other end of the PF (where we are maximizing Y_1) contains no choices about which Y_2 to select. For intermediate locations on the PF, to find points on the PS involves looking across the two contours associated with $|Y_2 - T| = c$, one Y_2 above the target, and one Y_2 below the target, and selecting the largest Y_1 value possible. The larger number of choices from which to choose often leads to disjointed locations in the PS as different regions in the input space emerge as best for balancing the response values. Unlike the MM scenarios described previously, where changes translate into easily characterized patterns in the PF and PS, the MT scenario has many more variations and is difficult to generalize. We have to restrict our discussion to the above mentioned cases due to the length constraint of the article. However, the R shiny apps will allow the practitioners to explore broader scenarios with different variations and/or constraints in different applications. A more detailed description of the R shiny app options is available in the Appendix.

Targeting-targeting (TT) optimization

The final scenario that we consider is when we are attempting to simultaneously achieve targets for both Y_1 and Y_2. We saw a substantial increase in complexity for anticipating where the Pareto set would be located as we moved from the MM to MT scenarios described in previous sections. This loss of easily recognizable patterns continues as we examine the TT case. We again consider the response surfaces shown in Figure 2, but now consider that the desired optimization is to achieve a target of Y_1 for $T_1 = 90$ and Y_2 for $T_2 = 47$. The top row of Figure 9 shows the contour plots for $|Y_1 - T_1|$ and $|Y_2 - T_2|$, with the white regions in each plot denoting where the estimated function achieves the target value for that response. When we examine

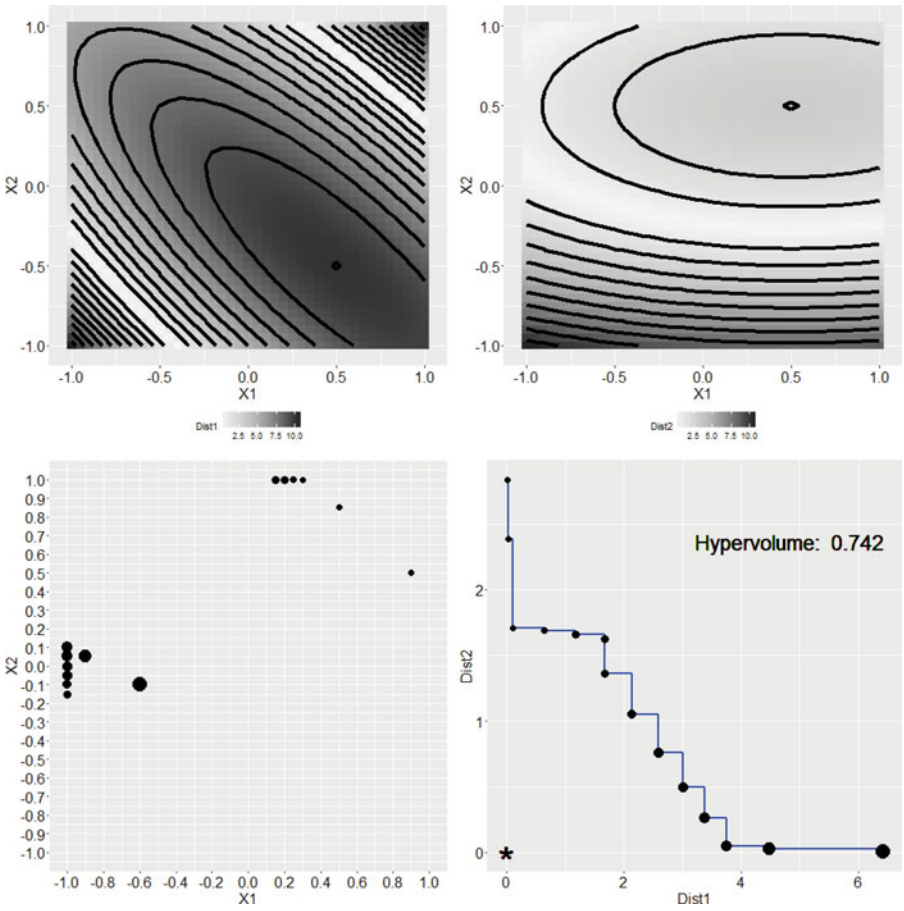

Figure 9. Summary of response surfaces and their associated Pareto set and Pareto front values for a TT optimization problem. Response surfaces for the two responses correspond to Figure 2, and target values for Y_1 and Y_2 are 90 and 47, respectively. The top row shows a contour plot of each criterion being optimized for Y_1 and Y_2. The bottom left plot shows locations in the input space of the Pareto set, while the bottom right plot shows the Pareto front. The size of dots in the bottom sub-figures is proportional to the value of $|Y_1 - T_1|$ with the smallest dot optimal for Y_1. Note that in the bottom right figure, the Utopia point (*) for this scenario is located in the bottom left corner.

the bottom left subplot of Figure 9, we see that there are multiple regions in the input space that have been selected as part of the PS. The smaller symbols correspond to best values for Y_1, and closely follow the contour $|Y_1 - T_1| = 0$ near the top right corner of the input space. Similarly, the larger symbols are close to the white contour for $|Y_2 - T_2|$. Intermediate-sized symbols are close to the left edge of the input space where some balance of good performance for both responses is possible.

As with the example in Figure 7, the separation between locations in the input space for different options on the PF can complicate the choice of a best location where the factor should be set. Examining the bottom right subplot of Figure 9, we note that there are substantially greater trade-offs between response objectives than in the MM case. Here, there are several locations in the input space where one response can be optimized at considerable expense to the other. Additionally, there are intermediate values where both responses are balanced between the ideal values. This degree of trade-off leads to a more moderate HV value of 0.742. In addition, we can see how the distance from the target differs considerably for the two responses, with the distance from target ranging between [0, 2.8] for Y_1 and [0, 6.3] for Y_2.

Trying to generalize patterns in the PS and PF for the TT scenario is challenging since there are a large number of factors affecting the nature of the solutions. If the target value for one or both of the responses lies outside the range of estimated values in the input space, then the problem translates to either a maximization or minimization for that response. In general if the targets lie within the estimated ranges for both responses, some generalization is possible. For the ends of the PF (where all the emphasis is placed on optimizing one of the responses), the general approach is to focus on contour $|Y_1 - T_1| = 0$ for the primary response and then find the best available value of the other response. For intermediate locations that achieve some balance between the response performance generally involves looking across the two contours associated with $|Y_1 - T_1| = c$ (above and below the target), and selecting the best value possible for the other response.

Discussion and conclusions

Since most experiments involve collecting data on more than one response, MRO is a common situation for many practitioners who need to select the most desirable input combinations for their product or process. While reducing the optimization process to some variation of aggregating the multiple responses into a single objective has often been used, we recommend careful exploration of alternatives to obtain the best solution that matches the priorities of the study. This article presented some exploration and summaries of the trade-offs patterns between two responses and provided some guidance on how to make the most desirable choices for three types of MRO problems with different combination of optimization goals.

To simplify the characterization of the different possible response surfaces, we used the canonical form of the second-order response surface model, as it compactly describes the nature of the stationary points and how the responses change as we move away from it. By examining different configurations of the nature and location of the stationary points, the magnitudes of the eigenvalues and the directions of the eigenvectors, we describe (1) how trade-offs and choices vary for different types of MRO problems and (2) how to identify contending solutions in PS. Clearly, there are many more combinations of inputs that might be explored, and so we encourage the reader to explore the interactive Shiny apps (at https://ycao.shinyapps.io/CAL_xx/, where xx = MM, MT, or TT for the three scenarios) for all the three types of MRO that generated the plots in this article. The Appendix provides the equations to convert from the canonical form to the standard equations for a second order response surface model. We believe that it is usually too simplistic to either focus on just a single response or to weight all responses equally when looking to find an optimal set of input-factor levels. Exploring and understanding the trade-offs is essential in MRO, because it highlights the relative changes in one response when another is altered. The Hypervolume (HV) metric provides a global summary that quantifies the degree of trade-offs between responses. Code for calculating HV is available on GitHub at https://github.com/statlife/Multiple-Response-Optimization.

Throughout the paper, we used the same sized grid for examining the Pareto sets and fronts. Depending on how precisely the input factors can be set in the process, it is helpful to match the chosen grid to what is possible in practice. If the grid is made coarser, then there are fewer points in the PS, and the PF is less rich. The diminished number of choices also means that while

the overall shape of the PF stays the same, the calculated value of the HV will become smaller.

An important aspect that we have not addressed in this article, but is worth emphasizing is the role that estimation plays in the optimization process. The PS and PF solutions discussed in this paper are based on having a single estimated equation for each of the responses. In general, the response surfaces that are obtained from an experiment are estimated with uncertain model parameters. Hence, some consideration should be given to alternative response surfaces that are also consistent with the observed data. Chapman, Lu, and Anderson-Cook (2014a, 2014b) provide two strategies for incorporating uncertainty into the optimization and understanding the impact that it can have on the attainable results of a selected set of input factor levels. More realistic decisions can be made with improved understanding of the associated uncertainty.

About the authors

Christine M. Anderson-Cook is a Research Scientist in the Statistical Sciences Group at Los Alamos National Laboratory. She is a Fellow of the American Statistical Association and the American Society for Quality. Her research areas include response surface methodology, multiple criteria optimization, design of experiments, reliability, and graphical methods.

Yongtao Cao is an Assistant Professor in the Mathematics Department of Indiana University of Pennsylvania. He earned his Ph.D. in Statistics at the University of Wyoming. His research areas include design of experiments, multiple criteria optimization and response surface methods.

Lu Lu is an Assistant Professor in the Department of Mathematics and Statistics at the University of South Florida. She earned her Ph.D. in Statistics at Iowa State University and was a Postdoctoral Research Associate at Los Alamos National Laboratory. Her research areas include reliability, multiple response optimization, design of experiments, and sampling.

References

Anderson-Cook, C. M., Y. Cao, and L. Lu. 2016. Statistics roundtable: Maximize, minimize or target. *Quality Progress* 49 (4):52–55.

Bader, J. M. 2010. *Hypervolume-based search for multiobjective optimization: Theory and methods*. Seattle, WA: Createspace Independent Pub.

Cao, Y., B. J. Smucker, and T. J. Robinson. 2015. On using the hypervolume indicator to compare pareto fronts: Applications to multi-criteria optimal experimental design. *Journal of Statistical Planning and Inference* 160:60–74.

Chapman, J. L., L. Lu, and C. M. Anderson-Cook. 2014a. Process optimization for multiple responses utilizing the pareto front approach. *Quality Engineering* 26:253–268.

Chapman, J. L., L. Lu, and C. M. Anderson-Cook. 2014b. Incorporating response variability and estimation uncertainty into pareto front optimization. *Computers & Industrial Engineering* 76:253–267.

Lu, L., and C. M. Anderson-Cook. 2012. Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering* 24:404–422.

Lu, L., and C. M. Anderson-Cook. 2013. Adapting the hypervolume quality indicator to quantify trade-offs and search efficiency for multiple criteria decision-making using pareto fronts. *Quality and Reliability Engineering International* 29:1117–1133.

Lu, L., C. M. Anderson-Cook, and T. J. Robinson. 2011. Optimization of designed experiments based on multiple criteria utilizing a pareto frontier. *Technometrics* 53:353–365.

Lu, L., J. Chapman, and C. M. Anderson-Cook. 2013. A case study on selecting a best allocation of new data for improving the estimation precision of system and subsystem reliability using pareto fronts. *Technometrics* 55:473–487.

Myers, R. H., D. C. Montgomery, and C. M. Anderson-Cook. 2016. *Response surface methodology: Process and product optimization using designed experiments*. 4th ed. New York: Wiley.

Poles, S., Y. Fu, and E. Rigoni. 2009. The effect of initial population sampling on the convergence of multiobjective genetic algorithms. In *Multiobjective programming and goal programming*, 123–133. Heidelberg: Springer Berlin.

Appendix

A1: Summary of Options in the Shiny Apps (https://ycao.shinyapps.io/CAL_MM/, https://ycao.shinyapps.io/CAL_MT/, and https://ycao.shinyapps.io/CAL_TT/)

1. Common features for all three apps for changing the underlying surfaces for responses Y1 and Y2: For each response, the following inputs can be manipulated to the response surface:
   - **X1 and X2 locations** (restricted to design region [−1,+1])—defines the location of the stationary point of each surface.
   - **Maximum response value** (range = [0,200/250])—sets the value of Y at the stationary point.
   - **Ratio of eigenvalues** (range = [−10,15])—defines the type of stationary point (positive =

maximum, 0 = ridge system, negative = saddle point) and relative rate of change moving away from stationary point (1 = circular contours, larger values = less circular ellipses).

- **Angle of maximum change for eigenvectors**—defines the direction of fastest change of the surface when moving away from the stationary point (0 = fastest change in horizontal direction, 90 = fastest change in vertical direction).

2. Feature that is specific for apps with optimization involving targeting (MT and TT)
   - **Target value** (range = [0,250])—defines the target value of response (criterion used for constructing the Pareto front is $|Y_i - \text{Target}|$).

A2: Equations for converting between response surface forms

1. Canonical to Standard:
   Elements of canonical form: ($\hat{y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2$ for each response)
   $x_{st} = \begin{pmatrix} x_{1s} \\ x_{2s} \end{pmatrix}$ = location of stationary point
   $\hat{y}_{st}$ = value of y at stationary point
   λ_1, λ_2= eigenvalues
   $w_1 = \begin{pmatrix} w_{11} \\ w_{12} \end{pmatrix}$ = standardized eigenvector associated with λ_1
   $w_2 = \begin{pmatrix} w_{21} \\ w_{22} \end{pmatrix}$ = standardized eigenvector associated with λ_2
   Conversion to standard form ($\hat{y} = b_0 + b_1 X_1 + b_2 X_2 + b_{11} X_1^2 + b_{22} X_2^2 + b_{12} X_1 X_2$ for each response):

$$b_0 = \hat{y}_{st} + \left(w_1' x_{st}\right)^2 + \left(w_2' x_{st}\right)^2$$

$$b_1 = -2w_{11}\left(w_1' x_{st}\right) - 2w_{21}\left(w_2' x_{st}\right)$$

$$b_2 = -2w_{12}\left(w_1' x_{st}\right) - 2w_{22}\left(w_2' x_{st}\right)$$

$$b_{11} = w_{11}^2 + w_{21}^2$$

$$b_{22} = w_{12}^2 + w_{22}^2$$

$$b_{12} = 2w_{11}w_{12} + 2w_{21}w_{22}$$

2. Standard to Canonical: see pp. 278–279 of Myers, Montgomery, and Anderson-Cook (2016).