Computers & Industrial Engineering 76 (2014) 253–267

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Incorporating response variability and estimation uncertainty into Pareto front optimization

Jessica L. Chapman [a,*], Lu Lu [b], Christine M. Anderson-Cook [c]

[a] Department of Mathematics, Computer Science, and Statistics, St. Lawrence University, United States
[b] Department of Mathematics and Statistics, University of South Florida, United States
[c] Statistical Science Group, Los Alamos National Laboratory, United States

ARTICLE INFO

Article history:
Received 21 November 2013
Received in revised form 28 July 2014
Accepted 30 July 2014
Available online 7 August 2014

Keywords:
Response surface
Multiple response optimization
Incorporating estimation uncertainty
Trade-offs
Graphical summaries

ABSTRACT

Pareto front optimization has been commonly used for balancing trade-offs between different estimated responses. Using maximum likelihood or least squares point estimates or the worst case confidence bound values of the response surface, it is straightforward to find preferred locations in the input factor space that simultaneously perform well for the various responses. A new approach is proposed that directly incorporates model parameter estimation uncertainty into the Pareto front optimization. This step-by-step approach provides more realistic information about variability in the estimated Pareto front and how it affects our decisions about the potential best input factor locations. The method is illustrated with a manufacturing example involving three responses and two input factors.

1. Introduction

When optimizing several estimated responses, the two-stage Pareto front approach (Lu, Anderson-Cook, & Robinson, 2011) to identify promising candidate solutions and then select a best match to user priorities can add structure and rigor to decision-making. Traditionally, Pareto front (PF) optimization approaches involving multiple estimated responses have focused on the maximum likelihood or least squares point estimates of the response at a given set of inputs (referred to as the "mean model" throughout this paper). However, uncertainty in the model parameter estimates can have an impact on which input factor combinations are identified as best. Since the responses are likely to have different natural variability in the operational space, the precision with which the parameters are estimated differs, making it difficult to anticipate effects on the Pareto front solutions. Naively treating the estimated response surfaces as fixed can lead to overconfidence in the conclusions and potentially sub-optimal input factor level choices which do not perform well when implemented in practice.

Costa, Espirito Santo, and Oliveira (2011) and Mattson and Messac (2005) propose visualization approaches to understand how uncertainty impacts the construction of the PF. Martins and Lambe (2013) provide a survey of design optimization architectures, while Yao, Chen, Luo, van Tooren, and Guo (2011) review strategies for uncertainty-based optimization. Hu and Youn (2011a, 2011b), Wei, Cui, and Chen (2008) and Chowdhury, Rao, and Prasad (2009) consider strategies for summarizing the impacts of uncertainty on complex systems and their reliability.

Chapman, Lu, and Anderson-Cook (2014) propose using the worst case bounds of prediction intervals as a simple way of incorporating uncertainty into the decision-making process. In this paper, we propose an alternative approach for quantifying and characterizing the impact of estimation uncertainty on solution selection. The uncertainty impacts both which solutions are located on the PF, as well as which solutions are best for the particular priorities of the study as measured by a desirability function with user-specified weightings of the different criteria.

To illustrate the proposed methodology, we consider the optimization of a chemical process described in Myers, Montgomery, and Anderson-Cook (2009), [p. 253] where three responses (y_1 = yield, y_2 = viscosity, y_3 = number-average molecular weight) are of interest. Two input variables (time, $\xi_1 \in [77\text{ min}, 93\text{ min}]$ and temperature, $\xi_2 \in [167\,F, 183\,F]$) can be adjusted to influence the responses. To estimate the relationships between inputs and responses, a 13-run central composite design (Myers et al., 2009, p. 297) for a circular coded region with maximum radius of $\sqrt{2}$ was run and data were collected for each response. After fitting quadratic response surface models and removing non-significant terms, the estimated mean models for each response are as follows:

* Corresponding author.

http://dx.doi.org/10.1016/j.cie.2014.07.028
0360-8352/

$$\widehat{y_1} = 79.94 + 0.995x_1 + 0.52x_2 + 0.25x_1x_2 - 1.38x_1^2 - 1.00x_2^2$$

$$\widehat{y_2} = 70.0 - 0.16x_1 - 0.95x_2 - 1.25x_1x_2 - 0.69x_1^2 - 6.69x_2^2$$

$$\widehat{y_3} = 3386.2 + 205.1x_1 + 177.4x_2$$

The goal of the optimization is to simultaneously maximize yield, y_1, and minimize both the molecular weight, y_3, and the distance from the viscosity to a target value of 65, $|y_2 - 65|$. The ideal solution is a combination of time and temperature that performs well for all three objectives. Since the three criteria cannot all achieve their optimum simultaneously, which location is selected depends on the relative importance that is placed on the different responses' performance.

Because there are objective and subjective aspects to selecting a best solution to an optimization problem, the PF approach in Lu et al. (2011) considers decision-making in distinct stages. Stage 1 is objective, since it removes all poor candidates that are strictly inferior to others. A solution is inferior if at least one solution exists that has all criterion values at least as good as the inferior solution and at least one that is strictly better. Eliminating these inferior choices is rational and simplifies subsequent steps by removing non-contenders from further consideration. The PF is comprised of all non-inferior solutions. Stage 2 is subjective as it considers how important good performance on the different criteria is to the decision-maker. It examines solutions on the PF and determines how well they match the priorities of the study. Clearly there are different ways to consider the subjective aspect of the decision-making. Our approach is to quantify the desirability of different options subject to different priorities, and then provide methods to explore the robustness of the solutions to changes in priorities. Graphical summaries of the different alternatives and how they compare can help guide the selection of which individual solution best suits the needs of the decision-maker and facilitate discussion with quantitative measures if several stakeholders have different priorities for the solution. Methods are adapted from Lu and Anderson-Cook (2012) and Lu, Anderson-Cook, and Robinson (2012).

The process for selecting a best overall solution is further complicated when the estimated responses have associated uncertainty, which suggests a range of plausible values for the model parameters that are consistent with the data observed. To capture this uncertainty, we use the estimated models to simulate a large number of response surfaces all consistent with the observed data. This collection of alternative solutions becomes the basis for examining the impact of estimation uncertainty on our conclusions. The overall goal of the selection process is to highlight a small number of combinations of input factor levels that give optimal performance for the responses of interest, subject to how we have chosen to prioritize them. To help with the discussion of the subjective Stage 2 when uncertainty is present, we have broken this stage into several sub-steps (2a–2c), each with distinct goals and customized graphical summaries. We now provide an overview of the different steps in the decision-making process, before illustrating the methods in detail with the example.

Step 0: Generate Alternate Response Surfaces Consistent with Data: The goal of this step is to generate a large number of sets of model parameter values that are consistent with the observed data. These values can then be used to obtain response surfaces representative of the plausible relationship between the inputs and the responses. These response surfaces serve as the basis for our understanding of the impact of estimation uncertainty. The new response surfaces each lead to different Pareto fronts, with the points on the front having different criteria values. Understanding which points are on the Pareto fronts more frequently and the likely range of criteria values can inform the decision process.

Step 1: Characterize the Pareto Front (objective): The goal of this step is to summarize the uncertainty associated with the PFs that is propagated from the estimation uncertainty in the individual responses. Using the PFs for each simulated surface, we summarize the frequency with which input factor combinations appear on the front. Those combinations which do not appear on the front frequently can be eliminated from further consideration. At the conclusion of this step, the decision-maker should see how the PF changes across the spectrum of anticipated response values as well as which locations are commonly chosen on the PF.

Step 2a: Identify Promising General Solutions (subjective): The goal of this step is to gain understanding about which locations are frequently selected as best for different weight combinations for the user-specified desirability function and scaling. To summarize this information, we combine the criteria into a single measure, identify how frequently different locations are best across all simulated response PFs and the entire set of weight combinations as well as how robust they are to different weightings of the criteria. Examining trade-offs between the criteria allows the decision-maker to understand how much compromise is needed on some of the responses to improve others. At the end of this step, the decision-makers should have improved understanding of which regions of the input factor space perform well for different weighting combinations as well as how frequently locations are best for some weighting.

Step 2b: Find Promising Solutions for More Focused Priorities (subjective): As the decision-makers narrow their search for a best solution to match their priorities, this step focuses on how frequently different solutions are identified as the best choice for a particular set of weights. Initially, weights for the desirability function can be partitioned into larger regions, and then subsequently a particular set of weightings of interest can be explored. For the selected range of weightings, we examine how frequently different solutions are selected as best. As different prioritizations of the criteria are considered, different locations in the input factor space are highlighted as common choices of best solutions. At the end of this step, the decision-maker has information about which locations are common choices for best for the specific priorities of the study.

Step 2c: Make Final Performance-Based Selection (subjective): Since the optimization of the product or process often necessitates selecting a single input factor combination from which to operate, this step guides the users to a final decision. Numerical and graphical summaries allow comparisons between individual solutions which inform the decision-makers of the relative merits of the available choices. Once commonly identified best solutions in the range of interest have been highlighted, evaluating and comparing their performance to the best available alternative for each weighting combination provides understanding about the merits of a solution. At the end of this step, the decision-makers should understand what choices are available and how they perform at optimizing the responses for the weightings of interest.

In the remainder of the paper, we describe the details of Steps 0 through 2c for the chemical process example. We illustrate how the numerical and graphical summaries in each of the steps can be used to identify promising candidates and eliminate non-contenders until a final solution is selected. The descriptive summaries also provide a quantitative means of justifying the choice. In Section 2, the simulation step is described. Section 3 discusses the objective Step 1 for characterizing the PF. Sections 4–6 describe the decision-making process of Steps 2a–2c that incorporates the user-specified desirability function for combining the measures and the priorities of the study as summarized by the weighing combinations of interest. Section 7 provides some discussion of extensions to the methods, while conclusions are given in Section 8.

2. Simulate multiple response surfaces

Due to sampling variability in the observed data, there is uncertainty associated with the estimated response surfaces based on the particular sampled data. If another set of data were collected, then we might obtain different estimates of the response surfaces, which might lead to a different set of choices on the PF and hence lead to a decision of a potentially different choice of operating conditions. To understand the impact of estimation uncertainty on the solution, we begin by generating new parameter values that are consistent with the current data. These new parameter values are used to approximate response surfaces that summarize the relationship between the responses and the input factors. The simulated response surfaces are central to the modified analysis process, which we break into two big steps for exploring (1) the impact of the variability in the PFs identified from the new estimated responses surfaces in the objective stage and (2) how this variability in PFs affects the subjective final solution based on understanding the trade-offs and the priorities of the study. In this section, we describe how to simulate new response surfaces.

First consider a linear model of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of responses, $\boldsymbol{X}$ is the $n \times p$ model matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of model parameters, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random errors. Under the assumption that the random errors are independent and identically distributed as $\epsilon_i \sim N(0, \sigma^2)$,

$$\hat{\boldsymbol{\beta}} \sim MVN(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}), \quad (1)$$

i.e., $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$ with a multivariate normal distribution with mean $\boldsymbol{\beta}$ and variance $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Further, $\hat{\sigma}^2 = MSE = SSE/(n-p)$ is an unbiased estimator of σ^2.

To generate new response surfaces, we repeatedly simulate new $\hat{\boldsymbol{\beta}}^*$'s that are consistent with the model parameter estimates. The steps for this process are as follows:

1. Identify and fit the appropriate model of the form $\boldsymbol{y}_r = \boldsymbol{X}_r\boldsymbol{\beta}_r + \boldsymbol{\epsilon}_r$ for each of the m response variables, $r = 1, \ldots, m$.
2. For each of the m response variables, randomly generate $\hat{\boldsymbol{\beta}}_r^*$ from $MVN\left(\hat{\boldsymbol{\beta}}_r, \hat{\sigma}_r^2(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1}\right), r = 1, \ldots, m$.
3. For each of the m response variables, use the simulated coefficients to approximate the response surface over a grid of points in the design region of interest. We use a grid of 630 locations inside the desired prediction region (a circle with maximum radius of $\sqrt{2}$) that correspond to multiples of 0.1 on the coded input scale determined by the Central Composite Design. Fig. 1 displays the grid locations that appear on PFs in the simulations.

To perform the modified analysis for the example process in the remaining sections, we repeat the above steps 500 times, as 500 simulations provided a good balance between stability of the results (when replicating the entire simulation–based analysis) and manageable computation time. This approach to understanding response variability captures the inherent uncertainty in the estimation of model parameters and allows direct translation into its impact on the estimated response surfaces. In the remaining sections we describe how this information is used to understand the impact of estimation uncertainty on the Pareto front and the final selected solution.

3. Characterize the Pareto front

This section describes the objective first step in the modified PF approach, which is to understand the general pattern of the PFs and how much variability is associated with the range of possible responses, as well as identifying promising locations and eliminating non-contenders from further consideration. Particularly, given the 500 response surfaces approximated over the grid of operating locations in the input factor region generated in Section 2, we identify the PF for each of the simulated response surfaces and then summarize across the 500 PFs to eliminate locations that do not appear on PFs often. These locations are not suitable to be considered for further selection regardless of the prioritization of the multiple responses, since it is more likely to find superior locations with strictly better response values.

The circle plot in Fig. 1 illustrates the relative frequency with which the locations on the grid appear on the PFs across the 500 simulations. The axes on the left and bottom show locations in their scaled units, while the top and right scales reflect the original units. The size of the point (location) in Fig. 1 is proportional to the number of times the location appears on PFs among the 500 simulations, with larger points corresponding to those appearing more often. The largest of the points correspond to the 43 locations which appeared on the Pareto fronts in at least 95% of the simulations. The 55 locations that did not appear on any of the PFs in the 500 simulations are not displayed in the circular region. The larger points located in the bow tie shaped region, approximately centered around (0,0) on the coded scale, are the contenders when estimation uncertainty is considered. The remaining locations which are closer to the edge of the circular region, with the exception of the region of $x_1 \in [-1.2, -0.7]$ & $x_2 \in [-1.2, -0.7]$, are not included on the PFs often, and hence should likely be eliminated from further consideration. The identified promising locations nicely overlap with the locations on the PF based on the mean model response surfaces (Chapman et al., 2014).

Fig. 2 shows the general shape and position of the PFs from the simulated response surfaces. The left panels display all the points (in gray) that have appeared on the PF in at least one of the 500 simulations with black points highlighting the 181 points on the PF based on the mean model response surfaces. We can see the range of response values that might be expected for points on the PFs. The estimated values range roughly from 75 to 81 for yield, 52 to 74 for viscosity, and 2750 to 3900 for molecular weight. In addition, we observe the variability in the PFs due to the estimation uncertainty of the response surfaces as well as their position relative to the mean model PF. The right panels display an approximate "frequency" with which points appear on the PFs, with dark regions appearing more frequently and corresponding to the region with larger points in Fig. 1.

This objective step allows us to examine the possible range of the response values and the amount of variability in the simulated PFs which is propagated from the uncertainty of the estimated response surfaces. In addition, we can identify solutions that are sensible to be considered further. This evaluation is independent of any subjective choices related to the users' priorities of the responses, and what form and scaling to select for integrating multiple responses into a single numeric summary for ranking solutions. This step is objective since solutions are valued only based on their approximated responses with estimation uncertainty incorporated. This forms a logical set of choices to select from and examined in subsequent steps for evaluating their trade-offs and relative performance when the subjective factors, such as decision-maker preferences, are brought into consideration. The choice of these graphical summaries allows for visualization of the locations in the design space, the range of responses possible from these locations, and how much fluctuation there is in the identified Pareto fronts between different simulated surfaces. This calibration translates into a fuller and more realistic understanding of what to expect from future data from this process.

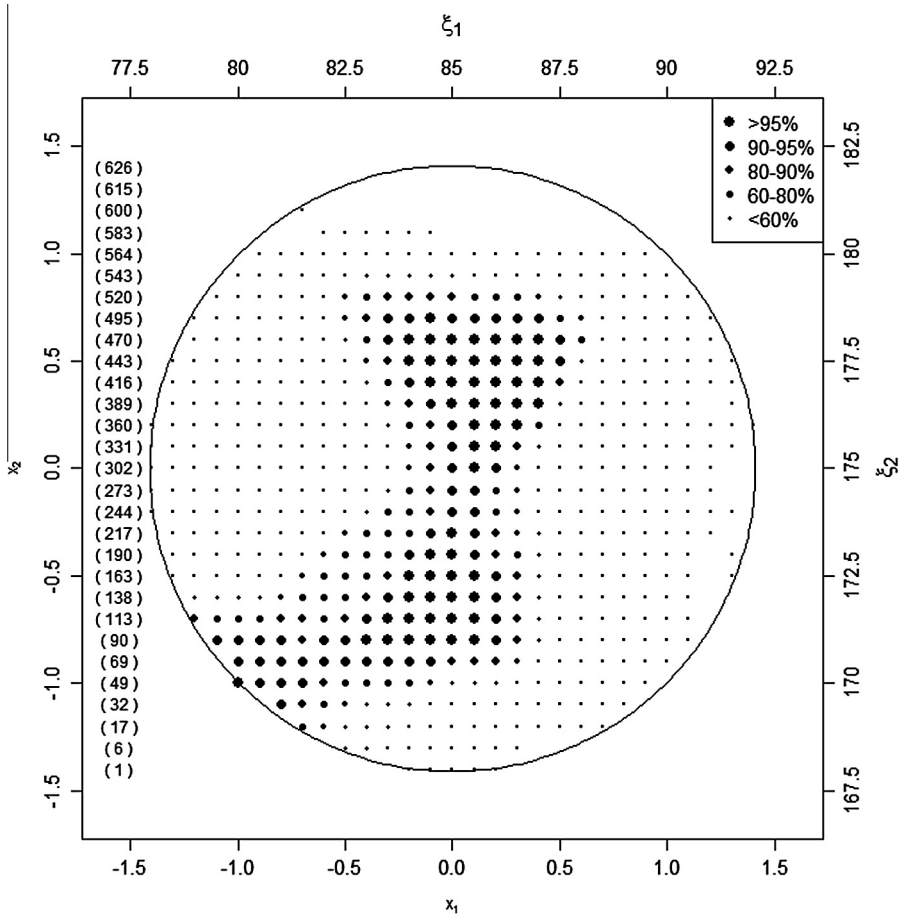

Fig. 1. Circle plot displays most of the grid of points at which the three responses are predicted. The size of the points is related to the proportion of times the design appeared on Pareto fronts in the 500 simulations (larger points appeared more often). The largest points appeared on the Pareto fronts in at least 95% of the simulations, while the smallest appeared on the Pareto fronts in less than 60% of the simulations. Points within the circle which did not appeared on any of the 500 Pareto fronts are not shown.

4. Identifying promising general solutions

The first objective step has successfully eliminated over two-thirds of the inferior locations and allows the experimenters to make a selection from a more manageable smaller set of candidates. In the second step of the decision-making, we consider the subjective aspects of the decision and bring in user priorities for evaluating and ranking the candidate locations. Often statisticians may feel like subjectivity is to be avoided, but subjectivity is essential for tailoring the solution to match study goals. However, an awareness of the impact of subjectivity is helpful. For instance, the desirability function approach with a single set of specified weights typically does not consider the impact of the chosen weights and identifies only a single solution with little or no understanding of the robustness of the final solution. In contrast, the Pareto front approach allows for the formal introduction of user priorities and a careful evaluation of their impact on the decision-making process.

The Utopia point approach (Lu et al., 2011) is used to select more promising locations from the set of most frequently appearing locations on the PFs. This method selects a smaller set of locations that are "closest" in distance to the "ideal" Utopia point solution on the desirability scale across the entire range of weighting choices. This measure of distance is dependent on a selected distance metric form as well as the scaling scheme chosen to convert the original response value onto the 0–1 desirability scale. Lu et al. (2011) prove that using the Utopia point approach with an L_1-norm distance metric selects the same optimal solutions as using an additive desirability function with the same scaling, expressed as

$$DF_{add}(j, \boldsymbol{w}) = w_1 C_1(j) + w_2 C_2(j) + w_3 C_3(j),$$

where $C_i(j)$ represents the scaled value of response i for location j, for $i = 1,2,3$, and $\boldsymbol{w}$ is the weight vector, representing the user-specified weight given to response i, with $\sum_{i=1}^{3} w_i = 1$ and $w_i \geqslant 0$. The commonly used multiplicative DF, expressed as

$$DF_{mult}(j, \boldsymbol{w}) = C_1(j)^{w_1} \cdot C_2(j)^{w_2} \cdot C_3(j)^{w_3}, \quad (2)$$

is equivalent to using the Utopia point approach with an L_1-norm on the log scale (Lu et al., 2011). For compatibility with the previous results by Myers et al. (2009) and Chapman et al. (2014), we choose to use the multiplicative DF in Eq. (2), which severely penalizes poor performance for any of the responses.

For many multiple objective optimization problems, the objective functions or criteria are assumed to be deterministic with no variability in the criterion values for every specific solution. The scaling is usually determined by the range of criterion values on the PF. However, for our case with estimation uncertainty in the response surfaces, it requires a wider scaling for incorporating this extra variability in the estimated response values. We choose to use 95% prediction bounds for determining the best and worst responses in our scaling. More specifically, for a general model form as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{Y}$ is an $n \times 1$ vector of the responses, $\boldsymbol{X}$ is the $n \times p$ model matrix, $\boldsymbol{\beta}$ is the $p \times 1$ parameter vector, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of independent and identically distributed random errors. A 95% two-sided prediction interval for the response using the distribution in Eq. (1) at a particular input location $\boldsymbol{x}_0$, with $\boldsymbol{x}_0' = (x_{01}, x_{02}, \ldots, x_{0p})$ in the model form, is given by

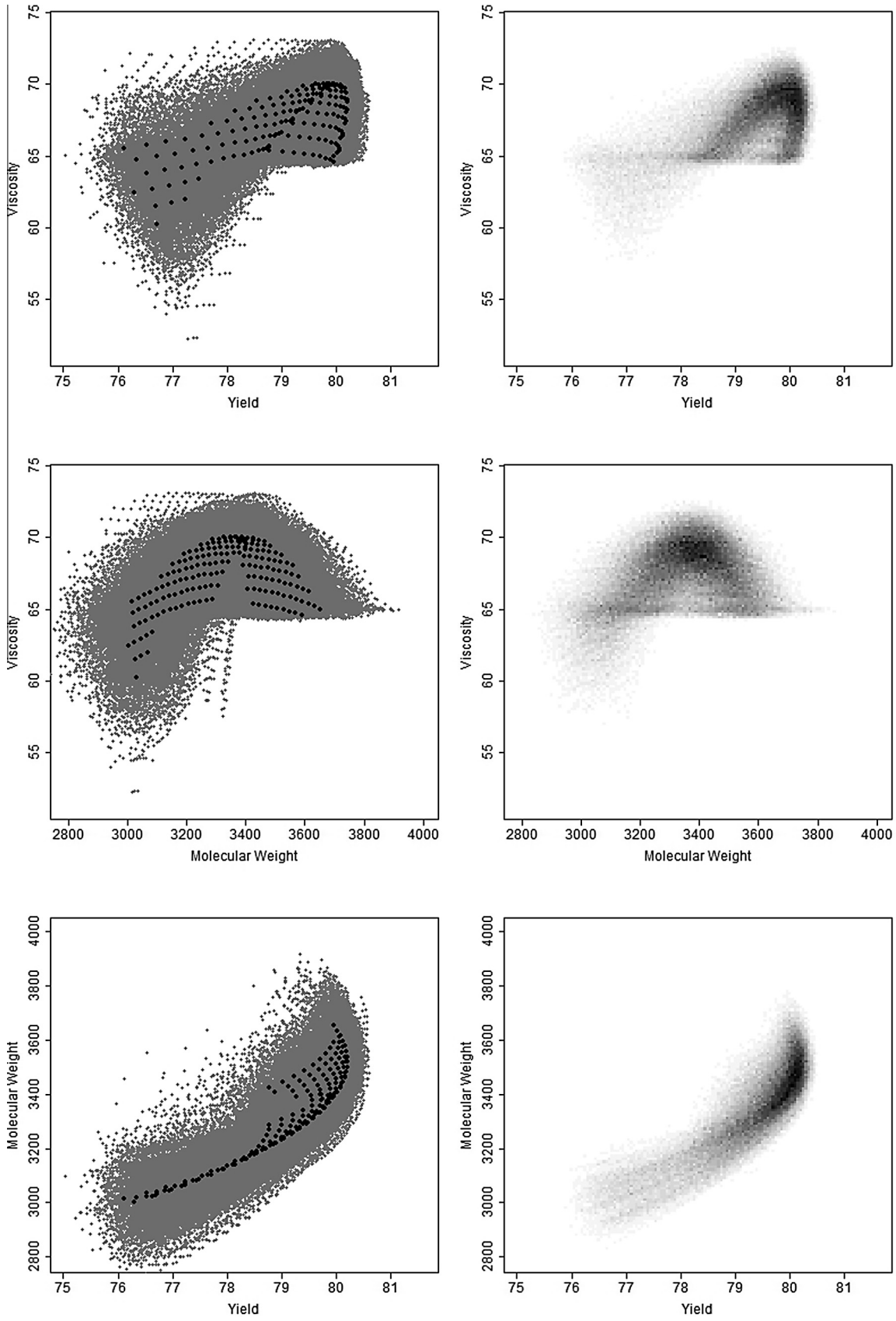

Fig. 2. (Left) Plots display the general shape of the PFs observed among the 500 simulations. The points on the mean-estimate PF are displayed with black symbols. (Right) Plots display the "frequency" with which points appear on the PFs, with dark regions appearing more frequently.

$$\boldsymbol{x}_0'\hat{\boldsymbol{\beta}} \pm t_{0.975,n-p}\sqrt{MSE\left(1+\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0\right)}. \quad (3)$$

In Eq. (3), $\hat{\boldsymbol{\beta}}$ and MSE are the least squares or maximum likelihood estimates of the model coefficients $\boldsymbol{\beta}$ and the estimated variance of the random errors, respectively. For example, for $\boldsymbol{x}_0' = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2) = (1, 0.5, -1, -0.5, 0.25, 1)$, a future observation on the yield is predicted to be in the range of (77.73, 79.18) at the 95% confidence level.

To choose the best and worst values for scaling the yield criterion, we use the original fitted response surfaces to compute 95% prediction bounds for the yield associated with each point on the grid of input combinations. Since the goal is to maximize yield, the largest upper bound obtained from all of the simulated prediction intervals is used as the best value for scaling purposes. Similarly, the smallest lower bound is used as the worst value for scaling. Since the molecular weight criterion is being minimized, we choose as the best value the smallest lower bound from all of the 95% prediction intervals. The worst value is selected as the largest upper bound from the prediction intervals for molecular weight. For the viscosity criterion, which is trying to achieve a target of 65, we use 0 as the best value because 65 is included in the range of plausible values. The prediction bound (either lower or upper) that has the largest absolute deviation from the target is used as the worst value for scaling the viscosity criterion. Because the available data were used to construct the prediction intervals for the responses at each grid point and were used to simulate new response surfaces, theoretically it is possible that a simulated response is outside the range of the values used for scaling. Thus, if a predicted response from one of the simulated response surfaces exceeds (is below) the best (worst) value, the scaled criterion value is set to 1 (0).

Based on the selected scaling and DF form, we first explore which locations on the mean model PF are selected as best for different possible weight combinations to provide some sense about the typical performance without incorporating the estimation uncertainty. Fig. 3 shows the mixture plot of the best locations for all possible weightings of the three responses based on the mean model. The vertices and the edges correspond to optimization based on a single criterion and two of the three criteria, respectively. A more detailed description of mixture plots is

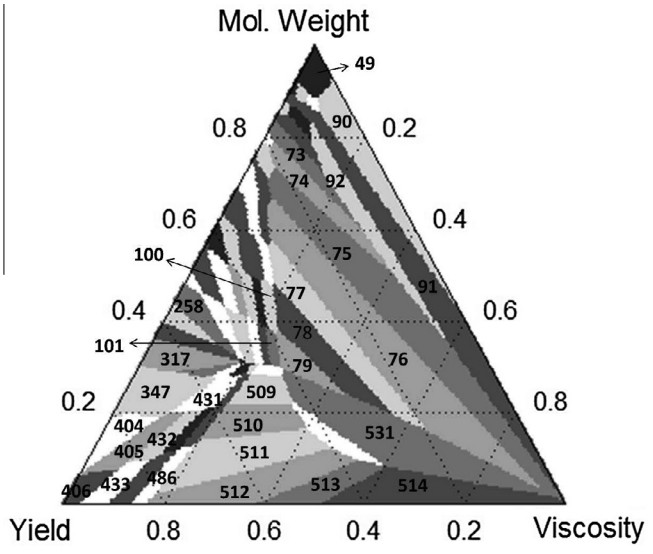

Fig. 3. Mixture plot for designs on the Pareto front for the mean model response surfaces when multiplicative desirability is used and the criteria values are scaled based on 95% prediction intervals for the responses at each location.

available in Cornell (2002), [p. 24]. Of the nearly 200 solutions on the front for the mean model, 22 are identified as being optimal for at least 1% of the total simplex area of weights. Additionally, a few solutions in the bottom left corner, which are optimal when yield is weighted dominantly (weighted above 60%), are also identified.

Table 1 provides more details about the performance of different locations. Locations 75 and 76 are the best solutions for the two largest weighting areas (both around 10%) with generally less than 30% weight for yield and a large variety of weights for the other two responses. Location 531 corresponds to the third largest area (7.7%) with at least 30% weight for viscosity and slightly more weight for yield than molecular weight. Locations 77–79 are adjacent solutions in the circle plot, which are optimal around the centroid region of the weighting simplex (equal weighting for all three criteria). Locations 509–512 are best solutions when molecular weight is weighted less than 30% and yield is weighted between 40% and 80%. Locations 431–433 are optimal when yield is weighted at least 50% and the other two criteria are weighted equally. Locations 406, 514, and 49 are the optimal solutions for univariate optimization of yield, viscosity, and molecular weight, respectively.

We next summarize the mixture area information for the simulated PFs in Fig. 4 and Table 1. Locations selected as best for at least some weight combination(s) in at least one of the 500 simulations are plotted in square shaped points. The size of the points is proportional (larger is better) to the number of times the location is best for at least one weight combination in the 500 simulations. The different gray-scale of points is used to reflect the average mixture area when the location is best for at least one weight combination. The darker points were selected as best for larger regions of weight combinations on average. The largest and darkest points are most desirable and labeled with their location number directly above the point.

There are two main areas where locations are observed to perform well across the simulations: the first area corresponds to $x_1 \in [-1, 0]$ ($\xi_1 \in$ [80 min,85 min]) and $x_2 \in [-1, -0.5]$ ($\xi_2 \in$ [170 F,175 F]), which includes the most balanced solutions as well as those having more emphasis on molecular weight; and the second area corresponds to $x_1 \in [0,0.5]$ ($\xi_1 \in$ [85 min,87.5 min]) and $x_2 \in [0,0.75]$ ($\xi_2 \in$ [175 F,178.75 F]), where a majority of the locations have better performance for yield and a couple have near-optimal viscosity. These labeled locations overlap well with the locations identified in the mixture plot for the PF based on the mean model (Fig. 3).

In addition, a trade-off plot with uncertainty bands for the response values for some selected locations from Fig. 4 is displayed in Fig. 5. The innermost vertical axes display the desirability scale while the outer vertical axes represent the criterion scales. Locations are ordered from worst to best performance on the yield response. Uncertainty bands are based on the 5% and 95% percentiles of the simulated responses for each location. For example, the estimated responses for yield, viscosity, and molecular weight for Location 49 are 0.224, 0.849, and 0.735 on the desirability scale, which correspond to 76.3, 62.5, and 3003.7 in the original response values, respectively. The uncertainty bands indicate that we could expect a yield between 76.0 and 76.7, a viscosity between 0.3 and 5.5 away from the target of 65, and a molecular weight between 2846.2 and 3144.3. We can see the uncertainty bands for the yield criterion are the narrowest, indicating that we may expect less variability in the estimated responses for yield. The molecular weight has generally the widest uncertainty band. The locations in the middle left region (77–80 and 100) have the most balance between responses, while the locations at both ends represent options with near-optimal performance for one or two of the responses, while sacrificing with poor performance for another response.

Table 1
Summary of mixture areas (expressed in 100% format) for selected locations. The area in the original multiplicative mixture based on the mean model as well as the percent of times the location was identified as best in the 500 simulations and summaries of the mixture areas from the simulations (when the location is best for at least one weight) are provided.

Loc	x_1	x_2	Mixture area in Fig. 3 (%)	Simulated fronts						
				% Best	Mean area	2.5%	25%	50%	75%	97.5%
49	−1.0	−1.0	0.99	88.4	1.71	0.03	0.58	0.99	1.69	8.42
77	−0.2	−0.9	3.67	67.4	2.76	0.03	0.36	0.99	3.00	15.45
79	0	−0.9	1.09	46.0	3.03	0.03	0.38	0.90	3.22	19.00
100	−0.1	−0.8	0.40	81.2	1.45	0.04	0.29	0.48	1.03	7.99
113	−1.2	−0.7	0.06	62.8	1.79	0.03	0.49	1.02	1.99	8.03
405	0.3	0.3	1.42	97.0	1.17	0.17	0.72	1.10	1.48	2.52
432	0.3	0.4	0.72	96.8	0.79	0.12	0.49	0.69	0.95	1.78
460	0.4	0.5	0.83	94.6	0.87	0.12	0.47	0.66	0.92	3.18
485	0.3	0.6	0.39	92.6	1.38	0.06	0.31	0.42	0.66	11.18
486	0.3	0.6	0.75	92.4	1.66	0.07	0.43	0.59	0.94	11.75
510	0.3	0.7	2.32	77.4	2.65	0.09	0.31	0.58	3.08	17.09
511	0.4	0.7	4.19	74.6	2.96	0.11	0.41	0.80	3.56	17.88
512	0.5	0.7	4.23	51.0	2.97	0.03	0.42	1.08	3.29	16.46
533	0.2	0.8	NA	41.0	3.18	0.06	0.30	1.64	4.53	14.00
535	0.4	0.8	NA	39.4	4.79	0.13	0.65	2.34	5.24	25.19

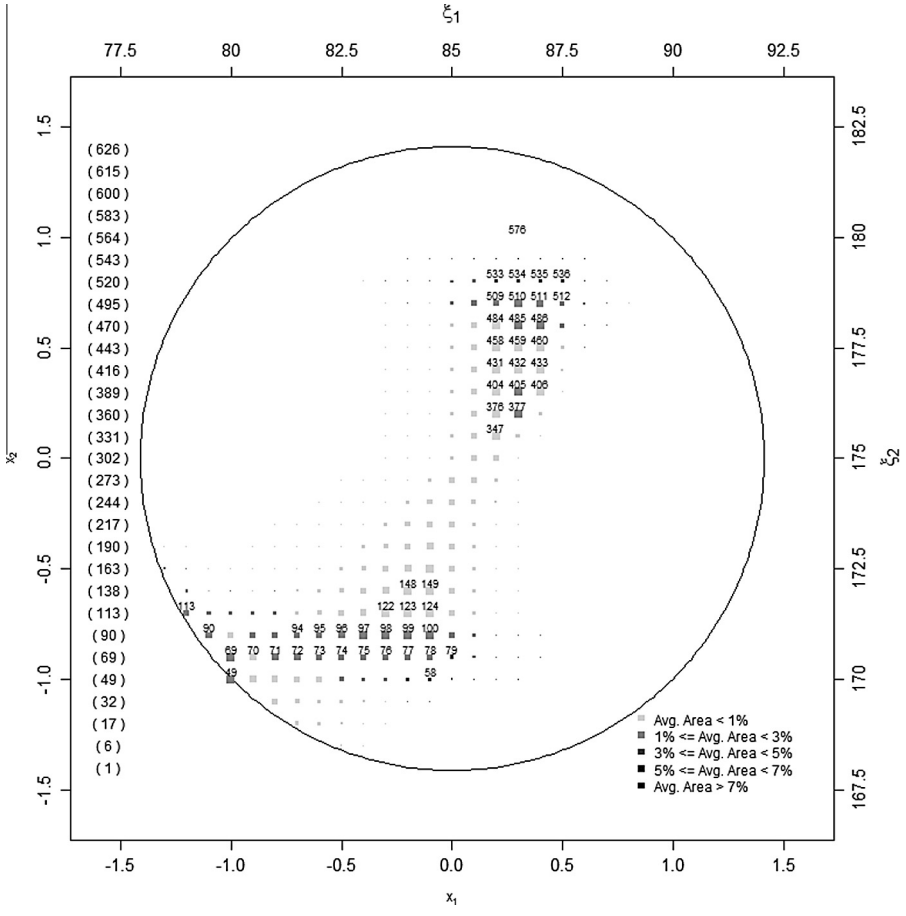

Fig. 4. Circle plot summarizing the mixture information for the locations. The size of the points is proportional to the number of times the design location is flagged as best for at least one weight combination in the 500 simulations (larger points were identified as best more often). The grayscale conveys the average multiplicative mixture area when the location is identified as best for at least one weight combination (darker points are best for larger regions of weight combinations). The largest and darkest points are labeled with their location number directly above the point.

In this step, we work to gain a better understanding of the general areas of promising solutions. The mixture plot provides a compact way of summarizing where individual solutions excel. The circle plot highlights both the frequency of selection into the set of superior choices and the robustness to perturbations of weightings. The trade-off plot provides a concise summary of the estimated response values with their uncertainties for particular input combinations of interest. Taken together this collection of summaries provides a diverse and complementary set of information to help select the most promising solutions when estimation uncertainty is desired to be addressed.

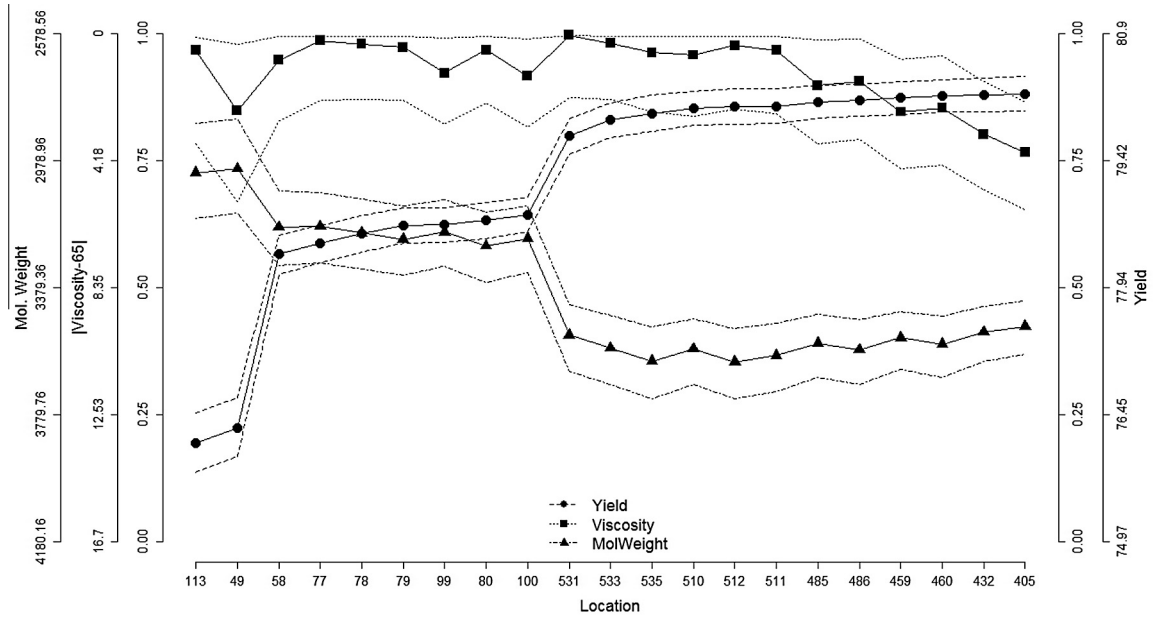

Fig. 5. Trade-off plot with uncertainty bands for selected locations. Locations are ranked from worst to best performance on the yield criterion. Uncertainty bands are based on the 5% and 95% percentiles of the simulated responses for each location.

5. Promising solutions for more focused priorities

Step 2a described in Section 4 allows us to gain an understanding of how much trade-off exists between the responses, how much uncertainty is associated with each response, and which regions of the input factor space perform well for different weighting choices across all possibilities. These insights are helpful for improving the decision-makers' understanding of the process/product and the interrelationship between the responses, as well as refining how to quantify priorities and study goals by understanding how elements interact.

Building on the previous steps, Step 2b in this section provides tools for helping decision-makers to narrow their choices down to a more focused region matching their priorities. The step starts by partitioning the entire weighing space into larger regions and identifying locations that appear more often as optimal solutions for the whole as well as individual weighting regions. This step is helpful for evaluating the global as well as local optimization performance of individual solutions and further selecting the most promising choices that achieve a good balance of the two aspects. Subsequently, decision-makers can transition to considering a narrower region of interest.

Fig. 6 illustrates a simple partition of the weighting space into four large regions. These regions represent different prioritization of the three responses, which are displayed in different shades of gray. Region 1 consists of weight combinations that emphasize yield over the other two criteria. Similarly, Regions 2 and 3 have primary emphasis on viscosity and molecular weight, respectively. Region 4, in the center of the simplex, consists of weight combinations that present more balance among the three responses.

The entire space consists of 496 weight combinations, with each entry of the weight vector being a multiple of 1/30. For each of the 500 simulations, the location for which each of the 496 weights is best is determined. Fig. 7 shows a stacked Pareto (bar) plot for the 20 locations that are identified as best most often for the whole weighting space across all 500 simulations and the 496 weightings considered. Each bar shows the average frequency per simulation that the location is best. Within each location, weights for which the location is best are classified as one of the four regions defined in Fig. 6. Hence smaller bars are stacked

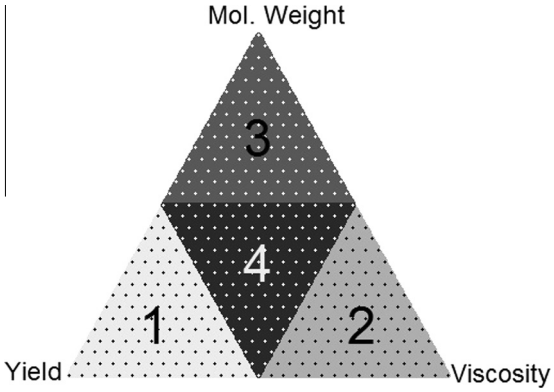

Fig. 6. Weight space partitioned into four regions. Region 1 consists of weight combinations that emphasize yield over the other two criteria. Region 2 emphasizes viscosity. Region 3 emphasizes molecular weight. Region 4 consists of the weight combinations that provide the most balance among the three criteria.

vertically to represent the average frequency per simulation for individual weight regions. For instance, Location 511 was identified as best for the most weights (approximately 14 per simulation). Of these weights, Location 511 is best for about six weights in Region 1, three in Region 2, and five in Region 4 per simulation.

Several patterns can be observed in Fig. 7. First, no location among the top 20 is best for weightings from all four regions. Typically, they are optimal for relatively large proportions of weight combinations from a couple of regions and then a smaller amount of weightings from a third region. This is a result of the substantial trade-offs between the three responses. There is no location that performs uniformly well for all of the weight regions. Secondly, most locations among the top 20 are best for a substantial portion of weightings from the center Region 4. This indicates that locations with relatively more balanced performance tend to have higher ranks in the global summary. Also, no locations among the top 20 are best for weightings from both Regions 1 and 3. This matches what we observe from Fig. 5 where yield and molecular weight exhibit the strongest trade-off among the three responses. Finally, all of the top 20 allocations are best for a small to moderate

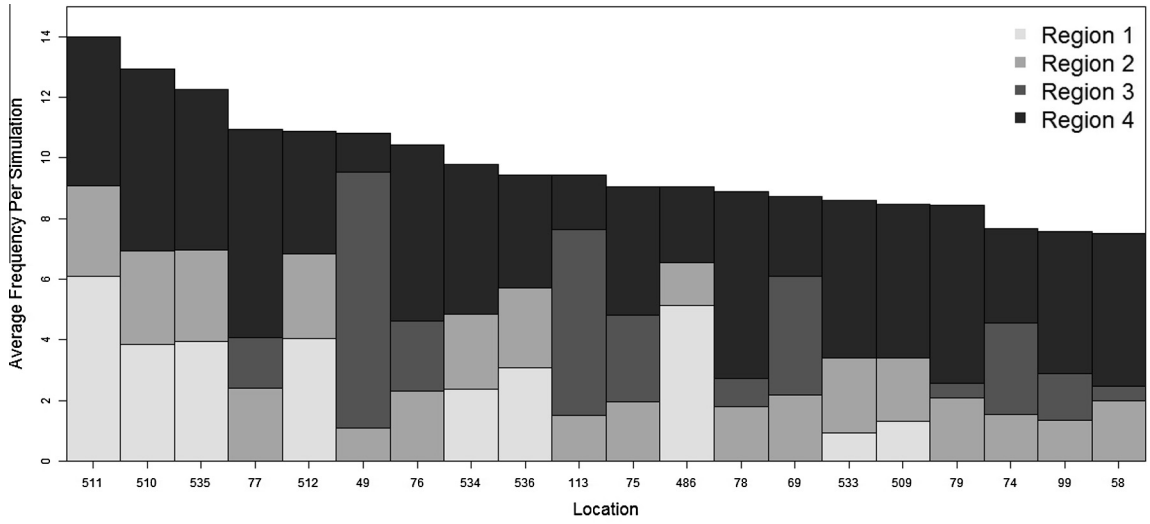

Fig. 7. The 20 locations identified as best most often across all simulations and the 496 weights considered. Within each location, weights for which the location is best are classified as one of the four regions defined in Fig. 6.

portion of weights from Region 2, which indicates it is very likely to find a best solution with good performance for viscosity since it has the least trade-offs with the other two criteria and does not require too much compromise when balancing between the choices.

In addition, there is substantial overlap between the locations identified in Fig. 7 and those from Fig. 3 based on the mean model, particularly for locations around the middle area of Fig. 3 with a reasonable size of the area (Locations 75–79 and 509–512). For locations that are in Fig. 7 but not labeled in Fig. 3, most represent adjacent locations to the labeled ones, which are expected to have not too different performance on average. Additionally, all of these locations are gathered around two bigger regions (the bottom left and the top right) in the bow-tie shaped region identified in Fig. 4.

Now that the decision-makers have an understanding of which regions of the input factor space perform well for different weighting choices, they are in a better position to consider a more focused region based on their own specific priorities. Suppose one decision-maker thinks all three responses should be weighted similarly, with the weight for each criterion ranging between 20% and 40%. This region of interest is displayed in Fig. 8(a) with 28 weight combinations included. This region is further divided into four smaller regions. Each of the Regions A–C has slightly less emphasis on molecular weight (blue), viscosity (green), or yield (red) while Region D (orange) has roughly equal weights for all three responses. This further partition offers potential to see more focused patterns. Fig. 8(b) shows the Pareto (bar) plot of the average frequency per simulation for the top 20 choices in this region. Again, the ranking for the most centered smaller region generally is similar to the summary over the entire region. We can still see trade-offs between molecular weight and yield with almost no top location doing well on both the red and blue regions. However, this pattern is not as strong as for over all possible weightings in Fig. 7. All of the top 12 choices are located around the bottom left portion of the bow-tie shaped region in the input factor space.

Table 2 summarizes the top location (appearing most frequently as the best) for each of the 28 weight combinations considered in this region. Locations 76–80, 100–101 and 508–509 appear frequently on the list. Due to the variability in the estimated response surfaces, the best location for a particular weight combination can vary across the simulations. Fig. 9 shows the truncated "distribution" of the optimal solutions for the 28 weight combinations in the region. Each row corresponds to a specific weight combination and the possible optimal locations are shown in descending order of the frequency of appearing as best in the 500 simulations. Locations with smaller than 2% frequency (i.e. best for less than 10 simulations) were truncated on the right hand side. Adjacent locations 77–79 and 100–101 appear most often as the top three designs for all weight combinations except in the blue region A, where locations 508–509 appear to be the top choices.

Now, suppose a different decision-maker believes that the yield criterion is the most important and should be weighted between 60% and 80%, with the remaining weight split between the other two criteria. Figs. 10 and 11 summarize the top choices for this scenario. This region is further divided into three regions depending on the weight that yield receives. Fig. 10(b) shows the top 20 most frequent best solutions averaged across the simulations and the weighting area. Locations 432 and 405 are the most promising solutions for the whole region as well as Regions E and F with bigger weights for yield. Locations 485 and 459 are the top choices for Region G with slightly less weight for yield. These locations are also the common top locations for the majority of the individual weight combinations in the region as seen in Table 3, as well as the common choices for the top three most frequent best locations for the 16 weight combinations shown in Fig. 11.

By the end of this step, the decision-makers have gained information about which locations are the most frequent choices as best in the focused region matching the priorities of the study. The approach outlined provides detailed information about individual input combinations, while still providing a higher level summary that connects smaller regions of the input space. In addition, the performance of individual solutions is summarized at different levels corresponding to varied degrees of focus on the weighting priorities, which offers the decision-maker a comprehensive understanding of how their level of weighting uncertainty may affect their decision. Being able to extract information at several levels provides information about robustness while still allowing an individual solution to be identified. Note that as a result of the continuity of the response surfaces under consideration, several adjacent locations are identified as common top choices. This is reassuring that we have found a common region of the optimal condition of the process/product for the particular priorities of the decision-makers.

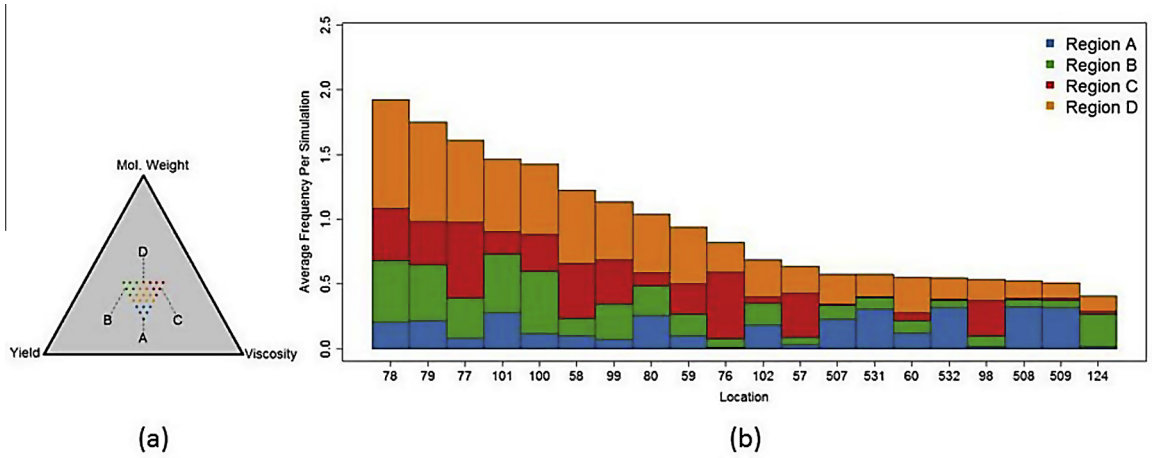

Fig. 8. (a) Region of weight space where each criterion receives a weight between 0.2 and 0.4. The weights in the blue region emphasize yield and viscosity over molecular weight, those in the red region emphasize molecular weight and viscosity over yield, and those in the green region emphasize yield and molecular weight over viscosity. The weights in the orange region represent the most balance between the three criteria. (b) The 20 locations that are flagged as best most often across all simulations and the 28 weights considered in this region. Within each location, weights for which the location is best are classified as one of the four regions defined in (a).

Table 2
The most frequent best locations in the region with all weights between 20% and 40%.

Weight index	Weight vector	Top location	Weight index	Weight vector	Top location
A1	(0.33, 0.4, 0.27)	79	C3	(0.23, 0.4, 0.37)	76
A2	(0.37, 0.37, 0.27)	80	C4	(0.27, 0.33, 0.4)	77
A3	(0.37, 0.4, 0.23)	508	C5	(0.27, 0.37, 0.37)	77
A4	(0.4, 0.33, 0.27)	508	C6	(0.27, 0.4, 0.33)	77
A5	(0.4, 0.37, 0.23)	509	D1	(0.3, 0.3, 0.4)	77
A6	(0.4, 0.4, 0.2)	509	D2	(0.3, 0.33, 0.37)	77
B1	(0.33, 0.27, 0.4)	78	D3	(0.3, 0.37, 0.33)	78
B2	(0.37, 0.23, 0.4)	100	D4	(0.3, 0.4, 0.3)	78, 79
B3	(0.37, 0.27, 0.37)	78	D5	(0.33, 0.3, 0.37)	78
B4	(0.4, 0.2, 0.4)	100	D6	(0.33, 0.33, 0.33)	78
B5	(0.4, 0.23, 0.37)	101	D7	(0.33, 0.37, 0.3)	79
B6	(0.4, 0.27, 0.33)	101	D8	(0.37, 0.3, 0.33)	79
C1	(0.2, 0.4, 0.4)	76	D9	(0.37, 0.33, 0.3)	78
C2	(0.23, 0.37, 0.4)	76	D10	(0.4, 0.3, 0.3)	101

6. Final performance-based selection

Given the few top choices identified at the end of Section 5, Step 2c guides the decision-makers to the final decision based on evaluating and comparing the actual performance of individual choices summarized over the chosen focused region as well as the entire weighting space (if global performance is of interest). The Fraction of Weight Space (FWS) plot with uncertainty bands offers a compact and convenient summary for this purpose.

The FWS plot for multiple criteria optimization without uncertainty was originally proposed in Lu, Chapman, and Anderson-Cook, 2013 and then adapted for summarizing over a focused region in Lu, Anderson-Cook, and Lin (2013) to provide a dimension-free summary (that flexibly allows for any number of criteria). The FWS plot quantifies the relative performance of individual solutions relative to the best available for each particular weight combination over the entire region of interest. The relative performance for a solution, j, at a weight combination, $\boldsymbol{w}$, is measured by its synthesized efficiency (Lu & Anderson-Cook, 2012) defined as $SE(j,\boldsymbol{w}) = \frac{DF(j,\boldsymbol{w})}{max_j(DF(j,\boldsymbol{w}))}$.

The FWS plot is built from the synthesized efficiency values for a fine set of weight combinations, such in as in Fig. 6. The lines for each solution show for what fraction of the weights (horizontal axis) the solution has synthesized efficiency at least as large as a certain percentage of the best possible performance (vertical axis). Fig. 12 displays the FWS plots with uncertainty bands for several selected locations. In these plots, the solid line represents the median (across the 500 simulations) of the empirical fraction of the weighting region for which the solution's synthesized efficiency as a certain percentage of the best possible, while the dashed lines represent the 5th and 95th percentiles from the simulations. The FWS plot provides an overall, quantitative summary of the performance of an individual solution across the weighting region of interest and can be used to easily compare several competing solutions.

Fig. 12(a) shows the FWS plot summarized over the focused region for the first scenario discussed in Section 5 where a decision-maker believes the weights for all three responses should be between 20% and 40%. The six most promising locations in Section 5 are examined here. We can see that Locations 77–79 and 101 all represent similar performance with above 90% synthesized efficiency for the median over the focused region and above 85% efficiency for the worst case percentile summary. Location 100 has almost identical performance to 101 and is not shown here. Locations 508–509 are similar to Locations 77–79 and 101, with slightly worse performance at the tail end of the FWS curve.

To compare the global performance of these locations, Fig. 12(b) shows the FWS plot summarized over the entire weight space. Again, Locations 77–79 perform almost identically while Location 101 (above Location 79 in Fig. 4) performs similarly with slightly

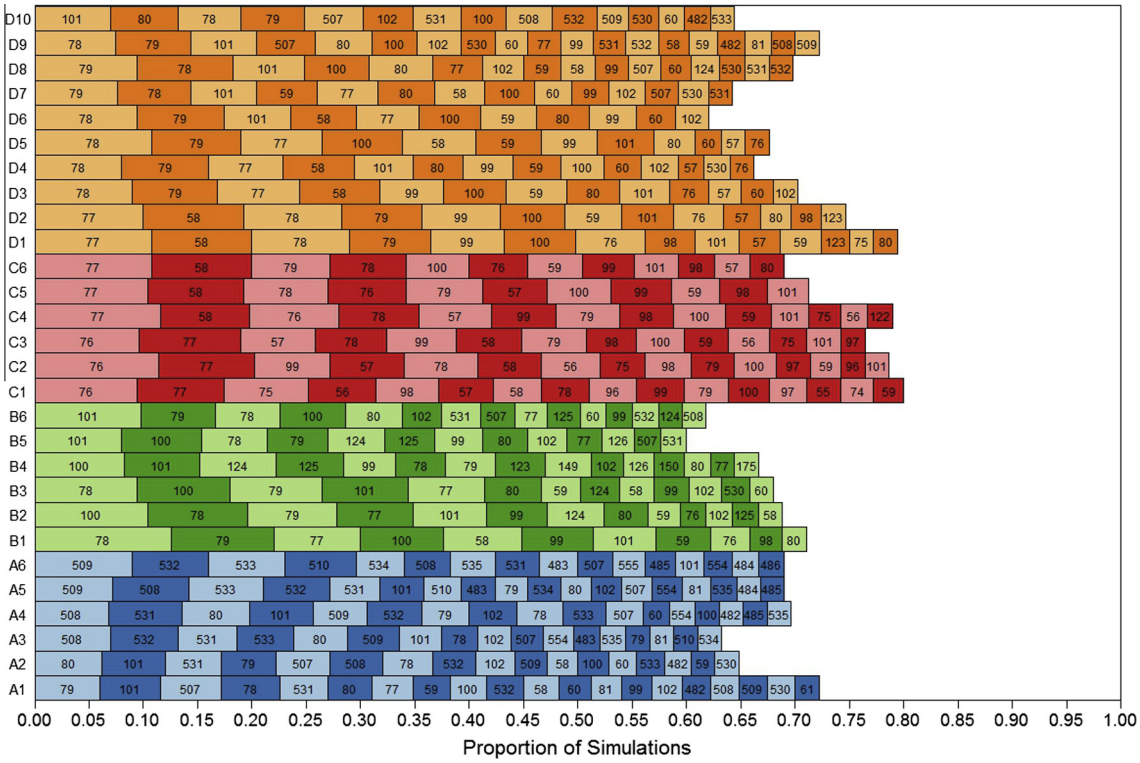

Fig. 9. Locations flagged as best in 2% or more of 500 simulations for weights considered when all criteria receive weights between 20% and 40%.

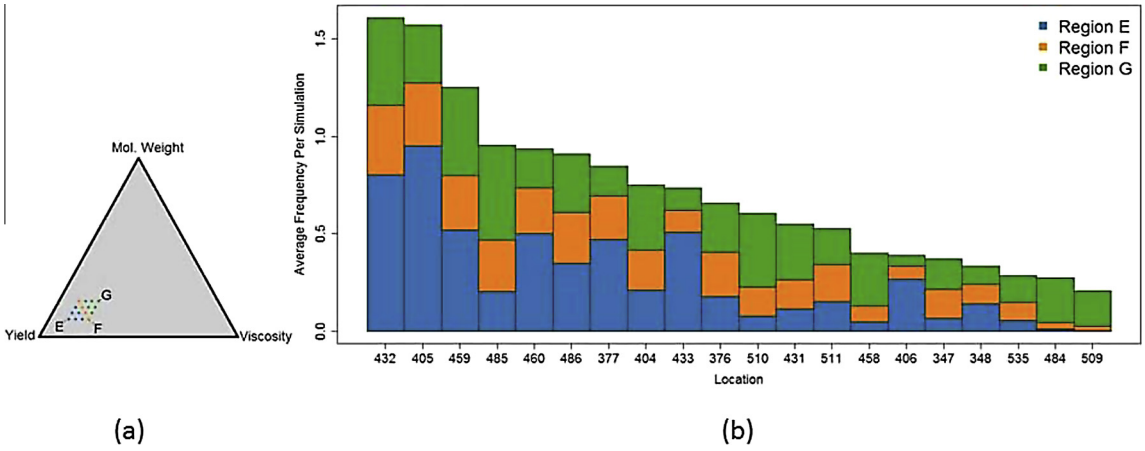

Fig. 10. (a) Region of weight space where the yield criterion receives between 60% and 80% of the weight while the remaining two criteria receive weights between 10% and 20%. The blue, orange, and green regions correspond to weighting combinations where yield receives more than 70%, exactly 70%, and less than 70% of the weight, respectively. (b) The 20 locations that are flagged as best most often across all simulations and the 16 weights considered in this region. Within each location, weights for which the location is best are classified as one of the three regions defined in (a).

better synthesized efficiency at the tail end of the FWS curve. All four locations have at least 75% synthesized efficiency over the entire weighting space. In contrast, Locations 508 and 509 do not have consistent good performance over all possible weightings. Their FWS curves drop quickly with the median and worst case efficiency near 60% and 40%, respectively.

Now, consider the second scenario where another decision-maker believes that yield should receive substantially more weight than the other two criteria. The FWS plot for the top performing locations identified by the second decision-maker is displayed in Fig. 13. First consider Fig. 13(a), where the fractions of the weighting space are computed using only weights in the desired region. Location 486 (with similar performance to Location 485) has the highest synthesized efficiency with narrowest uncertainty band across the focused region. Location 460 has pretty comparable performance for the median and best case summary. However, the worst case summary has slightly lower synthesized efficiency around the tail end of the curve. The other locations have shown consistently lower synthesized efficiency and more uncertainty than Locations 485 and 486.

To compare the global performance of these six locations, Fig. 13(b) considers the FWS plot summarized over the entire weight space. Locations 485, 486, and 406 perform similarly, with curves that drop quickly to around 60% and 40% for the median and

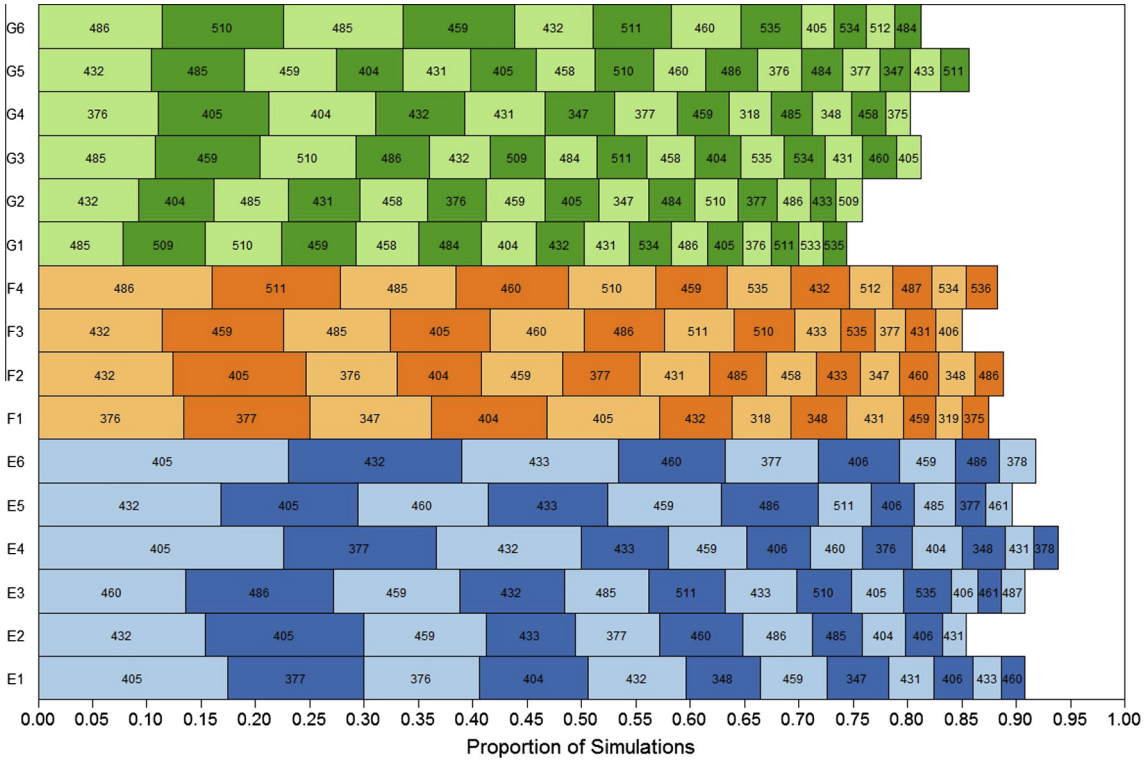

Fig. 11. Locations flagged as best in 2% or more of simulations for weights considered when the yield criterion receives between 60% and 80% of the weight while the remaining two criteria receive weights between 10% and 20%.

Table 3
Weights considered in focused investigation when yield is weighted 60 – 80% and both viscosity and molecular weight receive weights between 10% and 20%.

Index	Weight vector	Top location
E1	(0.73, 0.1, 0.17)	405
E2	(0.73, 0.13, 0.13)	432
E3	(0.73, 0.17, 0.1)	460, 486
E4	(0.77, 0.1, 0.13)	405
E5	(0.77, 0.13, 0.1)	432
E6	(0.8, 0.1, 0.1)	405
F1	(0.7, 0.1, 0.2)	376
F2	(0.7, 0.13, 0.17)	432
F3	(0.7, 0.17, 0.13)	432
F4	(0.7, 0.17, 0.13)	432
G1	(0.6, 0.2, 0.2)	485
G2	(0.63, 0.17, 0.2)	432
G3	(0.63, 0.2, 0.17)	485
G4	(0.67, 0.13, 0.2)	376
G5	(0.67, 0.17, 0.17)	432
G6	(0.67, 0.2, 0.13)	486

lower bound cases. While the curves for the three remaining solutions are lower in all cases, early on. They do not drop as low as Locations 485, 486, and 460 at the tail end of the curve.

By the end of this step, the decision-makers are prepared to make a final decision as to the optimal solution. The performance of the most promising solutions from Step 2b are compared over the focused region of interest using the FWS plots. Modifications of these plots allow examination of the robustness of the solutions relative to a broader range of weights. The FWS approach provides an adaptable summary which allows straightforward comparisons between competing solutions that is suitable for varied degrees of focus on the weighting preference and diverse dimensions of problem. In the case of the two decision-makers, the summaries can facilitate a discussion about the priorities of the study and which solution to choose as optimal settings. If the priorities of the study remain, as the first decision-maker suggests with all criteria receive 20–40% weight, then Location 77 (or 78–79), with mean estimated responses (and 95% prediction intervals) of $\hat{y}_1 = 78.5(77.8, 79.2), \hat{y}_2 = 65.2(59.2, 71.2), \hat{y}_3 = 3185.5(2784.1, 3586.9)$, might be selected. If, on the other hand, it is decided that yield is most important, Location 486 (or 485), with estimated responses (and 95% P.I.) of $\hat{y}_1 = 80.1(79.4, 80.8), \hat{y}_2 = 66.6(60.7, 72.4), \hat{y}_3 = 3574.6(3180.3, 3969.0)$ using Eq. (3) could be chosen. Alternatively, if there is uncertainty or disagreement about the weightings of the three criteria, a more robust choice, such as Location 100 (or 101) with estimated responses (and 95% P.I.) of $\hat{y}_1 = 78.8(78.1, 79.5)$, $\hat{y}_2 = 66.5(60.5, 72.3), \hat{y}_3 = 3244.3(2826.6, 3620.9)$, might be preferable.

7. Discussion

While the richness of the summaries is helpful for highlighting response uncertainty and making more informed decisions, the method is computationally more demanding than just focusing on the mean model. R code for producing all of the plots in the paper is available by request from the authors. There are two major components related to the computational time necessary for this simulation-based approach. The first component, simulating 500 new response surfaces and finding the Pareto front associated with each, requires less than an hour of computation time on a standard desktop computer. The second component, finding the mixture information for each simulated front, is more time-consuming, requiring nearly 12 h for the 500 response surfaces. However, considering the high consequence of selecting the right operating conditions, investing time in obtaining a more thorough understanding of the impact of variability and estimation uncertainty on decision-making and optimization is worthwhile.

In addition, this chemical industry problem focuses on two inputs and three responses. We now describe how the various

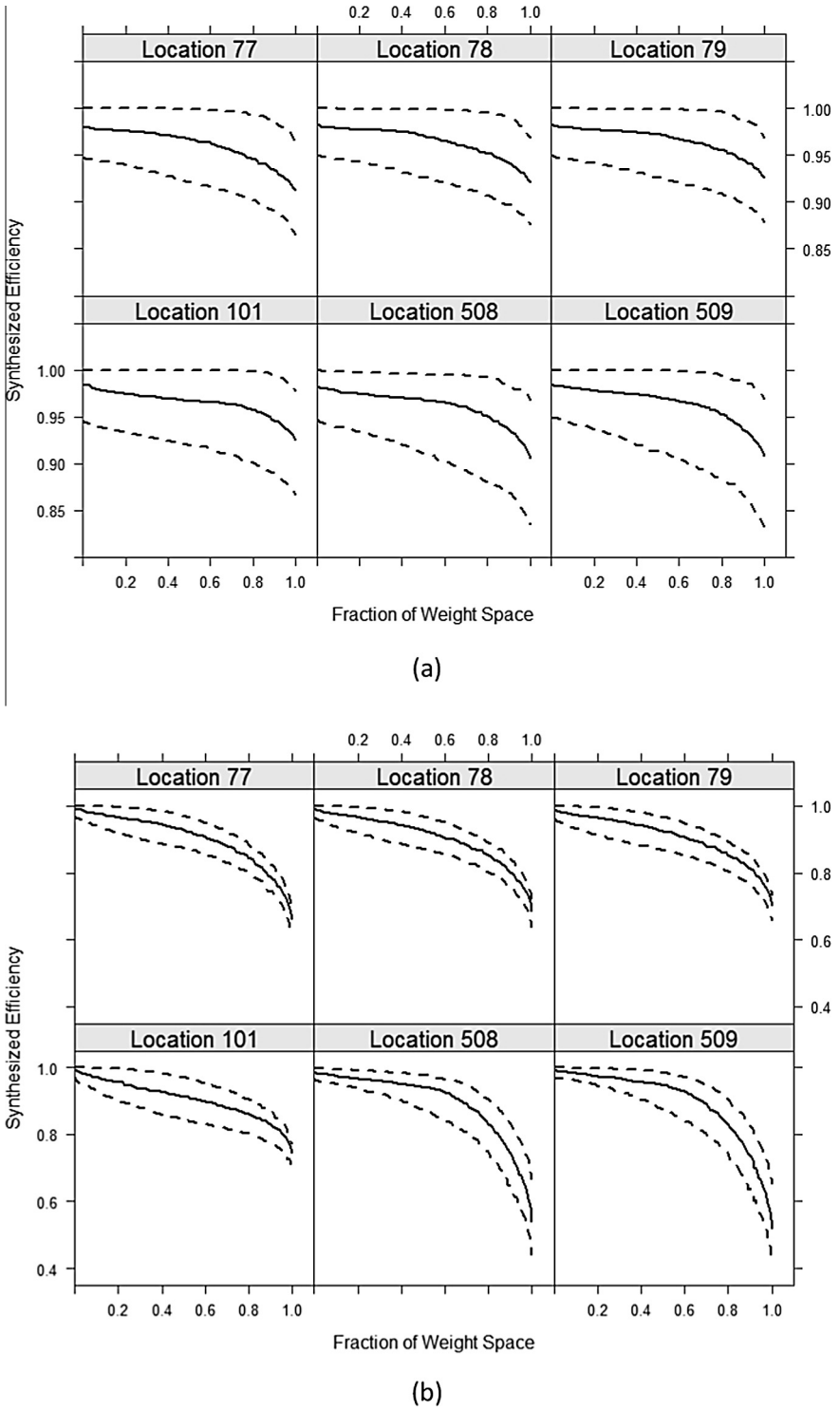

Fig. 12. (a) FWS plot for the focused weighting space for locations selected based on the weight region where all criteria receive between 20% and 40% weight. (b) FWS summarized over the entire weighting space for the same locations.

summaries scale as we change these problem attributes. In general, we suggest that the decision-maker think carefully about expanding the number of criteria/responses over which to optimize as this typically leads to increasingly severe trade-offs between choices and can lead to an overall mediocre performance on many of the responses. Increasing the number of inputs is generally not a problem, except that it complicates the visualization of the different solutions. Step 0 remains largely unchanged if the number of inputs and responses are changed. Each of the responses needs to be estimated for the grid of candidate locations for each set of generated model parameters. Similarly, if the design region has constraints then the only change required is to create a list of eligible input combinations which can be considered as possible solutions for the optimization.

In Step 1 if the number of responses increases, then the Pareto front is likely to become exponentially richer as increasing the number of criteria considered can quickly increase the total number of solutions on the front. If the number of inputs increases, then

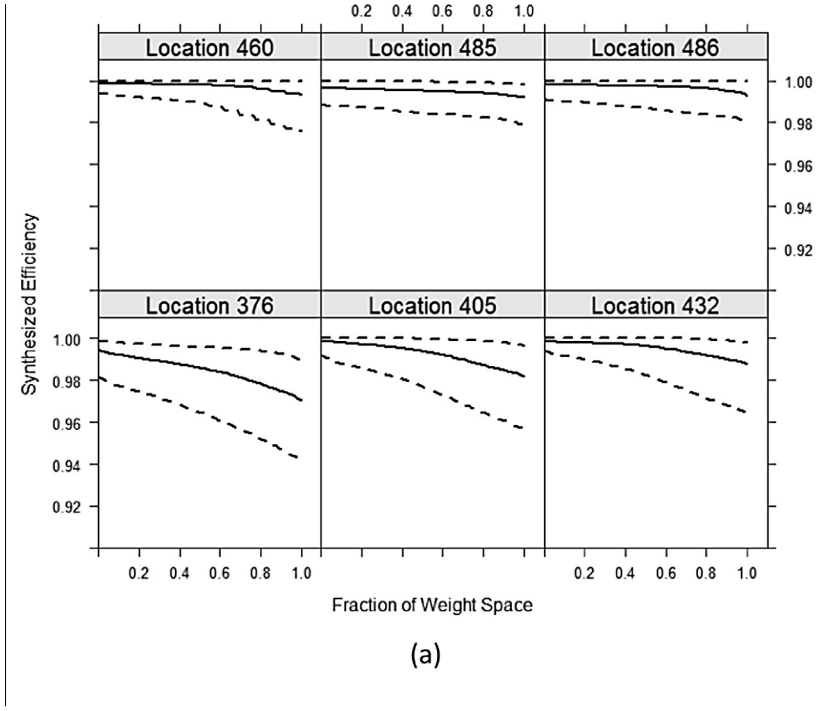

Fig. 13. (a) FWS plot for the focused weighting space for locations selected based on the weight region where yield receives 60–80% of the weight and the remaining two criteria receive 10–20% of the weight. (b) FWS summarized over the entire weighting space for the same locations.

the circle plot of Fig. 1 for visualizing the most promising solutions and their relative locations becomes more difficult. Several alternatives for summarizing this information include (a) a table with the input combinations sorted by the fraction of times that each is located on the Pareto front, or (b) creating multiple "slices" of the circle plot where some of the input levels are held fixed, and two of the inputs are shown with circle plots. This would allow summarization of the overall patterns and promising locations, but would require more interactive exploration. The pairwise plots of Fig. 2 would continue to be an option as the number of responses increases, and might be summarized with the more traditional upper triangular grid of plots.

In Step 2a, the mixture plot of Fig. 3 can be expanded to show the weight combinations for up to 4 responses with the sliced approach illustrated in Lu and Anderson-Cook (2013). Similar adaptations of the circle plot discussed for Fig. 1 would be required for Fig. 4 if the number of inputs was increased. The trade-off plot in Fig. 5 could be expanded to include one or two additional responses, however, beyond 4 responses this might become too cluttered.

The Pareto (bar) plots for the overall (Fig. 7) and specific regions (eg. Fig. 8) of Step 2b to focus on more promising locations scale well, with the major complication coming from connecting the location number to a particular input combination in a higher

dimensional design space. Similarly, the ordered frequency plot of the optimal locations for specific weights (eg. Fig. 9) is also easily adapted to either increasing numbers of inputs or responses. The FWS plots (eg. Fig. 12) in Step 2c are flexible enough to accommodate any numbers of responses and inputs.

8. Conclusions

The methods described in this paper illustrate some key features of optimizing responses when there is uncertainty about the estimated response surfaces characterizing them. In particular, it is important to remember that the mean model characterization of each response surface using the point estimates is based on our best guesses of the model parameters, but there are numerous other values which are also plausible and consistent with the observed data. Since different amounts of estimation uncertainty could be associated with different responses and the amount of uncertainty can vary across different design locations (the predicted responses are associated with more uncertainty when the location is further from the center of the design region), it is unclear how estimation uncertainty affects the decisions made when compared with the mean model solutions. Hence, when we seek to optimize several responses simultaneously, this uncertainty should be examined and its impact should be included in the decision-making process.

Using a Pareto front approach allows for inferior solutions to be removed from further consideration and then graphical summaries of the trade-offs between the remaining promising solutions can be studied and compared. Chapman et al. (2014) complements the mean model optimization by examining the worst-case prediction interval values for each of the responses at each design location. This provides a simple approach to including the variability, while keeping the amount of additional information to a minimum.

In this paper, we use simulations based on likely model parameter values to generate multiple sets of responses. By exploring how the Pareto front changes across these values, a realistic characterization of the inherent variability can be obtained, and a set of promising candidate design locations can be identified and examined for their expected performance. When we compare the results from the different approaches, we find that both the mean model approach and the worst-case prediction interval values identify similar solutions for many of the different weight combinations for prioritizing the three responses in our case study. The new simulation-based approach also highlights similar regions, but shows how model parameter uncertainty can change the range of anticipated response values at a given location and also shift the Pareto fronts. The results highlight how several adjacent locations are likely to produce desired values for a given combination of weights. Note that due to the existence of estimation uncertainty, the range of actual possible response values is wider than for the mean model. If there is uncertainty with this choice, a sensitivity analysis investigating its impact on optimization could be performed.

The outlined approach breaks the decision-making process into manageable steps that sequentially explore different aspects of the choice, moving from objectively removing less promising candidate locations to subjectively highlighting locations that are consistent with the study goals. In addition, the process transitions from general results across the entire range of weighting combinations to user-specified ranges which closely match study priorities. We think that obtaining an overall sense of the trade-offs between the responses is good calibration for the subsequent specific choices, since it highlights the synergistic or antagonistic relationship between responses as well as how robust locations are to different priorities. While the choice of which summaries to consider in the subjective steps can be tailored to individual preferences, the set of graphical and numerical summaries described in this paper provides a coherent and logical progression through the decision-making process.

Including the estimation uncertainty of the response surfaces associated with different criteria can help decision-makers better understand what input combinations are most promising for simultaneously balancing different objectives and what values are possible for each of the responses at a chosen input combination in the real world full of variability and parameter estimation uncertainty. This added realism improves understanding of what is really known about the responses and what can be expected under future production conditions.

References

Chapman, J. L., Lu, L., & Anderson-Cook, C. M. (2014). Process optimization for multiple response utilizing the Pareto front approach. *Quality Engineering, 26*, 253–268.

Chowdhury, R., Rao, B. N., & Prasad, A. M. (2009). High dimensional model representation for structural reliability analysis. *Communications in Numerical Methods in Engineering, 25*(4), 301–337.

Cornell, J. (2002). *Experiments with mixtures: Design, models, and the analysis of mixture data* (3rd ed.). New York: Wiley.

Costa, L., Espirito Santo, I.A.C.P., & Oliveira, P. (2011). Uncertainty on multi-objective optimization problems numerical analysis and applied mathematics. In: American Institute of Physics conference proceedings (pp. 775–778).

Hu, C., & Youn, B. D. (2011a). An asymmetric dimension-adaptive tensor-product method for reliability analysis. *Structural Safety, 33*(3), 218–231.

Hu, C., & Youn, B. D. (2011b). Adaptive-sparse polynomial chaos expansion for reliability analysis and design of complex engineering systems. *Structural and Multidisciplinary Optimization, 43*(3), 419–442.

Lu, L., & Anderson-Cook, C. M. (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., & Anderson-Cook, C. M. (2013). Balancing multiple criteria incorporating cost using Pareto front optimization for split-plot designed experiments. *Quality and Reliability Engineering International* (in press).

Lu, L., Anderson-Cook, C. M., & Lin, D. (2013). Optimal designed experiments using a Pareto front search for more focused desirability function weights. *Computational Statistics and Data Analysis, 71*, 1178–1192.

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

Lu, L., Anderson-Cook, C. M., & Robinson, T. J. (2012). A case study to demonstrate Pareto frontiers for selecting a best response surface design with simultaneously optimizing multiple criteria. *Applied Stochastic Models in Business and Industry, 28*, 206–221.

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

Martins, J. R. R. A., & Lambe, A. B. (2013). Multidisciplinary design optimization: A survey of architectures. *AIAA Journal, 51*(9), 2049–2075.

Mattson, C. A., & Messac, A. (2005). Pareto frontier based concept selection under uncertainty, with visualization. *Optimization and Engineering, 6*, 85–115.

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology*. Hoboken, NJ: Wiley.

Wei, D. L., Cui, Z. S., & Chen, J. (2008). Uncertainty quantification using polynomial chaos expansion with points of monomial cubature rules. *Computers and Structures, 86*(23-24), 2102–2108.

Yao, W., Chen, X., Luo, W., van Tooren, M., & Guo, M. (2011). Review of uncertainty-based multidisciplinary design optimization methods for aerospace vehicles. *Progress in Aerospace Sciences, 47*(6), 450–479.