Impact of Response Variability on Pareto Front Optimization

Jessica L. Chapman,[1]* Lu Lu[2] and Christine M. Anderson-Cook[3]

[1]*Department of Mathematics, Computer Science, and Statistics, St. Lawrence University, Canton, NY 13617, USA*

[2]*Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620, USA*

[3]*Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA*

Received 19 June 2014; revised 26 February 2015; accepted 1 June 2015
DOI:10.1002/sam.11279
Published online 1 September 2015 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: A two-stage Pareto front approach can improve the process of making a decision about which input values simultaneously optimize multiple responses. However, ignoring estimation uncertainty and natural variability in the responses can potentially lead to suboptimal choices about those input values. A simulation-based approach is used to quantify and examine the impact that variability has on the superior solutions identified on the Pareto front and their performance. Because each optimization scenario has its own unique characteristics, including responses with different amounts of natural variability, the impact of variability on the solutions varies from situation to situation. We study how varying the amount of response variability affects the locations identified for the front and the characteristics of the most promising solutions on the front. We illustrate the method with an application involving process improvement through variance reduction. © 2015 Wiley Periodicals, Inc. Statistical Analysis and Data Mining: The ASA Data Science Journal 8: 314–328, 2015

Keywords: multiple response optimization; response surfaces; estimation uncertainty; simulation

1. INTRODUCTION

When optimizing a process with multiple responses, combinations of input factors are sought to simultaneously optimize all responses. A common practice for making this selection includes conducting a designed experiment to estimate the multiple response surfaces within the operating space of the input factors and then using the desirability function (DF) approach [1] to find a solution that optimizes an overall metric that combines the multiple responses for a set of predetermined weights. However, this method can be sensitive to the subjective choices of the user-specified priorities of the responses (as summarized by the weights), the scaling scheme for converting the response values to desirability values between 0 and 1, and the metric for integrating multiple responses into a single quantitative measure.

Chapman *et al*. [2] used a Pareto front (PF) approach ([3–5]) to find promising input combinations using a structured decision-making process, which allows practitioners to explore all promising solutions (on the PF) and examine their trade-offs and performance as well as understand the impact of various subjective choices made throughout the process. The PF approach has been used extensively in many different disciplines, such as engineering and computer science, as a tool for multiple response optimization. However, the majority of the literature has focused on how to find the PF with its most promising choices without providing more insights on how to proceed from those choices to a final decision. Lu *et al*. [3] first introduced the method in design selection and proposed a structured process with statistical tools developed ([3–5]) for guiding informed decision-making. The process can be divided into two stages: (1) an objective stage where the PF consisting of all superior solutions that are not strictly outperformed by any other solutions based on all responses is identified by strategically eliminating all non-contending (i.e. inferior) solutions and (2) a subjective stage that introduces user-specified choices to hone in on the most appropriate solution to match the user's priorities. For every solution on the PF, no uniformly better options

* *Correspondence to:* Jessica L. Chapman (jchapman@stlawu.edu)

exist. Hence the PF provides an objective set of promising solutions that are sensible to be considered in further steps. By seeing the complete set of promising solutions, the users are offered more flexibility on understanding the potential impact from various subjective choices and hence more confidence in making an informed decision.

The simplest implementation of this approach optimizes the estimated response surfaces using the point estimates of the parameters (referred to as the mean model) and examines a fine grid of input factor combinations to find the Pareto optimal solutions. However, due to sampling variability, a range of values for the model parameters are plausible to be consistent with the observed data. Hence, the uncertainty in the estimated response surfaces introduces uncertainty in the decisions for optimizing the multiple responses.

To incorporate estimation uncertainty into the decision-making process, Chapman *et al*. [6] expand the two-stage PF approach to quantify the variability in the identified PF to evaluate its potential impact and facilitate an informed decision. More specifically, the method simulates a large number of response surfaces that are consistent with the observed data and finds the PF associated with each set of the simulated response surfaces. By characterizing the PFs across a large number of simulations, the objective stage can be expanded to also summarize the frequency with which locations appear on the fronts and examine the uncertainty in the PFs, which is propagated from the uncertainty in estimating the models. Furthermore, in the subjective stage, promising solutions based on user-specified priorities, scaling, and metric choices are explored and summarized across all simulated PFs. By understanding the robustness of different solutions to the subjective choices and across the simulations, decision makers can identify solutions that are best, most frequently, for their particular weighting priorities.

Since each optimization situation has its own unique characteristics, how the variability in the responses impacts results can vary considerably. Different responses have different degrees of natural variability, which may have substantial impact on the variability associated with the PF choices as well as the final decisions. Additionally, the role of variability interacts with other features impacting the PF, including how much the response values vary across the operating space and the degree of trade-off between the responses. For instance, in a simplistic example, if all of the responses are optimized at the same location, then changes in identifying the PF would only be a function of how much uncertainty there is around that best location. On the other hand, if the best location for each response lies in different regions of the operating space, then different amounts of variability can shift the front more as the trade-offs between responses change.

In this paper, we evaluate the impact of differing amounts of variability in individual responses on the identification of the PF (the objective first stage in the approach [6]) and on the characteristics of the promising solutions based on flexible weighting choices (evaluated early in the second subjective stage). We focus primarily on these portions of the process because they identify and characterize a subset of input combinations from which all subsequent decisions are made. We consider optimizing the same chemical process described in Myers *et al*. ([7], p. 253) and conduct a simulation study to investigate how changing the amount of variability for each of the responses affects the choice of the PFs and further decisions about the optimal operating conditions.

The chemical process considered has two input variables ($\xi_1 =$ time and $\xi_2 =$ temperature) and three responses of interest ($y_1 =$ yield, $y_2 =$ viscosity, and $y_3 =$ number-average molecular weight). The operating region was selected to include the time of the process (between 77 and 93 minutes) and the temperature (between 167 and 183 degrees Fahrenheit). A 13-run central composite design (CCD) ([7], p. 297) for a circular coded region with maximum radius of $\sqrt{2}$ was run to estimate the three independent response surfaces. After fitting quadratic models and removing nonsignificant terms for all responses, the estimated response surfaces, in terms of x_1 and x_2, the coded time, and the temperature variables, respectively, are given in Eq. (1) for yield, Eq. (2) for viscosity, and Eq. (3) for molecular weight.

$$\hat{y_1} = 79.94 + 0.995x_1 + 0.52x_2 + 0.25x_1x_2 - 1.38x_1^2 - 1.00x_2^2 \quad (1)$$

$$\hat{y_2} = 70.0 - 0.16x_1 - 0.95x_2 - 1.25x_1x_2 - 0.69x_1^2 - 6.69x_2^2 \quad (2)$$

$$\hat{y_3} = 3386.2 + 205.1x_1 + 177.4x_2 \quad (3)$$

The goal of optimization is to maximize yield, y_1, and minimize both molecular weight, y_3, and distance from the viscosity to a target value of 65, $|y_2 - 65|$. Fig. 1(a) displays the overlaid contours for the mean model estimated by Eqs. (1)–(3). For yield, the maximum occurs around $(\xi_1, \xi_2) \in [85, 89] \times [173, 179]$, with lower response values for locations farther away from this region. As both time and temperature decrease, molecular weight becomes smaller and more desirable. Viscosity achieves its target value (lightest gray contour in Fig. 1(a)) when the temperature is near either 171 or 178 degrees Fahrenheit and becomes less desirable as it moves away from these regions. Therefore, there is no location that is universally best for all responses. Simultaneously optimizing the three responses requires making some trade-offs, especially

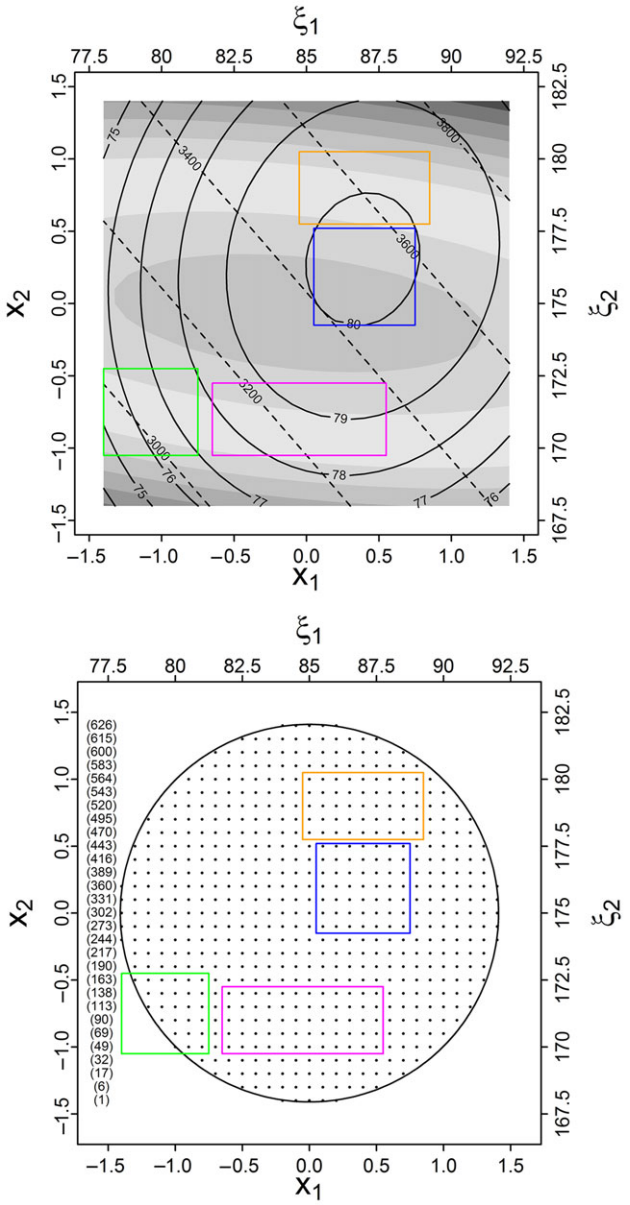

Fig. 1 (Top) Overlaid contours for the three estimated response surfaces (mean model), as estimated from Eqs. (1)–(3). Grayscale contours are for viscosity's distance from target (with the lightest gray corresponding to the ideal values), oval contours are for yield (to be maximized), and diagonal line contours are for molecular weight (to be minimized). The colored rectangles identify regions with good performance in different combinations of priorities: green: good for molecular weight and viscosity; orange: good for yield and viscosity; blue: excellent for yield but poor for others; pink: most balanced for the three responses. (Bottom) Circular region defined by the points used in the CCD. Numbers along the left edge identify the specific points on the grid. All subsequent circle plots are plotted with the same scale. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

between yield and molecular weight, as the optimal locations for one of the two responses are generally associated with poor performance for the other.

In Fig. 1(a), four regions are highlighted in different colors that correspond to locations where there is good performance for different priorities. The rectangular region in green at the bottom left identifies locations that perform well on molecular weight and viscosity, but poorly on yield. The pink region at the bottom consists of locations with more balanced performance across the three responses. The blue region toward the top right has locations that perform extremely well on yield but not as well on the other two responses. The orange region above the blue region generally has better performance on yield and viscosity than molecular weight.

To create a set of solutions from which the PF can be identified, Chapman *et al*. [2] describe creating a set of grid points within the desired operating range. Here, we construct a grid of 630 points located inside the circular coded region defined by the points used in the central composite design (CCD) (Fig. 1(b)). We assume that a distance of 0.1 for the coded values is the smallest possible increment when adjusting the factor levels, and thus the adjacent points in the same row or column are separated by 0.1. Beginning in the bottom row and moving from left to right, the points in Fig. 1(b) are labeled from 1 to 630; once the end of a row is reached, the labeling scheme continues with the leftmost point in the next lowest row. The regions identified in Fig. 1(a) are also marked in Fig. 1(b) to connect the factor input values with the estimated responses in the mean model. For each of the three responses, Table 1 summarizes the smallest and largest estimated responses, using Eqs. (1)–(3), for the locations in Fig. 1(b) (with the corresponding grid point identified in parentheses).

As seen in Table 1, the three responses have different variability of the response values across the operating space. Yield has the smallest absolute range in the estimated responses (4.44), followed by viscosity (14.88), and molecular weight has the largest range (752.8). However, the absolute range of the estimated responses should be interpreted in the context of their natural variability. The estimated models for the three responses (Eqs. (1)–(3)) have substantially different mean square errors (MSE), which, assuming adequate models with no substantial lack of fit, measure the natural variability in each response. The ratio of the range of the estimated responses relative to its estimated variability (square root of the MSE) provides a measure of the signal-to-noise ratio. A large ratio indicates that the response changes considerably across the operating region relative to its variability (noise). We would anticipate that responses with a large signal-to-noise ratio might influence the PF quite differently than a response that

Table 1. Summaries of the range of estimated response values, using Eqs. (1)–(3), relative to the estimated variability as measured by $\frac{\max(\hat{\mathbf{y}})-\min(\hat{\mathbf{y}})}{\sqrt{\text{MSE}}}$ for the three responses in the original study.

	Estimated response surfaces				
Response	$\sqrt{\textbf{MSE}}$	Min $\hat{\mathbf{y}}$	Max $\hat{\mathbf{y}}$	Range $\hat{\mathbf{y}}$	$\frac{\max(\hat{\mathbf{y}})-\min(\hat{\mathbf{y}})}{\sqrt{\text{MSE}}}$
Yield	0.2665	75.78 (244)	80.21 (406)	4.44	16.66
Viscosity	2.2749	55.15 (630)	70.03 (287)	14.88	6.54
Molecular weight	165.6231	3003.66 (49)	3756.46 (519)	752.8	4.55

appears relatively flat across the operating space relative to its variability. For our example, the yield response has a substantially larger signal-to-noise ratio (16.66) than the other two responses. Viscosity has slightly larger signal-to-noise ratio (6.54) than molecular weight (4.55).

The amount of variability in the responses also interacts with trade-offs between the responses. To quantify the degree of trade-off between the three responses as well as between each pair of responses, we can calculate the hypervolume under the PF (HVUPF) [8]. With a chosen scaling that maps the best performance for a criterion to desirability value 1 and the worst performance to desirability value 0, the shape of the PF relative to the rectangular region $[0, 1]^m$ can indicate the amount of trade-off between the m criteria under consideration. The closer the PF is to the Utopia point (which corresponds to simultaneously achieving the optimal values for all criteria), the less trade-off there is between the criteria, indicated by larger HVUPF values. The HVUPF is a single numerical summary with values between 0 and 1 that measures the hypervolume of the region between the origin (worst case for all criteria) and the PF within the rectangular region $[0, 1]^m$.

For our example, defining the best and the worst estimated responses in the operating region to have desirability values 1 and 0, respectively, the HVUPF for all three responses (corresponding to the volume under the PF within $[0, 1]^3$) is 0.73. The HVUPF values for each pair of the responses (corresponding to the area under the PF within $[0, 1]^2$) are 0.97 for yield and viscosity, 0.79 for yield and molecular weight, and 0.99 for viscosity and molecular weight. This indicates only a little trade-off between viscosity and either yield or molecular weight, while there is much more trade-off between yield and molecular weight. Therefore, it is possible to achieve quite good performance simultaneously on viscosity and yield (such as the orange region in Fig. 1(a)) or on viscosity and molecular weight (such as the green region in Figure 1(a)); however, considerable trade-offs between the responses are required when considering all three responses simultaneously.

To study the impact of the amount of response variability on the decisions made using the PF approach, we evaluate eight (2^3) combinations of degrees of variability for the three responses by adding or removing 50% of the observed variability from the mean model from Eqs. (1)–(3) (i.e. considering either 0.5MSE or 1.5MSE for each response). For each of the eight scenarios, we conduct a simulation study to generate a large number of response surfaces consistent with the estimates of the model parameters and the particular amount of variability. Compared to the unchanged response variability scenario, we can investigate how changes in the degree of natural variability in the responses are propagated to the PF choices. Additionally, in some processes, there may be opportunities to invest in reducing the natural variability of one or more responses through process improvement. Understanding how variation impacts the choices of the best input factor locations can help to guide decision-making about where to invest in these reductions.

The remainder of the paper is organized as follows. Section 2 details the simulation study. Section 3 describes the results of the simulated PFs across the different scenarios. Section 4 presents results from the subjective stage of the PF optimization process for a particular choice of the DF form. In Section 5, we consider a process improvement scenario and describe how a user might decide which variability reduction strategy is most beneficial. Section 6 provides some discussion and concluding remarks.

2. SIMULATIONS WITH VARIED VARIABILITY FOR MULTIPLE RESPONSES

To investigate the impact of response variability, simulations are conducted to generate new response surfaces that are consistent with the mean models in Section 1 (Eqs. (1)–(3)), but with different levels of variability. These simulations are modeled after 'Step 0' in [6]. To begin, let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ denote the general form of the linear models for all responses, where $\mathbf{y}$ is the vector of responses, $\mathbf{X}$ is the model matrix, $\boldsymbol{\beta}$ is the vector of model parameters, and $\boldsymbol{\epsilon}$ is the vector of random errors assumed to be normally and independently distributed with a common variance σ^2. Then, the least squares estimates of the coefficient parameters in $\boldsymbol{\beta}$ follow the multivariate normal distribution,

$\hat{\boldsymbol{\beta}} \sim MVN(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, and $\hat{\sigma}^2 = MSE = \frac{SSE}{(n-p)}$ is an unbiased estimator of σ^2, where p is the number of terms in the model.

To generate new response surfaces, we repeatedly simulate new $\hat{\boldsymbol{\beta}}^*$s from the multivariate normal distribution with mean $\hat{\boldsymbol{\beta}}$, as in the mean models (Eqs. (1)–(3)), but different choices for the variance. We opted to generate new model parameters centered around the estimated parameter values to simulate the increased or decreased variability in the responses, instead of working directly with new observations. This approach has the advantage of keeping the mean model the same for all of the different scenarios considered, which simplifies the comparison of results. Recall that the variability of $\hat{\beta}$ is a direct function of σ^2, which is estimated from the observations. Since our focus is on understanding the impact of changes in variability, this allows us to generate models of the response that differ by only the magnitude of the natural variability.

The detailed steps for the simulation are described as follows:

1. For each of the three responses, randomly generate $\hat{\boldsymbol{\beta}}_j^*$ from $MVN(\hat{\boldsymbol{\beta}}_j, \hat{\sigma}_j^2 d_j(\mathbf{X}_j'\mathbf{X}_j)^{-1})$, $j = 1, 2, 3$, where $d_j \in \{0.5, 1.5\}$ is a multiplying factor of the variance that determines the natural variability of the response variable.

2. For each response, approximate the response surface by estimating the response values using the simulated coefficients from Step 1 over a grid of points in the chosen operating region (Fig. 1(b)).

Eight combinations of multiplicative factors of the variance for the three responses are evaluated with $(d_1, d_2, d_3) \in$ {(0.5,0.5,0.5), (1.5,0.5,0.5), (0.5,1.5,0.5), (1.5,1.5,0.5), (0.5,0.5,1.5), (1.5,0.5,1.5), (0.5,1.5,1.5), (1.5,1.5,1.5)}. For each scenario, 500 simulated response surfaces are generated using the process described above. In the following sections, these results are compared to the original variability scenario with 500 simulations generated as $(d_1, d_2, d_3) = (1, 1, 1)$.

3. IMPACT ON THE PF

Based on the simulated response surfaces obtained following the procedure described in Section 2, this section describes the objective step [6] where the PFs made up of superior locations are identified and the range and variability of the PF choices among the simulations are evaluated. Thus, this step highlights the locations on the PFs as well as how frequently they appear on the PFs. In practice, non-contending locations that do not, or rarely, appear on the PFs should be eliminated from further consideration. Hence, this step is useful for identifying the input combinations, which give consistently superior performance for many combinations of simulated responses. In studying the different variability scenarios, interest lies in understanding how the proportion of times that a location is found on the PF changes.

Fig. 2 shows the locations that appear at least once on the PFs in the 500 simulations for the unchanged variability scenario (the top right panel) and the eight scenarios with different variability for individual responses (the top left and bottom right panels). The layout of all scenarios is illustrated in the bottom left panel of Fig. 2, which shows the geometry for a two-level, three-factor full factorial design. The three responses correspond to the three factors, and the low and high factor levels correspond to 0.5 and 1.5 times the original variability, respectively. The centroid of the cube corresponds to the original variability scenario, which is labeled with index number 0. The corners of the cube represent the eight combinations of different variability levels, which are labeled with index numbers 1–8. The four circle plots in the top left panel correspond to the corner points on the front face of the cube, which are simulated with 50% of the original variability for molecular weight, while the four circle plots in the bottom right panel correspond to the corner points on the back face of the cube, all of which have 150% of the original variability for molecular weight. Within each group of four subplots, the rows correspond to changes in the variability of viscosity response and the columns match changes in the variability of yield response.

In each circle plot, the size of the points is proportional to the frequency with which each location appears on the PFs in the 500 simulations. Hence, larger points appeared more often on the PFs. In the top right panel corresponding to the unchanged variability scenario, the points appearing more often on the PFs (at least 20% out of 500 simulations) form a bow tie–shaped region with larger points in the center and smaller points on the edge. The upper and lower regions of the bow tie shape correspond to the lightest gray (ideal) contour for the viscosity response in Fig. 1(a). Compared with Fig. 1, the points that appear on the PFs most often (more than 90% of the 500 simulations) mainly correspond to the four regions identified in the overlaid contour plots. This indicates that the locations with generally 'good' performance based on a wide variety of different priorities tend to appear on the PF with higher frequency.

The figures in the top left (with smaller variability for molecular weight) and bottom right (with larger variability for molecular weight) show patterns in the PF appearance frequency for locations similar to those from the original variability scenario. In addition to indicating the PF appearance frequency using the size of the points, the locations in the circular operating region are displayed in different

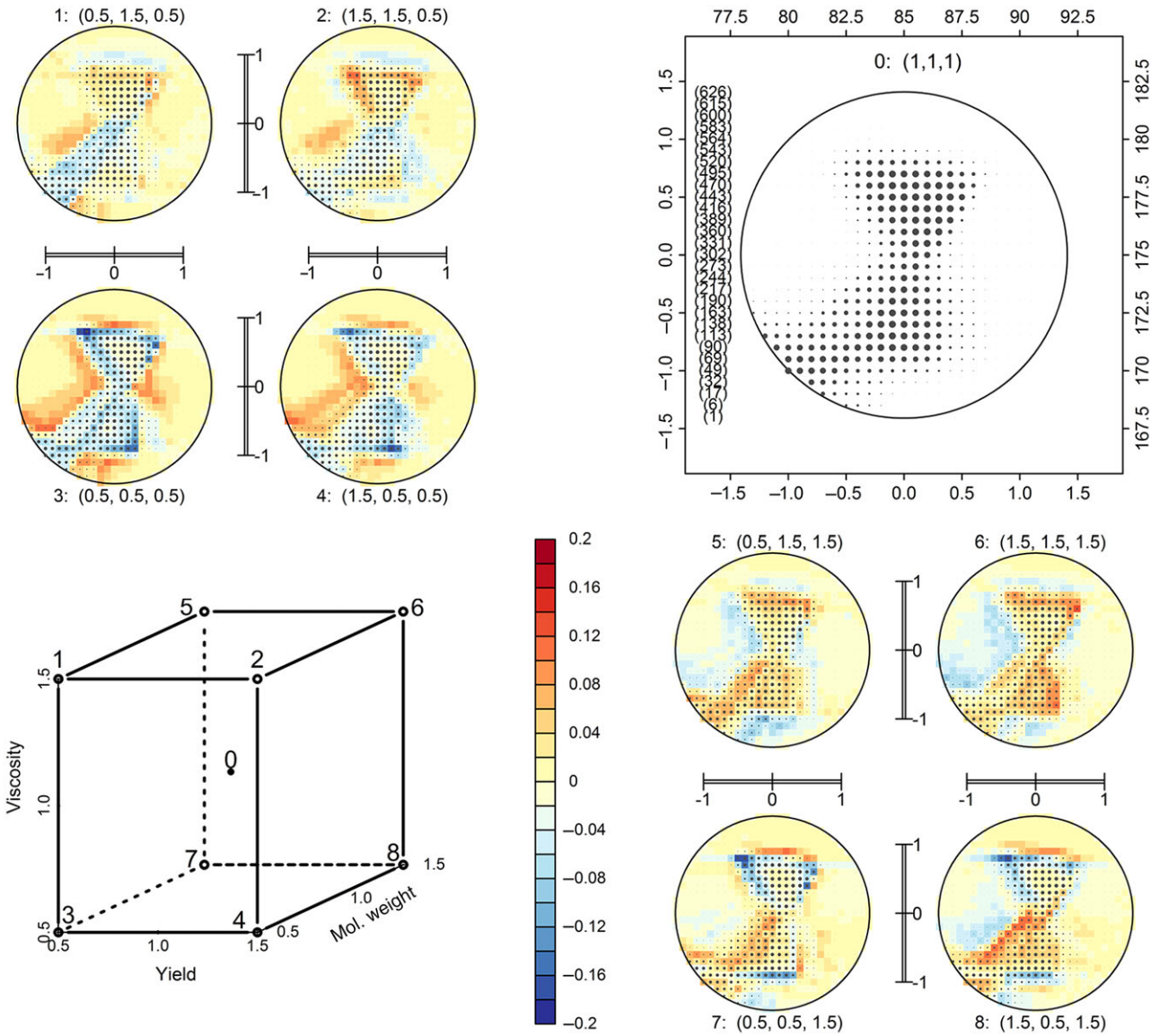

Fig. 2 (Bottom Left) A geometric representation of the two-level, three-factor full factorial simulation study design. (Remaining plots) Circle plots showing the possible input combinations, with the size of the points being proportional to the frequency with which the location appears on PFs in the 500 simulations for the original variability scenario (top right) and the variability scenarios (top left and bottom right). In each of these plots, the legend connects the plot to the geometric representation subplot from the bottom left panel, with both the cube index number and the 3-tuple indicating the MSE multiplier for each response variable. The color scale for the top left and bottom right panels illustrates how the frequency for that case differs from that for the original variability case, with locations shaded in blue appearing more frequently in that case than in the original case and locations shaded in red/orange appearing less frequently. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

colors to illustrate changes in their frequency relative to the original variability scenario. The warm colors (red/orange) represent decreases in how frequently the locations appear compared to the original variability scenario, with darker red colors corresponding to larger changes. The cool colors (blue) represent increased frequencies compared to the original variability scenario, with darker blue corresponding to higher frequencies than those in the original case.

We can observe several patterns in Fig. 2. First, the larger changes (corresponding to darker colors) tend to occur more often around the edges of the bow tie–shaped PF region. Second, reducing the variability of molecular weight tends to increase the PF appearance frequency for points located around the bottom left region of the operating region (corresponding to the green and pink regions in Fig. 1), which is evidenced by the large amount of blue coloring

in these regions in the figures in the top left panel. On the other hand, increasing the variability of molecular weight tends to reduce the PF appearance frequency for points in these regions, based on the red and orange colors in all figures in the bottom right panel. The points in the green region in Fig. 1 have the best performance for molecular weight on average, but relatively poor performance for the other two responses. Hence, reducing the variability of molecular weight tends to favor these locations more often. While the locations in the pink region of Fig. 1 tend to balance the performance of the three responses, molecular weight has the smallest signal-to-noise ratio, and reducing its associated variability seems to have the most impact on the PF appearance frequency of the locations in this region.

Third, reducing the variability of viscosity tends to result in increased PF appearance frequency for points in the orange region of Fig. 1 because these locations correspond to having the best performance on viscosity. Fourth, reducing the variability of yield tends to be associated with a slightly increased PF appearance frequency for points in the blue region in Fig. 1, which corresponds to locations with the best performance on yield and poor performance on molecular weight. However, the degree of change is generally smaller than changing the variability of either molecular weight or viscosity by the same amount, since yield has the largest signal-to-noise ratio and hence benefits the least from reducing its variability. Lastly, reducing the variability for at least two of the three responses is generally associated with larger increases in the PF appearance frequency for locations in the pink region of Fig. 1 with more balanced performance across all the responses.

These observations at the objective stage show that the impact of the degree of response variability depends on both the signal-to-noise ratios of the responses and the inherent trade-offs between the responses. Hence, it is helpful to understand this impact for a particular application before making any decisions that may impact a process' associated uncertainty.

4. IMPACT ON PROMISING SOLUTIONS IDENTIFIED IN THE SUBJECTIVE STAGE

The locations identified to have higher PF appearance frequencies provide us an objective set of promising choices. In Stage 2 of the decision-making process [6], these locations are investigated further after a series of subjective choices are made by the user. These choices include how to scale the responses, how to combine the scaled criteria into a single metric, and how the user's priorities can be expressed as relative weights on the different criteria. Here, we focus on the early part of Stage 2, rather than making an actual decision, to investigate the overall impact of response variability for a specific set of user choices. In practice, if the user is uncertain about any of these choices, a sensitivity analysis should be performed.

For each of the 2^3 variability scenarios, we scale each criterion using bounds obtained from 95% prediction intervals for each response. This allows a unified scaling scheme across the large number of simulations. For each of the linear models estimated in Section 1 (Eqs. (1)–(3)), a 95% two-sided prediction interval for any of the responses at a specific input location $\boldsymbol{x_0}$, with $\boldsymbol{x_0'} = (x_{01}, x_{02}, \ldots, x_{0p})$ in the model form, is given by $\boldsymbol{x_0'}\hat{\boldsymbol{\beta}} \pm t_{0.975,n-p}\sqrt{d * MSE(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}$ ([7], p. 35), where $d \in \{0.5, 1, 1.5\}$ is the variance multiplier for a particular response variable. Then, the best (mapped to 1) and the worst (mapped to 0) possible values used for scaling are identified by summarizing these prediction bounds over all grid points in Fig. 1(b). For yield (to be maximized), the best possible value is defined to be the largest upper bound resulting from the prediction intervals across the entire input space and the worst possible value is defined to be the smallest lower bound. For molecular weight (to be minimized), we use the smallest lower bound and the largest upper bound across the entire input space as the best and the worst possible values, respectively. For viscosity (to hit a target at 65), we use 0 as the best possible value for that criterion as many prediction intervals include the target, and the worst value is chosen by considering all the upper and lower prediction bounds across the entire input space and using the one furthest from 65 to compute the largest distance from the target. It is theoretically possible that a simulated response is outside the range specified by the best and the worst possible values from the prediction intervals. If a simulated response exceeds the best possible value, the scaled criterion is set to 1. Similarly, a simulated response below the worst possible value is set to 0.

To be consistent with previous results ([2,6,7]), we combine the three scaled responses using a multiplicative DF of the form

$$\mathrm{DF}_{\mathrm{mult}}(l, \mathbf{w}) = C_1(l)^{w_1} \cdot C_2(l)^{w_2} \cdot C_3(l)^{w_3}, \qquad (4)$$

where $C_j(l)$ is the scaled value for response j ($j = 1, 2, 3$) at location l on the grid in Fig. 1(b), and $\mathbf{w} = (w_1, w_2, w_3)$ is the relative weighting of the three responses under consideration, with $\sum_{j=1}^{3} w_j = 1$.

After choosing the scaling and DF form, we allow some flexibility in specifying the weighting preference, since making a single specific choice is difficult, especially when multiple decision makers are involved. To do this, we consider all possible weightings of the three scaled responses by examining a space of 20,301 weight combinations, with each entry of the weight vector $\mathbf{w}$

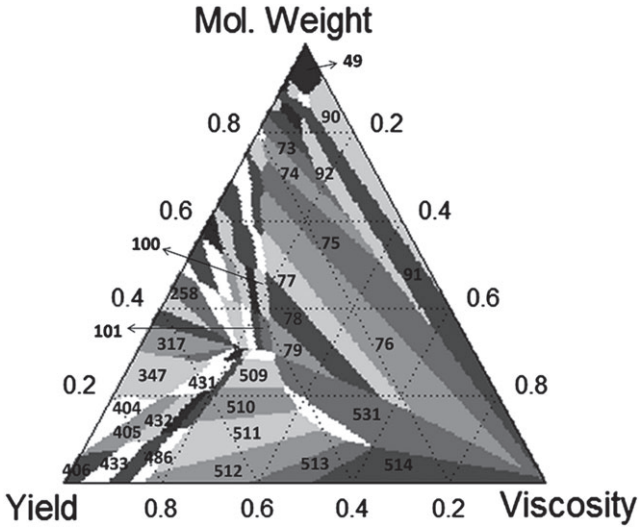

Fig. 3 Mixture plot for designs with best desirability for different weighting choices based on the mean model response surface when multiplicative desirability is used and the criteria values are scaled based on 95% prediction intervals for the responses across the locations in the design space.

being a multiple of 1/200. For each weight combination, we can find the location that optimizes the desirability among all candidate locations and then summarize the best solution for the different weight choices in a mixture plot ([3,9]).

Fig. 3 shows the mixture plot for optimizing the mean model for the original variability scenario. Every point in the triangle corresponds to a weight combination. The vertices and the edges correspond to optimizing based on a single criterion and two of the three criteria, respectively. Adjacent weight combinations corresponding to the same optimal location are displayed in the same gray scale. The area of a region shows the proportion of weight combinations for which the corresponding location has the best desirability. Hence, a larger area corresponds to a more robust location that is best for more weight combinations. For example, Locations 75 and 76 (both in the pink region of Fig. 1) are each best for about 17% of the weights and account for the largest areas in Fig. 3. Location 531 (in the orange region of Fig. 1) is optimal for about 12% of the weights and corresponds to the third largest area. These represent some of the more robust locations with optimal performance for different weighting priorities.

We note that the visualization of the optimal choices gets more complicated for more than three responses. Lu and Anderson-Cook [10] adapt the mixture plot for applications considering four criteria simultaneously. To handle cases with more than four responses, other devices such as tabular lists or dynamic graphics would be necessary to highlight top choices in a more effective manner. In general, we caution against trying to optimize over too large a number of responses simultaneously, as this often leads to mediocre values for many of the responses. It is recommended to spend additional time in the early stages of the optimization process to make strategic decisions about the most important priorities.

Since the PF changes across simulations for each scenario, the locations with the best desirability at each weight combination, and thus a location's optimal area, change across the simulations as well. Hence, we quantify the average robustness of locations by calculating their average area across the 500 simulations, which is summarized in Fig. 4. While Fig. 2 illustrates the objective result of how often each location in the input space appears on the PF, Fig. 4 focuses on a subjective aspect of the fraction of prioritization weightings for which that location can be considered best. In each of the circle plots, the size of the points is proportional to the PF appearance frequency and the shades of yellow-green-blue represent small-to-large average areas for each location. The maximum average area across all eight scenarios is only about 3.5%, which is quite small. This is a result of several factors. One factor is that there may be quite a few locations that are optimal for only small proportions of weights and hence have small optimal areas. Another factor is each location's PF appearance frequency; if, for a single simulation, a location is either not on the PF or does not have the best desirability for any weight combinations, then the area contributed to the average for that simulation is zero.

In the original variability scenario (top right panel of Fig. 4), the locations with the largest average area ($\geq 2.1\%$) are in the interior of the orange and pink regions of Fig. 1 (corresponding to the lightest gray contour for ideal viscosity). The more robust locations, with average area no less than 0.7%, are all in the four highlighted regions from Fig. 1. As we reduce the variability of molecular weight (top left panel of Fig. 4), there are more locations with higher average areas in the green region of Fig. 1, which favors the molecular weight more than other two responses. When the variability of viscosity is reduced ($d_2 = 0.5$, in the top left and bottom right panels of Fig. 4), more locations in both the orange and pink regions (best for viscosity) have larger average areas. Changing the yield's variability results in the least obvious change in the average areas due to its large signal-to-noise ratio. Among the eight variability scenarios, the largest average areas ($\geq 2.1\%$) occur when the variability associated with viscosity is decreased, as evidenced by more locations being colored with shades of dark green and blue.

By introducing the subjective choices and evaluating the robustness of locations to varied weighting choices while taking into account the uncertainty of the responses, a summary like Fig. 4 is helpful for reducing the contending

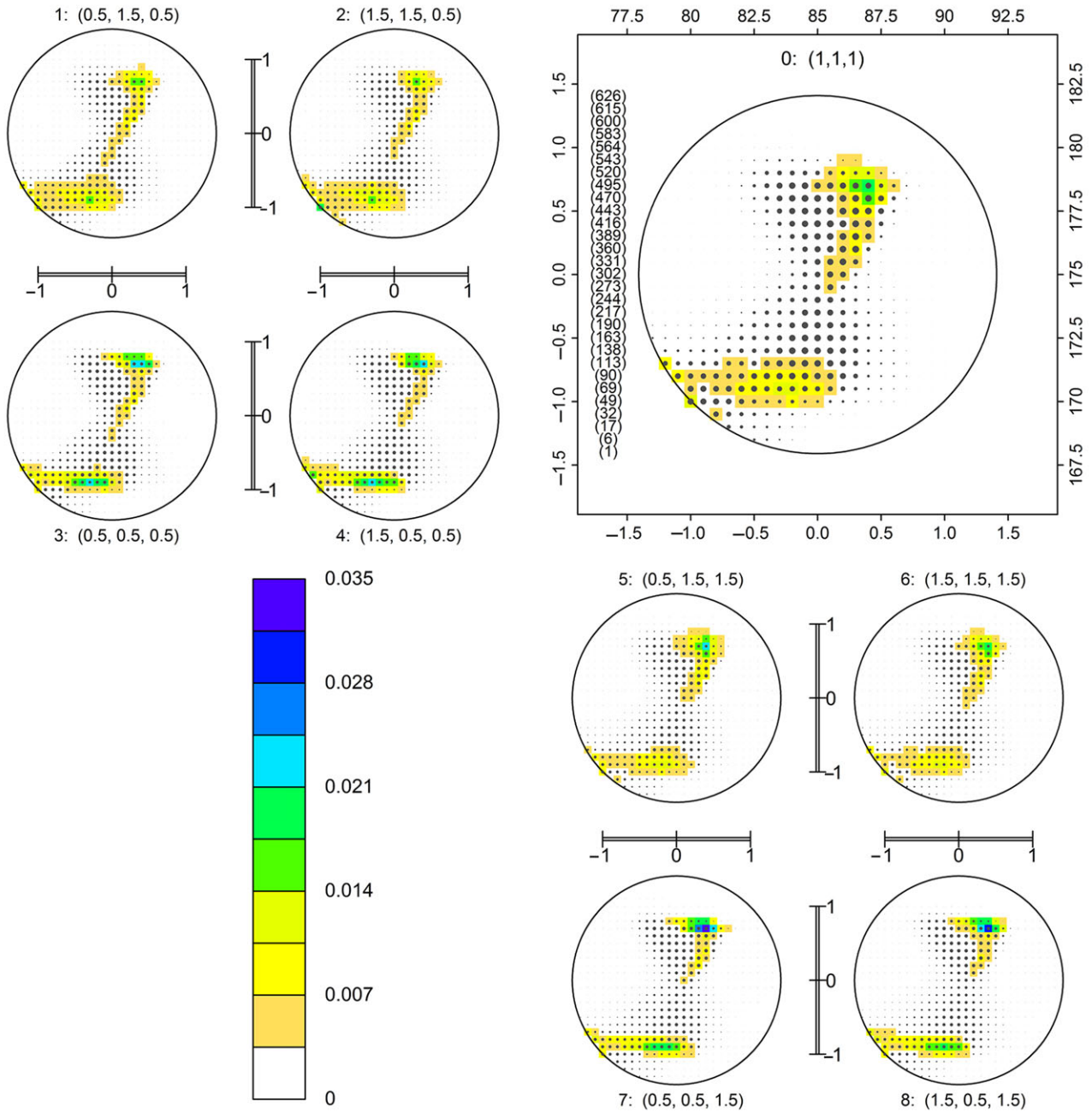

Fig. 4 Circle plots, with the size of the points being proportional to the frequency with which the location appears on PFs in the 500 simulations, that examine the average (across the 500 simulations) mixture area, with shades of green and blue representing larger average areas. In each of these plots, the legend connects the plot to the geometric representation from Fig. 2 (bottom left panel), with both the cube index number and the 3-tuple indicating the MSE multiplier for each response variable. (Top Right) Circle plot for the original variability scenario. (Top Left and Bottom Right) Circle plots for the 2^3 variability scenarios. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

solutions to an even smaller subset of the most promising choices with the greatest robustness across a large number of simulations. The choice of a location is ultimately dependent on how much we value the different responses as captured by their different weightings. As noted in Section 3, the impact of variability of the responses on the robustness of locations is affected by the intrinsic trade-offs between the responses as well as the degree of variability relative to the range of response values across the desired operating region.

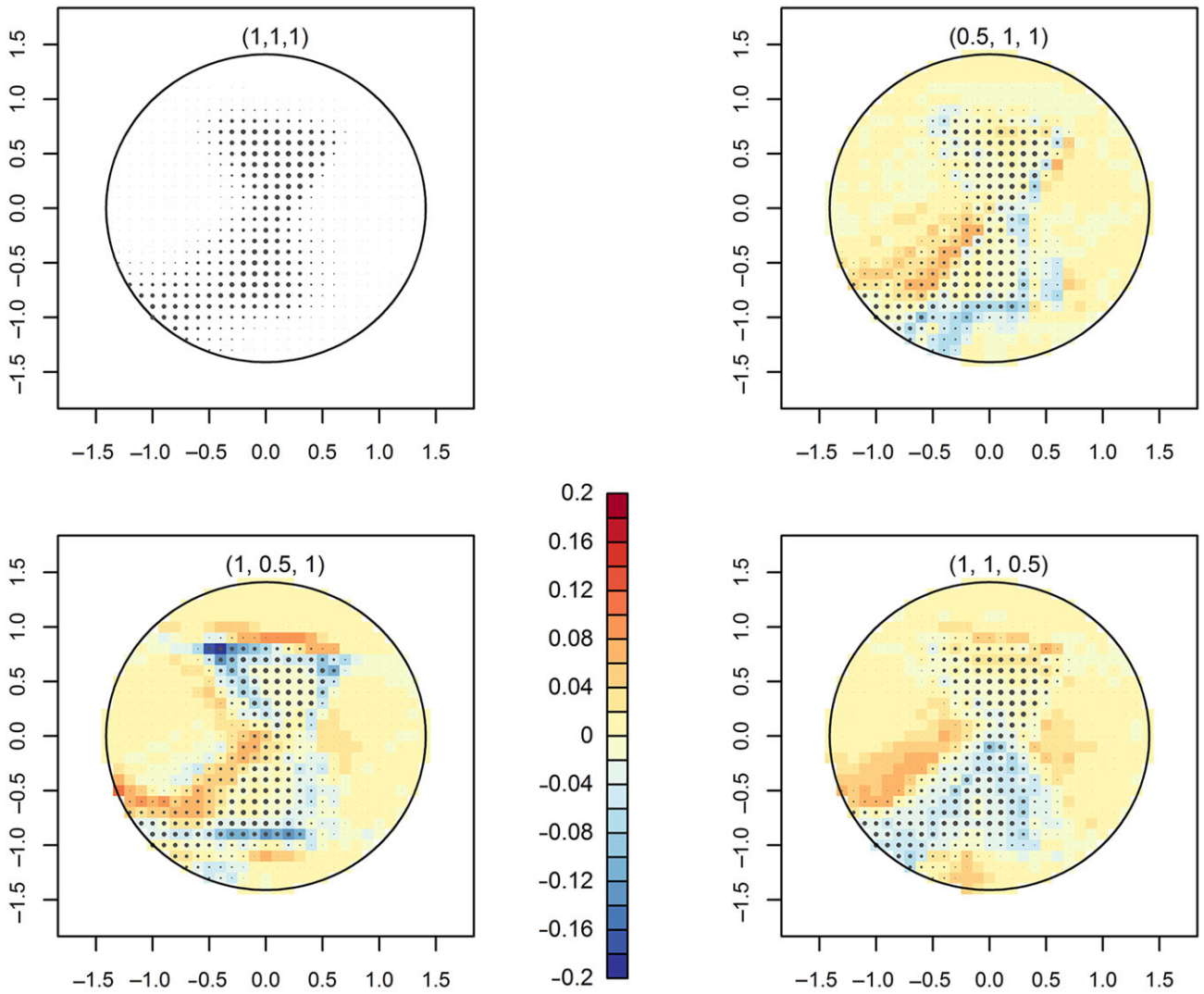

Fig. 5 Circle plots showing the possible input combinations, with the size of the points being proportional to the frequency with which the location appears on PFs in the 500 simulations. In each of these plots, the legend displays the 3-tuple indicating the MSE multiplier for each response variable. (Top Left) Circle plot for the original variability scenario. (Remaining Panels) Circle plots for the variance reduction cases (Top Right): decreasing yield's variability by 50%; (Bottom Left): decreasing viscosity's variability by 50%; (Bottom Right): decreasing molecular weight's variability by 50%. The color scale illustrates how the frequency for each case differs from that for the original variability case, with blues indicating an increase in frequency and reds a reduction in frequency relative to the original case. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Given the identified PF with the objective set of promising solutions, there are different ways to incorporate user priorities for approaching a final solution. One possibility is to use a full Bayesian approach to quantify the user priorities through a probability distribution and guide the subjective decision-making process in a more rigorous manner. This would allow optimization, not just on a single combination of weights for prioritization of the response values but would also allow a region to be identified, and performance across this region to be considered. This examination of multiple combinations of weights, each with a potentially different relative emphasis, could provide a posterior distribution of candidate locations for evaluation. This method may be suitable for scenarios when there is a general agreement on the primary region of weighting preference, which can be efficiently captured in a single probability distribution. However, it would not be suitable for scenarios when there may be disagreements in the weighting preferences among different subject matter experts. In this case, there could be multiple regions to explore before a consensus could be achieved. Another alternative to the full Bayesian approach for

handling more focused users' priorities by using probability distributions was explored in Lu, Anderson-Cook, and Lin [11].

5. PROCESS IMPROVEMENT THROUGH VARIANCE REDUCTION

Process improvement can often be achieved through variance reduction [12]. For our example, we assume that the cost of reducing the variability of each response is the same, and given an assumed budget, we look at the impact of a 50% reduction in the variance of one of the responses. We note that it would be straightforward to consider any total cost and relative cost structure, by defining alternatives that are within budget and comparable based on the cost of implementation. An advantage of our simulation-based approach is that different alternatives can be compared without having to actually manipulate the system, and thus it is possible to first gain an understanding about how the changes will impact results before those changes are made.

Fig. 5 shows the circle plot for the original variability scenario (top left panel) as well as the three alternatives under consideration. As with Fig. 2, the size of the points is proportional to the frequency with which different locations in the operating space occur on PFs in the 500 simulations. The color scale indicates changes in the PF appearance frequency with blues denoting an increase in frequency and reds denoting a reduction in frequency relative to the original variability scenario. When the variability of yield is reduced by 50% (top right in Fig. 5), many of the frequencies for locations in the operating space stay relatively consistent with the original variability case, generally changing by less than 8%. The larger increases in frequencies occur close to the blue and pink regions from Fig. 1, where yield has better performance. Reducing the variability of either viscosity or molecular weight results in larger changes in the PF appearance frequencies. When the variability of viscosity is reduced, locations close to the desired target viscosity (the lightest contour in Fig. 1(a)) are on the PF much more frequently. This 'tightening up' of the PF with locations that appear a moderate number of times on the front being reduced, and more commonly appearing locations increasing their frequency, shows the edges of the PF locations more clearly. Although not as dramatic, we see a similar pattern for the reduction of molecular weight's variability. Commonly selected regions are on the PF even more frequently, and rarely chosen locations are selected less frequently. For molecular weight, the change in variability tends to lead to more locations in the green and pink regions of Fig. 1 appearing on PFs with increased frequency.

In terms of choosing between the gains from any of the variance reduction process improvements, investing in yield seems to have the least clearly defined benefits. This response already has the best signal-to-noise ratio, and further improving that ratio does not substantially change the frequency with which locations are selected for the PF. Instead, selecting variability reduction for either viscosity or molecular weight gives more substantive gains in clearly identifying the preferred locations on the PF. The choice of the more desirable one among these two for the practitioner should be based on the regions of greater emphasis in terms of preferred weighting of the responses.

In addition to appearing frequently on the PFs across the simulations, the preferred location (input factor levels) for operating this process should 1) perform well for a range of weights believed to be the right prioritization for the three responses and 2) perform well across a larger range of weights close to the ideal combination. This second requirement is helpful because it is often difficult to precisely specify the preferred weight combination, particularly if several decision makers are involved. By identifying the solutions that are robust across the simulations, decision makers can identify locations that are most frequently the best for their weighting preferences. Fig. 6 examines the robustness of the locations over a range of desirability weights, using the same DF and scaling as in Section 4. The color scale in Fig. 6 shows the average area across all 500 simulations. Similar to the distinction between Figs. 2 and 4 earlier, Fig. 6 focuses on the subjective aspect of how robust a given input location is across different user weightings of the responses, while Fig. 5 was focused on the objective phase and quantified the frequency with which a location was selected on the PF. When we examine the benefits of process improvement by investing in variation reduction, we again see relatively little benefit when yield is improved (the top left and right panels of Fig. 6 look similar). There are some minor changes in the average size of areas for a few locations in the orange and pink regions of Fig. 1, but in the blue region, where yield performs best, there is relatively little change.

In contrast, reducing the variability of viscosity (bottom left panel of Fig. 6) increases the robustness over different weight combinations for locations in the pink region in Fig. 1, where the responses have the most balanced performance, and in the orange region, where yield and viscosity perform better than molecular weight. When the variability of molecular weight is reduced (bottom right panel of Fig. 6), there are also some gains in robustness for locations in both the pink and orange regions, although perhaps less dramatic than when the improvement is made in viscosity. Again, it appears that there is the least gained by improving yield, and consideration of either of the other two responses

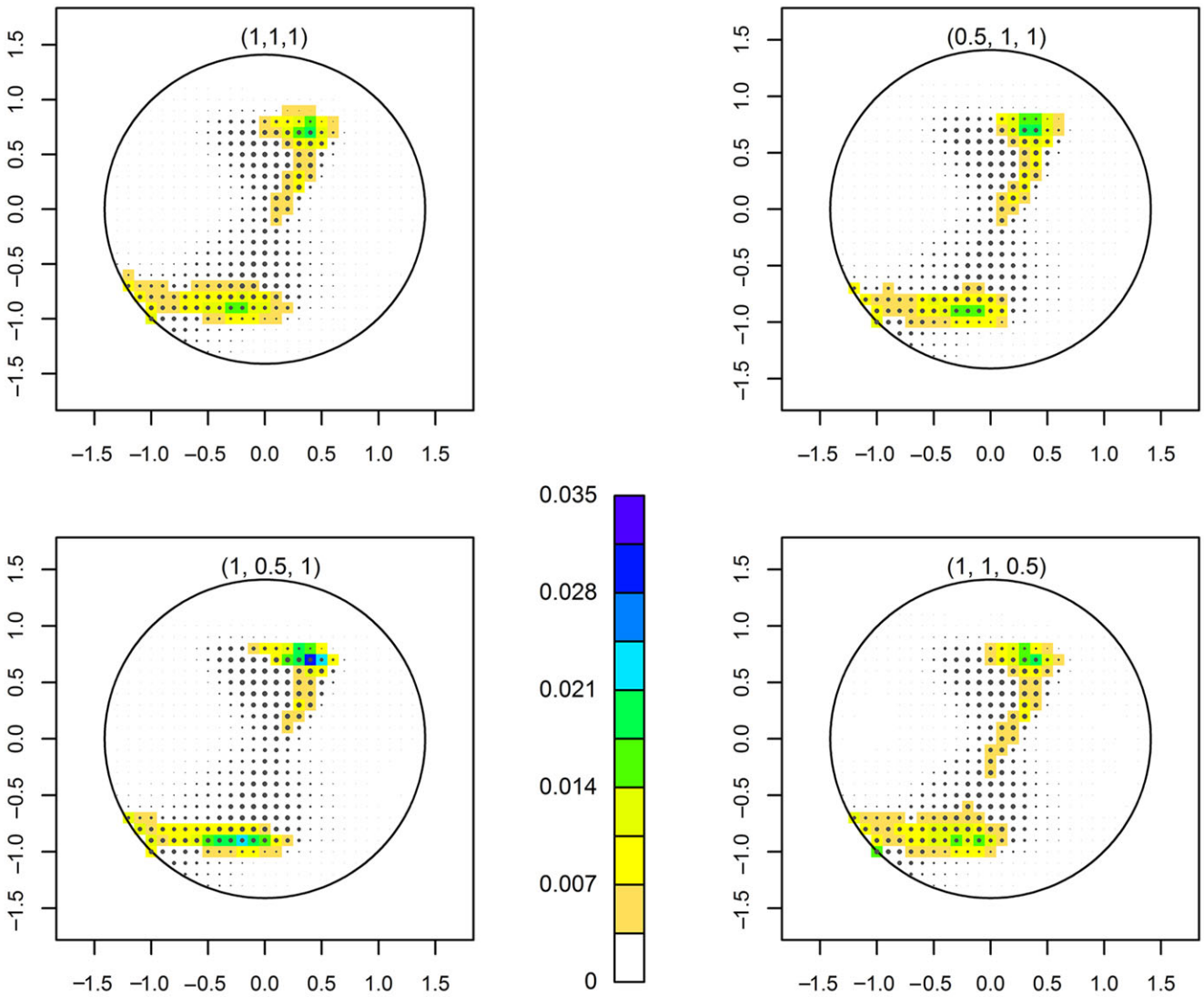

Fig. 6 Circle plots, with the size of the points being proportional to the frequency with which the location appears on PFs in the 500 simulations, that examine the average (across the 500 simulations) mixture area, with shades of green and blue representing larger average areas. In each of these plots, the legend displays the 3-tuple indicating the MSE multiplier for each response variable. (Top Left) Circle plot for the original variability scenario. (Remaining Panels) Circle plots for the variance reduction cases (Top Right): decreasing yield's variability by 50%; (Bottom Left): decreasing viscosity's variability by 50%; (Bottom Right): decreasing molecular weight's variability by 50%. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

should be chosen based on the user's preference of weighting.

Finally, Fig. 7 shows a trade-off plot with uncertainty bounds [6] for the unchanged variability scenario, and Fig. 8 displays similar trade-off plots for the three variability reduction scenarios. These plots show the range of values for each criterion, with the associated desirability scaling used. The scaling for the reduced variability scenarios changes slightly across the various plots. The locations highlighted in these figures were chosen from across the four highlighted regions of Fig. 1. The first four locations (113, 49, 90, and 93) are located in the green region where yield performs most poorly. The next four locations (77, 79, 100, and 125) are found in the pink region where there is balanced performance among the three responses. The next four locations (531, 535, 511, and 486) are from the orange region where yield and viscosity perform well, and the last four locations (460, 432, 377, and 405) are from the blue region where yield has its best values. Examination of the trade-off plots (Fig. 8 vs. Fig. 7) shows that the 50% reduction in variance does result in some reduction in the range of values anticipated (changes in the y

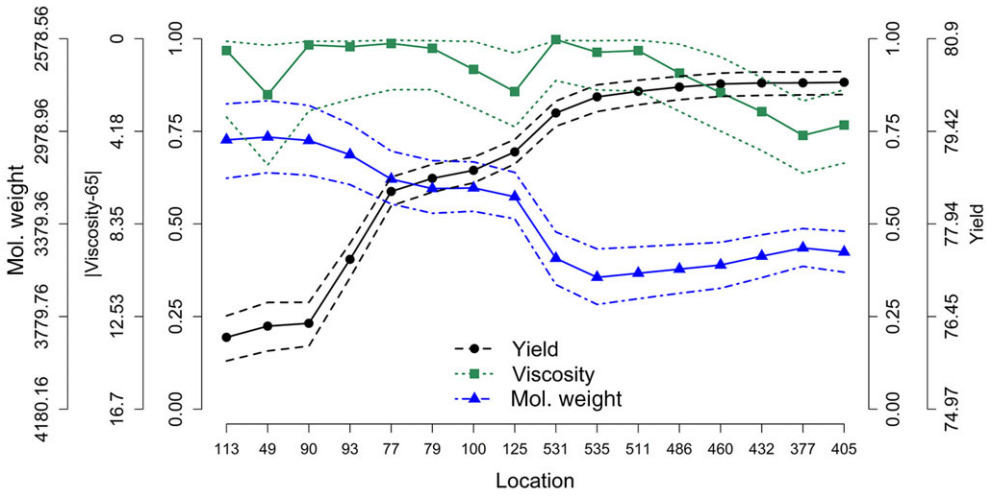

Fig. 7 Trade-off plots with uncertainty bands for select locations for the simulation for the original variability scenario. The locations are ranked from worst to best performance on the yield criterion. Uncertainty bands are based on the 5% and 95% percentiles of the simulated responses for each location. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

axes), with slightly narrower uncertainty bounds for the response with the altered variance. The trade-off plot can be easily adapted for applications with additional responses (incorporating additional curves for each new response), although we again caution against increasing the total number of objectives of the optimization as it may dilute each response's contribution and lead to a very mediocre overall solution.

Overall, it can be helpful to use the simulation study to examine how potential improvements to the process through variability reduction impact the user's ability to confidently select a preferred location to optimize a combination of the responses. In this example, we have illustrated how reducing the variability of the response with the largest signal-to-noise ratio is expected to result in the least change in the PF compared to improving either of the other two responses.

6. CONCLUSIONS AND DISCUSSION

PF optimization allows the user to identify promising locations in the operating space that perform well based on the estimated models for the responses of interest. Examining just the mean model for each of the responses (e.g. Eqs. (1)–(3)) for constructing the PF and for making decisions about the preferred location ignores the uncertainty associated with model estimation. Taking into account this uncertainty through simulations helps give a deeper understanding of the performance that can be expected when a selected optimal location is implemented.

In this paper, we examine the relationship between the variability of the responses and how the PF changes. There are multiple interdependent aspects that influence the changes. The signal-to-noise ratio, which considers the range of response across the operating space relative to the natural variability, plays an important role. The results associated with the responses with a small signal-to-noise ratio are more likely to change than those for the responses where the signal dominates natural variability. However, the relative priorities of the user are also important for determining if a change in the variability of different responses is likely to be influential. Both the objective first step of the PF optimization process, which selects the front, and the initial subjective stages of incorporating a DF summary of the combined responses show changes depending on how the size of the natural variability changes.

By using simulations to examine variability in the estimated response surfaces, we can have a more realistic quantification of its impact on the identified PF as well as the final decision. The outlined approach paired with graphical tools is helpful for guiding a structured process for implementing the method. When process improvement is considered, comparing alternatives with similar associated costs to evaluate how the PF changes can provide a cost-effective way of determining which response should be the focus of variation reduction. Considering both which responses are valued most highly and which have the smallest signal-to-noise ratio can suggest where to invest in process improvement.

While the richness of the summaries is helpful for facilitating informed decision-making, the computational demand increases quickly as the number of input factors and/or the number of responses increase. The identification of the PF can be adapted for higher dimensional cases in a straightforward way. However, visualization of the different

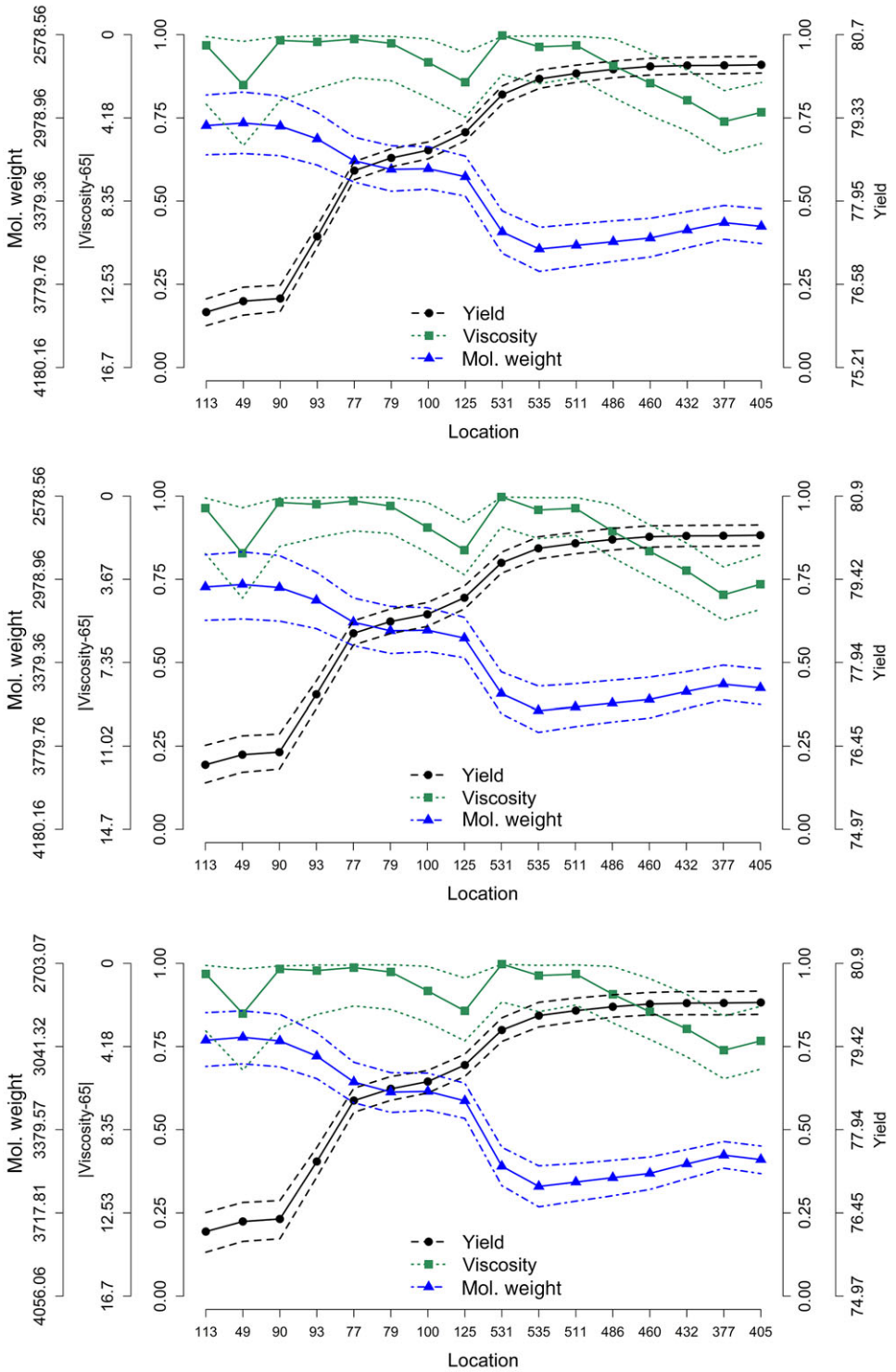

Fig. 8 Trade-off plots with uncertainty bands for select locations for the (top) case where yield's variability is reduced by 50%, (middle) case where viscosity's variability is reduced by 50%, and (bottom) case where molecular weight's variability is reduced by 50%. The locations are ranked from worst to best performance on the yield criterion. Uncertainty bands are based on the 5% and 95% percentiles of the simulated responses for each location in each of the three variability reduction scenarios. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

solutions can get complicated with more inputs and responses. In general, we suggest that the decision maker think carefully about expanding the number of responses over which to optimize as this typically leads to increasingly severe trade-offs between choices and can also lead to the overall mediocre performance on many of the responses.

If the dimension of the input space increases, then some of the visualization tools described in the paper become more problematic. As the number of factors increases, the size of the grid over which to search can increase dramatically. To speed up the optimization process, it may be helpful to define an initial sparse grid that allows full exploration of the space, but without overburdening the required computational time. Once promising regions of the design space have been identified, then a finer grid of input combinations can be explored in those focused regions of the design space. Displays of the results from the study of variation can be adapted to tables, which show promising input factor combinations along with the response values and the frequency with which they appear on the PF. In addition, there may be some factors over which the responses are somewhat robust, allowing some reduction in the dimension of the space to be displayed. Finally, an additional dimension of display can be added by taking slices of the response space (fixing one or more factor at a series of values and showing a set of plots across the slices).

A key message of the PF optimization approach for responses is that incorporating the associated uncertainty of the parameter estimates can lead to better understanding of the results that are possible across the different locations in the design space, and hence lead to a more realistic decision in practical applications.

REFERENCES

[1] G. Derringer and R. Suich, Simultaneous optimization of several response variables, J Qual Technol 12 (1980), 214–219.

[2] J. L. Chapman, L. Lu, and C. M. Anderson-Cook, Process optimization for multiple responses utilizing the pareto front approach, Qual Eng 26 (2014), 253–268.

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

[4] L. Lu and C. M. Anderson-Cook, Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature, Qual Eng 24 (2012), 404–422.

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

[6] J. L. Chapman, L. Lu, C. M. Anderson-Cook, Incorporating response variability and estimation uncertainty into pareto front optimization, *Comput Ind Eng* 76 (2014), 253–267.

[7] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, Response Surface Methodology, 3rd ed., Hoboken, NJ, John Wiley & Sons, 2009.

[8] L. Lu and C. M. Anderson-Cook, Adapting the hypervolume quality indicator to quantify trade-offs and search efficiency for multiple criteria decision-making using pareto fronts, Qual Reliab Eng Int 29 (2013), 1117–1133.

[9] J. Cornell, Experiments with Mixtures: Design, Models, and the Analysis of Mixture Data, 3rd ed., New York, John Wiley & Sons, 2002.

[10] L. Lu and C. M. Anderson-Cook, Balancing multiple criteria incorporating cost using pareto front optimization for split-plot designed experiments, Qual Reliab Eng Int 30 (2014), 37–55.

[11] L. Lu, C. M. Anderson-Cook, and D. Lin, Optimal designed experiments using a pareto front search for more focused desirability function weights, Comput Stats Data Anal 71 (2013), 1178–1192.

[12] S. H. Steiner and R. J. Mackay, Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes, Milwaukee, WI, American Society for Quality, 2005.