Received: 5 June 2017 | Accepted: 15 June 2017

DOI: 10.1002/qre.2193

RESEARCH ARTICLE

WILEY

Uncertainty analysis of trade-offs between multiple responses using hypervolume

Yongtao Cao[1] | Lu Lu[2] | Christine M. Anderson-Cook[3]

[1]Department of Mathematics, Indiana University of Pennsylvania, Indiana, PA, USA

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

[3]Los Alamos National Laboratory, Los Alamos, NM, USA

Correspondence
Christine M. Anderson-Cook, Los Alamos National Laboratory, Los Alamos, NM, USA.
Email: candcook@lanl.gov

Abstract

When multiple responses are considered in process optimization, the degree to which they can be simultaneously optimized depends on the optimization objectives and the amount of trade-offs between the responses. The normalized hypervolume of the Pareto front is a useful summary to quantify the amount of trade-offs required to balance performance across the multiple responses. To quantify the impact of uncertainty of the estimated response surfaces and add realism to what future data to expect, 2 versions of the scaled normalized hypervolume of the Pareto front are presented. To demonstrate the variation of the hypervolume distributions, we explore a case study for a chemical process involving 3 responses, each with a different type of optimization goal. Results show that the global normalized hypervolume characterizes the proximity to the ideal results possible, while the instance-specific summary considers the richness of the front and the severity of trade-offs between alternatives. The 2 scaling schemes complement each other and highlight different features of the Pareto front and hence are useful to quantify what solutions are possible for simultaneous optimization of multiple responses.

KEYWORDS

global and instance-specific normalized hypervolumes, model parameter uncertainty, multiple response optimization, Pareto front, scaling choices

1 | INTRODUCTION

Since responses from most experiments cannot be simultaneously optimized, finding an appropriate balance between the performance of the responses is generally required when optimizing a product or process. A good initial step in identifying promising input combinations is to construct a Pareto front (PF)[1-5] that shows the range of possible nondominated choices of response values. The Pareto set (PS) is the corresponding collection of solutions, here, the combination of input factor values that objectively provide the best trade-off among responses. Understanding the trade-offs between multiple responses can help make an informed and justifiable decision to sensibly balance performance between the responses.

One approach to support the decision maker is to use local criteria to summarize the trade-off information[6] by examining differences between individual solutions. Another strategy is to use a quantitative summary of the degree of trade-off across all alternative solutions, like the hypervolume (HV).[7-9] This helps to understand if near-optimal values of the responses are possible, or if improvement of one response comes only at the cost of substantial compromise of the others. The HV considers both the proximity of the PF to the ideal result of

simultaneously optimizing all the criteria, as well as the richness or cardinality of the front.[10]

Historically, a single value of the HV has been reported as a fixed value for the degree of trade-off. However, when the response values have been estimated from data, this summary is too simplistic and hides the fact that the response surfaces[11] have been estimated. The response surfaces characterize the relationships between the inputs and outputs and have associated uncertainty from the natural variability in measured responses and potential lack of fit of the assumed models. Chapman et al[4] show that the PF and accompanying PS can change substantially for different estimated response surfaces consistent with the observed data. Not taking into account the impact of the uncertainty when optimizing can lead to overly optimistic assessment of optimal solution's value and location in the input space. In this work, we show how the simulation-based approach[5] for selecting an optimal solution allows exploration of the impact of the uncertainty on the degree of trade-off between responses, as captured by the HV values.

The HV of the PF based on scaling the best possible value for a given response with a value of 1 and the least desirable value with 0 is referred to as the normalized HV, ie, NHV.[10] This is similar to the scaling used with desirability functions.[12] Hence, the NHV can be thought of as the area (for 2 responses), volume (for 3 responses), or HV (for more than 3 responses) of a unit region $[0,1]^k$ (for k responses) for which attainable solutions are available. Note that the outer edge of the PF area is connected by step functions (not a smoothed line) between adjacent solutions to reflect the discontinuity of attainable solutions in the input space. If the NHV achieves a value close to 1 (near-optimal values can be obtained simultaneously for all the responses), then relatively little trade-off between responses is required, and solutions are likely to be highly desirable. As the NHV value becomes smaller, increasing trade-offs and compromise between responses are required. When the PF contains only 2 solutions, each with the best value for one criterion and the worst value for other criteria, the NHV is 0, which corresponds to the maximum trade-off between responses (optimizing one criterion requires the worst performance on the PF for others).

Figure 1 reviews some basics of PFs and illustrates the calculation of the NHV for a PF for maximizing 2 criteria. First, the PF is comprised of the 4 solid circles, which are the nondominated solutions. The remaining 7 open circles represent other possible solutions not on the PF. For each of these solutions, there is at least one point on the PF for which the criteria values are at least as good, with one criterion strictly better. The points on the PF are joined by a step-function (not by straight lines) since there are no choices between the solutions, but rather only those noted with circles. In this paper, we consider the trade-off as how much of a decrease in performance from one criterion is needed to improve in the other. For example, if we start with the best possible value for criterion 1 (at the far right of the PF), then the horizontal distance moved left to get to the next point on the PF is what we sacrifice to obtain a gain in the vertical direction for criterion 2. The triangle represents the Utopia point, the set of solution values that corresponds to simultaneously optimizing all criteria. Note that this is not an available solution in most applications.

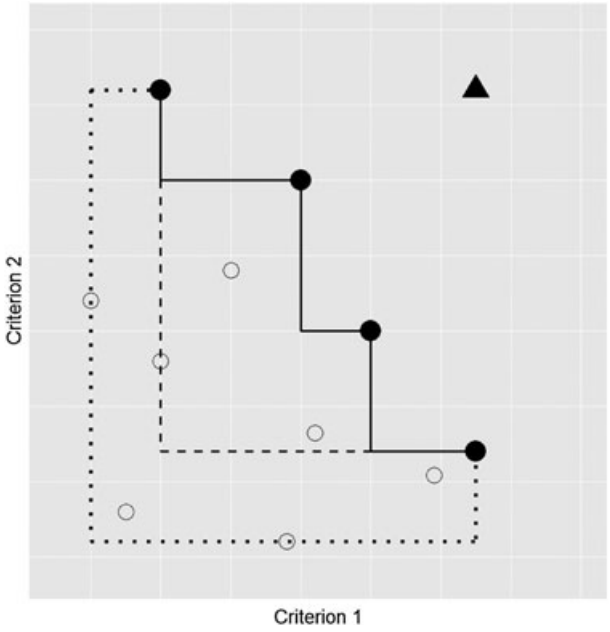

FIGURE 1 Illustration of how hypervolume can be calculated for a simple 2-criterion Pareto front. Points on the Pareto front are black filled circles. Dominated points not on the PF are open circles, and the unattainable Utopia point is the triangle in the top right corner

The dotted lines represent one possible choice of scaling for the NHV, where the scaled value of 1 for each criterion is assigned to the best achievable value, and zero is assigned to the worst observed value from all possible solutions. The NHV value for this choice is the scaled area between the dotted lines and the solid lines of the PF. A more typical choice for normalization is to keep the choice of 1 for the scaled value the same (the best observed) and to set 0 as the worst value on the PF. The dashed lines indicate the choice of 0 for this case, which leads to the NHV value for the area between the dashed lines on the bottom left and the solid lines of the PF. Note scaling using points on the PF provides a normalized quantification of the trade-off among only competing solutions. This has a different frame of reference than when the scaling is based on the entire solution space. The scaling based on the entire set of solutions usually

results in higher NHV value, which means less trade-off between the criteria. Hence, choosing an appropriate and meaningful scaling for each application leads to more precise understanding of the amount of trade-off between the multiple objectives under consideration.

For the simple example in Figure 1, the criterion values are fixed, with no uncertainty about the PF or its associated NHV for a chosen scaling. Now, consider the case of optimizing multiple responses, where each location in the input space has an estimated value of each response. There is uncertainty associated with these estimated values, which means that the points in Figure 1 might move left or right because of uncertainty from criterion 1, or up and down for uncertainty from criterion 2. Depending on how much each point might reasonably move, the shape of the PF might change, and different points might be included. A simulation-based approach[5] allows exploration of different instances of the PF and associated PS.

Anderson-Cook et al[13] describe differences in the optimization process for a single response depending on whether the goal is to maximize, minimize, or achieve a target value. When looking for extremes (maximizing or minimizing), the optimum generally occurs at a stationary point in the region, or at the perimeter of the design space. When looking to attain a target, it may or may not be possible to hit the target, and if the target value does not lie in the range of responses, then the optimization reduces to an extreme optimization. It is also possible to have more than one solution that achieves the target, if a contour for the response matches the desired value. The role of uncertainty also changes with the type of optimization. Anderson-Cook et al[14] examine patterns for dual-response problems on the basis of the nature of the optimization. In this paper, we consider the Myers et al[11(p333)] example with 3 responses, with one of each type of optimization goal.

The remainder of the paper is organized as follows: Section 2 describes the chemical experiment on which the case study is based. Section 3 outlines 2 simulation-based choices for the scaling, global and instance-based, that serve different purposes. Section 4 illustrates graphical summaries for improved understanding of the NHV for 2 responses. Section 5 extends the methods for considering trade-offs between 3 responses. Section 6 discusses general patterns observed and concludes the paper.

2 | PROCESS OPTIMIZATION BASED ON A CHEMICAL EXPERIMENT

Myers et al[11] (Table 7.4, p333) considered a chemical process involving 2 input variables ($\xi_1 =$ time and $\xi_2 =$ temperature) and 3 responses ($Y_1 =$ yield, $Y_2 =$ viscosity, $Y_3 =$ number − average molecular weight). The goal was to choose the best set of operating conditions to simultaneously maximize yield, hit a target for viscosity at 65, and minimize the molecular weight. The region of inputs explored included time between 77 and 93 minutes and temperature between 167°F and 183°F. It was believed that either a second-order or simpler model is appropriate for all responses in the design region. Hence, a 13-run central composite design[11(p393)] was conducted. The fitted response models using least squares estimation for all 3 responses are given by

$$y_1 = 79.94 + 0.995x_1 + 0.52x_2 + 0.25x_1x_2 - 1.38x_1^2 - 1.00x_2^2,$$
$$y_2 = 70.0 - 0.16x_1 - 0.95x_2 - 1.25x_1x_2 - 0.69x_1^2 - 6.69x_2^2 \text{ and}$$
$$y_3 = 3386.2 + 205.1x_1 + 177.4x_2.$$

Note that $x_1 \in (-1.5, 1.5)$ and $x_2 \in (-1.5, 1.5)$ in the above equations are the coded variables for the input factors, which were calculated from the raw values with formulas: $x_1 = (\xi_1 - 85)/5$ and $x_2 = (\xi_2 - 175)/5$.

The contour plots for the estimated response surfaces are shown in Figure 2. The maximum yield occurs around

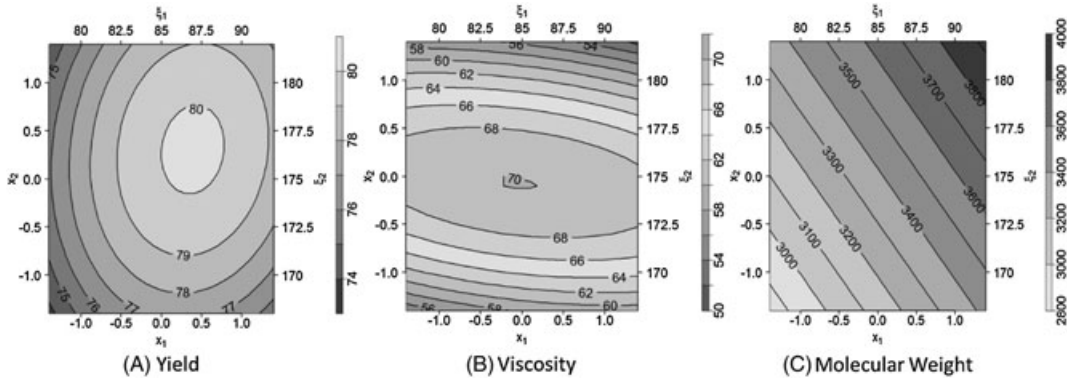

FIGURE 2 Contour plots for each of the estimates response surfaces for the 3 responses: A, yield, B, viscosity, and number-average molecular weight. Lighter colors indicate better values for each criterion

the ellipsoid region at $x_1 \in [0,0.8]$ and $x_2 \in [-0.1,0.7]$. The best viscosity (target at 65) occurs near $x_2 = \pm 0.8$. The molecular weight improves as both of x_1 and x_2 approaches their minimum values in the bottom left corner of the design space. Since optimal values for different responses are located at different regions, simultaneously optimizing all 3 responses requires trade-offs to be made between the responses, especially between yield and molecular weight, as their optima are located in opposite directions in the design space.

Note that the fitted models are based on point estimates of the parameters. Under the assumption of independent and normally distributed random errors, the sampling distribution of the least squares estimates of the model coefficients is $\widehat{\boldsymbol{\beta}} \sim MVN\left(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)$, where $\boldsymbol{X}$ is the model matrix, $\boldsymbol{\beta}$ is the vector of model coefficient parameters, and $\widehat{\boldsymbol{\beta}}$ denotes the vector of the least squares estimates. Chapman et al[5] proposed a simulation approach for quantifying uncertainty associated with the fitted responses surfaces and the multiple response PF. In particular, a large set of model coefficients are simulated from $MVN\left(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)$, which are all consistent with the observed data. Then, the response surfaces based on the simulated coefficients are used to identify possible PFs, and quantifying the trade-off with its associated uncertainty. Since the preferred choice depends on how we value the trade-off between the different response values, Section 3 introduces 2 different scaling choices sensible for the simulated-based approach that serve different purposes for understanding the uncertainty associated with the optimization.

3 | QUANTIFYING TRADE-OFFS USING HV WITH DIFFERENT SCALING CHOICES

In this section, we present concepts related to the HV and scaling issues for normalizing the HV in the simulation setting when optimizing multiple responses with associated uncertainty. For cases of maximizing or minimizing, the response values are used directly. For achieving a target value, a new response (to be minimized) is constructed by using the distance between the response and the target. The idea of HV is documented in the receiver operating characteristic[15,16]) and HV literature. But to avoid arbitrary HV values, the NHV is preferred in practice. More importantly, as discussed in Section 1, the NHV has a useful interpretation when quantifying trade-offs between responses.

To calculate the NHV under the PF, a scaling is selected and new scaled values of the criteria constructed. For values y_{best} to be assigned a value of 1 and y_{worst} to be assigned 0, a transformation of the following form

$$Z_i = \frac{y_i - y_{worst}}{y_{best} - y_{worst}} \tag{1}$$

is used for all points on the PF. This creates values of Z_i between 0 and 1. If it is possible to obtain values of y_i outside of the range defined by y_{best} and y_{worst}, then these values should be assigned a 1 or 0 value, depending where they are outside of the range. This allows for all criteria to be scaled in the interval [0,1]. Since the final decision is made from the Pareto optimal set, the trade-offs associated with this set of solutions are directly relevant in seeking appropriate balance between different responses. We recommend the normalized scaling using the worst observed values from the PF to define 0 for each criterion, and the values of the Utopia point (the best values for all criteria) to correspond to the value 1, as it strictly quantifies the trade-offs among competing solutions.

For the two criteria case, there is a unique ordering of the points on the PF. We can construct a series of rectangles between the PF and the axes of origin with distinct values of one criterion used as dividing points to define the set of rectangles. The NHV is then the sum of the rectangle areas. This transformation restricts the value of Z_i to lie between 0 and 1; the NHV represents the fraction of the unit square below the PF with attainable solutions.

Adaptations to the normalized scaling are needed for the simulation-based approach. We define 2 ways of scaling to associate an NHV value with the uncertainty. To simplify the discussion, we assume that the responses to be optimized all involve maximization. For minimization or hitting a target, negative values of the quantity to be maximized can be used.

First, we consider a "global scaling," where we look across all of the simulated PFs to find the overall maximum and minimum values. These are assigned as y_{best} and y_{worst}, respectively. The transformation from the original values to the scaled criteria defined in (1) is applied to obtain values in [0,1]. The NHV found for each simulated PF with this scaling is called the "global NHV (GNHV)."

In contrast, another version is to scale based on only the range of individual simulated PF, which is referred to as the "instance-specific scaling." The y_{best} and y_{worst} values are identified on the basis of each individual simulation. Hence, for this scaling, there are guaranteed to be some solutions in each simulation receiving the extreme values 0 or 1 for at least one response. The transformation in (1) is again applied to obtain scaled criteria values. The NHV constructed from these values is called the "instance-specific NHV (ISNHV)."

More in-depth comparisons between the GNHVs and ISNHVs considering pairs of responses as well as all 3 responses from the chemical experiment example are presented in Sections 4 and 5. Here, we describe a few important differences between the 2 scalings. First, because the PF is scale-independent, the same solutions with the same response values are used for each scaling. In this way, we summarize the same information and differences in the GNHV and ISNHV values are attributable to only the choice of scaling.

For each simulation, the attainable ranges from the 2 scalings are quite different. For the ISNHV, the best value of each individual response will be mapped to 1, and the worst to 0. Hence, the available range for each simulation spans the entire range. For the GNHV, we expect that only the simulation(s) with the overall best observed value of the response have a possible value of 1. Similarly, only the simulation(s) that contain the overall worst observed value achieve a value of 0. With this scaling, the best and worst values are bounds for the most optimistic and pessimistic conditions. Hence, for the GNHV, we anticipate that for many simulations, the range of values spans only a portion of the [0,1] range.

In terms of the variability that we anticipate between the GNHV and ISNHV across simulations, the values of GNHV are more likely to be consistent, since the same scaling is universally applied. The ISNHV can have dramatic differences in the values since the chosen best and worst values are based on just a single PF and are likely to fluctuate more.

When we compare the PF for the mean model of each response (using the point estimates for all the model parameters) to the results from the global and instance-specific scalings, we anticipate some differences as well. If we consider each simulation as a plausible set of values consistent with the observed data, then the PF for each simulation gives us a sense of how much variability we might anticipate for the optimization. The ISNHV values can be thought of as characterizing the range of what might be possible when looking at just one result at a time. We have a single number associated with the mean model NHV, and the ISNHV quantifies how consistent that the mean model would be if we had run another experiment from the same process. When we compare the mean model NHV to the GNHVs, we are more likely to see that the mean model NHV is centrally located relative to the simulation values. It is unlikely to have best and worst values that span the entire normalized range between 0 and 1, since the mean model rarely achieves these extremes when compared to the entire population of simulations. In Section 4, we revisit these findings in the context of specific examples from the chemical process case study.

4 | EXAMINING HV PATTERNS FOR 2 RESPONSES FROM THE CHEMICAL PROCESS EXAMPLE

In this section, we examine patterns in the NHVs of the PF across simulations for pairs of responses using the global and instance-specific scaling schemes. This considers pairwise trade-offs between the responses, and the associated uncertainty attributable to the natural variability of the responses and the model fitting.

Anderson-Cook et al[13] demonstrated different patterns and complexity of impacts of types of optimization on the uncertainty of the selected optimal solutions for a single response. When the goal involves maximization or minimization, there is usually less uncertainty associated with the identified optimal solutions compared with hitting a target. There are often more varied possibilities for the latter case and the identified solutions may not be even in adjacent regions. Anderson-Cook et al[14] further explored how identified solutions are affected by the shapes of the response surfaces (mountain, valley, or ridge), the relative location of the individual optimal regions and different signal-to-noise ratios across types of optimization for known response surfaces without considering estimation uncertainty. The type of optimization plays an important role for both the inherent trade-off between the responses as well as its uncertainty. In this section, we explore patterns in the NHV of the simulated PFs for pairs of responses, which represent three typical combinations of optimization objectives: maximization vs minimization (Y_1 and Y_3), maximization vs target (Y_1 and Y_2) and target vs minimization (Y_2 and Y_3).

4.1 | Optimizing Y_1 = yield and Y_3 = number − average molecular weight

First, we consider optimizing Y_1 (maximizing) and Y_3 (minimizing) simultaneously. Figure 3 shows a scatterplot of GNHV vs ISNHV values across 1000 simulations, with marginal histograms for each type of NHV along the axes. The mean model NHVs are shown with black lines on the histograms, and a solid black point in the scatterplot. Several interesting patterns can be observed. First, from the marginal histograms, the mean model GNHV is around 0.52, which is considerably smaller than the mean model ISNHV at around 0.66. This is because the global scaling uses a wider global range for the responses across all simulated PFs as the unit scale, so the relative area of the obtainable solutions under the mean model PF is naturally smaller with the larger unit square under the global scaling.

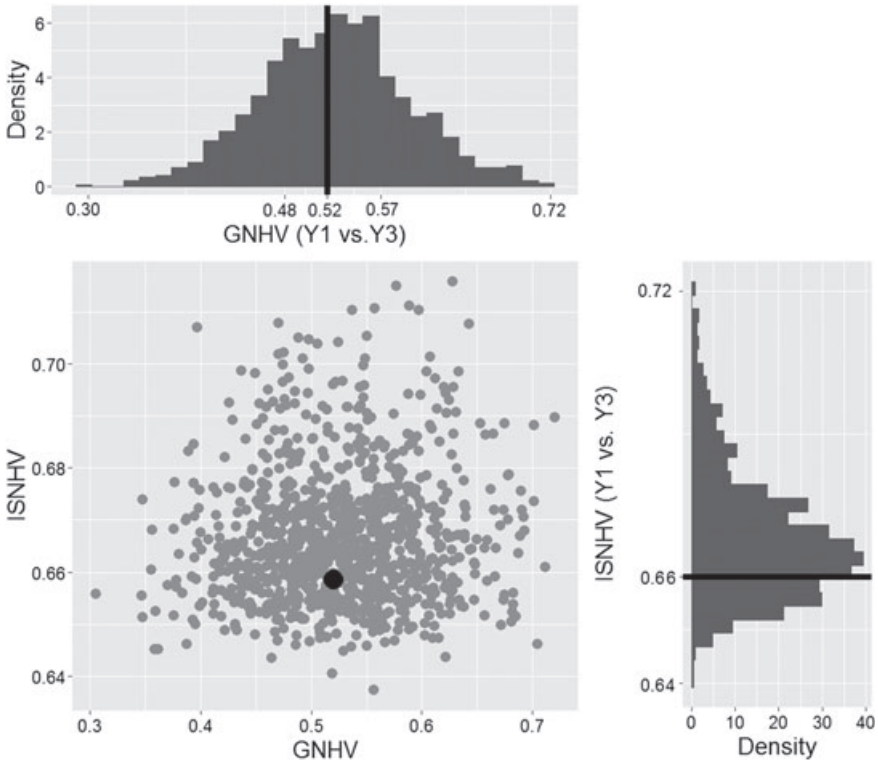

FIGURE 3 Scatterplot and distribution of global and instance-specific hypervolume values for the PFs for yield (Y_1) and molecular weight (Y_3)

Second, if we look across the NHV values for the simulated response surfaces, the distribution of the GNHV values is more symmetric and the mean model GNHV is located closer to the center of the distribution than for the ISNHV distribution, which is skewed towards larger values. This matches our intuition about the unified scale of the global scaling, with the mean model naturally capturing the central tendency of the overall distribution. The ISNHVs are a function of independent choices of ranges and scaling, which leads to the mean model just being a particular instance that could be anywhere in the set of simulated values.

Third, the GNHV values range between 0.3 and 0.72, which is much wider than the range of the ISNHV between 0.63 and 0.72. In this case, there is less variation in the size of the area under the individual PF relative to the unit square bounded by its own range of values. The larger variation in the GNHV comes from more changes in the location of the individual simulated PFs in the unit square bounded by the extreme observed response values.

Next, we consider the correlation between the 2 NHVs. For Y_1 and Y_3, the correlation is close to zero ($r = .065$), indicating virtually no connection between the 2 values. When we consider what the 2 measures are capturing, this lack of association becomes a bit more intuitive. The value of the GNHV is largely capturing the location of the PF relative to the simulation extremes. Hence, a PF that is relatively close to the combination of y_{best} values observed across all simulations results in a high GNHV value. The value of the ISNHV is not connected to any other simulations, and hence, this measure focuses on the degree of trade-off and diversity of available solutions within that single simulation. A front able to simultaneously approach the y_{best} values for that simulation results in a high ISNHV value.

We can see that with different scaling choices, the uncertainty of the NHV values are affected by different factors. For the ISNHV, the areas with attainable solutions are calculated relative to the range of individual PF values. Hence, the uncertainty of the ISNHV characterizes the variability in the degree of trade-off across all solutions on the PF, which is largely tied with the richness or cardinality of the PF. However for the GNHV, due to the use of the universal scaling bounded by the extreme values from all the simulations, the uncertainty of the GNHV is primarily affected by the relative position of the individual PF among all possible PFs. If the responses include at least one with large variation such as Y_3, the change of the PFs across all simulations can be large. In this case, the degree of trade-off is largely affected by the front cardinality and results in relatively small variation across simulations, as evidenced by the consistent patterns among sampled PFs from Figure 4. There is more variation with the GNHV values than the ISNHV values.

Figure 4 shows a few representative samples of the simulated PFs across the simulations for the different scalings. Throughout the remainder of the paper, all

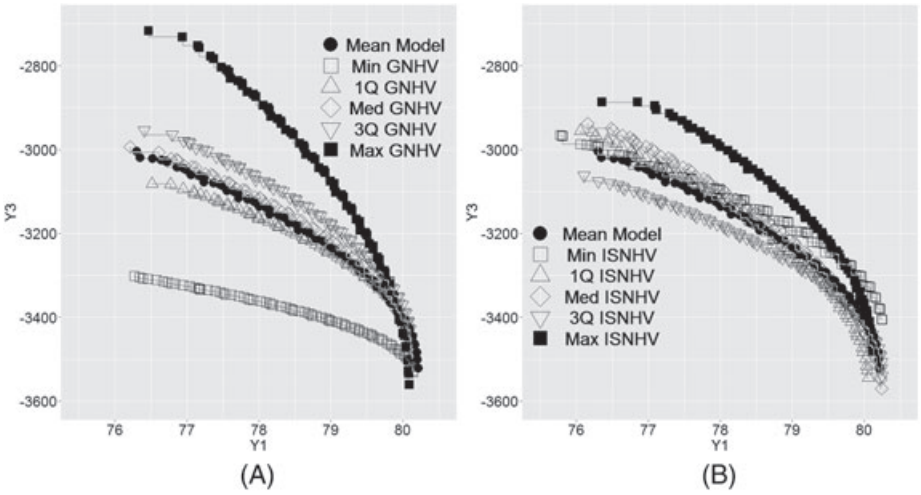

FIGURE 4 Samples of Pareto fronts from the simulations for optimizing yield (Y_1) and molecular weight (Y_3) based on A, global scaling and B, instance-specific scaling

figures transform the responses to convert the optimization to maximization. For ease of presentation, we convert minimizing Y_3 to maximizing $-Y_3$ and turn targeting Y_2 to 65 to maximizing $-|Y_2-65|$. The mean model front and the fronts corresponding to the five number summary (minimum, 1st quartile, median, 3rd quartile, and maximum) based on the NHVs are shown with different symbols. For the global scaling, all sample PFs overlap in the bottom right corner corresponding to near optimal (at or above 80) value for yield (Y_1) but with a poor (greater than 3500) value for the molecular weight (Y_3). The top left end of the PF moves upwards as the sample percentile increases. Hence, the value of the GNHV is largely driven by the estimated best attainable Y_3. The minimum and maximum GNHV fronts bound the range. The pattern matches Figure 3 with the near symmetric distribution of GNHV around the mean model value. However, when the instance-specific scaling is used, the sample fronts show more differences on both ends. The relative positions of the fronts are no longer related to the natural order of the percentile summaries, since the size of the ISNHV is not directly related to the location of the PF relative to other simulations. For instance, the 3rd quartile ISNHV front is below all other fronts except for the bottom right part, while the minimum ISNHV front is consistently above the 3rd quartile ISNHV front (ie, better Y_3 values). But since it has a much narrower range of values for Y_3 and a sparser front, it results in smaller relative area below the PF for the instance-specific scaling.

To connect the different PFs with the corresponding PS of optimal solutions, Figure 5A shows the PS for the mean model as well as the minimum and maximum NHVs using both scalings (Figure 5B-E). The PS for the mean model forms a fan-shaped region from close to the center region to the bottom left edge of the design region for the original experiment. Recall from Figure 2 that the best values of yield are located near the center of the region, while the best values for molecular weight are in the bottom left corner. Hence, for the mean model, the locations associated with the top left corner of the PF in Figure 4 are at the edge of the design region, the bottom right of the PF are near the center.

The PS for the minimum ISNHV is also a fan-shaped region but shifted clockwise, while the PS for the maximum ISNHV shifts counterclockwise. By overlaying the PSs with the contour plots in Figure 2, we see that all sets have similar ranges for Y_1 but have considerably different ranges for Y_3. In particular, the PS for minimum ISNHV is associated with a narrower range of Y_3 with all larger values (smaller/better values on the original scale), while PS for maximum ISNHV is associated with a wider range of Y_3. This matches with the corresponding sample PFs shown in Figure 4. The PS for the minimum and maximum GNHVs show similar patterns except that the PS for the minimum GNHV is closer to the PS for the mean model.

From both Figures 4 and 5, we can see that the uncertainty of the PF for Y_1 and Y_3 and its associated PS seems to be affected more by the variance from the estimated response surface for Y_3. This is explained by the much larger square root of the mean square error (MSE) (165.6 for Y_3 compared with 0.27 for Y_1) and smaller signal-and-noise ratio (4.6 for Y_3 compared with 16.7 for Y_1) that are associated[5] with Y_3 than Y_1. Figure 6 shows the range of the response values across all 1000 simulated PFs. The plots for Y_1 (top panel) and Y_3 (bottom panel) are both sorted on the basis of the minimum values (corresponding to the worst values) on the PF. We can see that there is little variation for the maximum (best) values on the PF for Y_1. Almost all simulations have the maximum Y_1 values at around or above 80. Alternately, the maximum

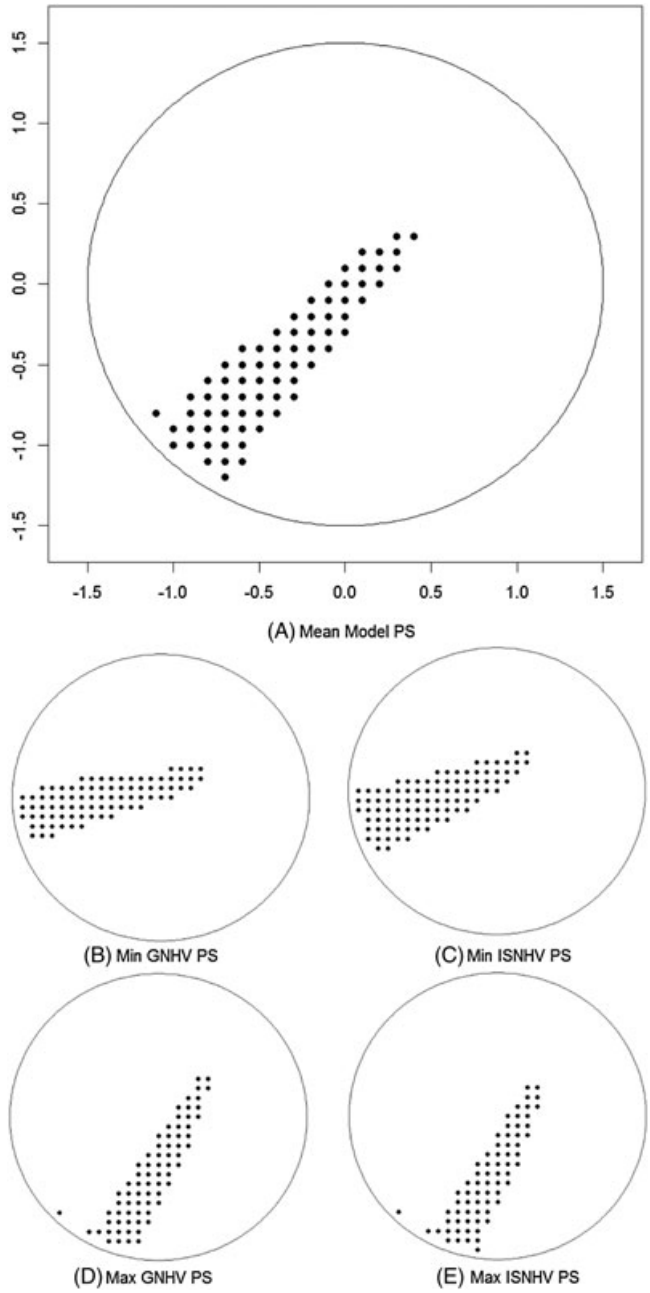

FIGURE 5 Pareto sets for yield (Y_1) and molecular weight (Y_3) for the mean model, and the minimum and maximum for each of the GNHV and ISNHV from the simulations

values for $-Y_3$ (corresponding to the best values) on the PF vary much more, which spreads over at least half of the overall range of the Y_3 values on the PFs across simulations.

4.2 | Optimizing Y_1 = yield and Y_2 = viscosity

In Section 4.1, we examined an optimization involving maximizing and minimizing. We now consider a different scenario that involves hitting a target value for Y_2 = viscosity, while maximizing Y_1 = yield. Figure 7 shows the scatterplot of GNHV vs ISNHV across the simulations with marginal histograms for each type of NHV values. Again, the mean model values are shown with black lines and a solid black circle. From the marginal histograms, the GNHV values range is approximately symmetrical between 0.824 and 0.991, with the center located around the mean model value at 0.92. In contrast, the ISNHV values are highly skewed and cover almost the entire range of possible values between (0, 0.991). A large number of simulations resulted in ISNHV values above 0.8, which are substantially larger than the mean model value of 0.635. This indicates a relatively small trade-off between the responses for many of the simulations, but there are also a small number of cases have a very large trade-off. From the scatterplot, there is again little correlation between the 2 types of NHV summaries (0.068). This is not surprising given that the 2 NHV summaries quantify different characteristics of the anticipated PFs.

Compared with the first case of optimizing Y_1 and Y_3, there are some similarity in the observed patterns, including the symmetry of the GNHV distribution, the more skewed distribution for the ISNHV, and near zero value of the correlation between the two NHV values. But we also observe some differences. First, the mean model GNHV for Y_1 and Y_2 (0.92) is substantially higher than the mean model ISNHV (0.635). It is due to the inclusion of the target value for Y_2 for almost all estimated response surfaces. Hence, it is possible to achieve a 0 value for $|Y_2 - 65|$ at a number of different locations in the input space (on the contour for 65), which allows for choices for Y_1. This leads to less trade-off and larger area under the PF. In addition, the mean model PF has end points relatively close to the global ideal values and so it compares favorably to the larger unit square. This can be clearly seen from the sampled PFs (including the mean model) shown in Figure 8, where proximity to the target value of 65 for Y_2 provides many choices from which to pick the preferred value for Y_1. Second, the range of GNHV is much smaller than ISNHV for the Y_1 and Y_2 pair, in contrast to the opposite pattern for Y_1 and Y_3. This is because the simulated PFs consistently contain a set of solutions close to the Utopia point in the top right corner of the unit square but have more variation on the ends of the anticipated PFs. For GNHV, the variation on the end points does not make much difference with the unified scaling based on the extreme values. However, ISNHV is highly dependent on the end points of the individual PFs. This leads to the large variation in choice of scaling for individual simulation cases, which translates into large variation of the ISNHV values.

An important contributor to the large variation in the end points of the PF is the sparsity of the PF associated with

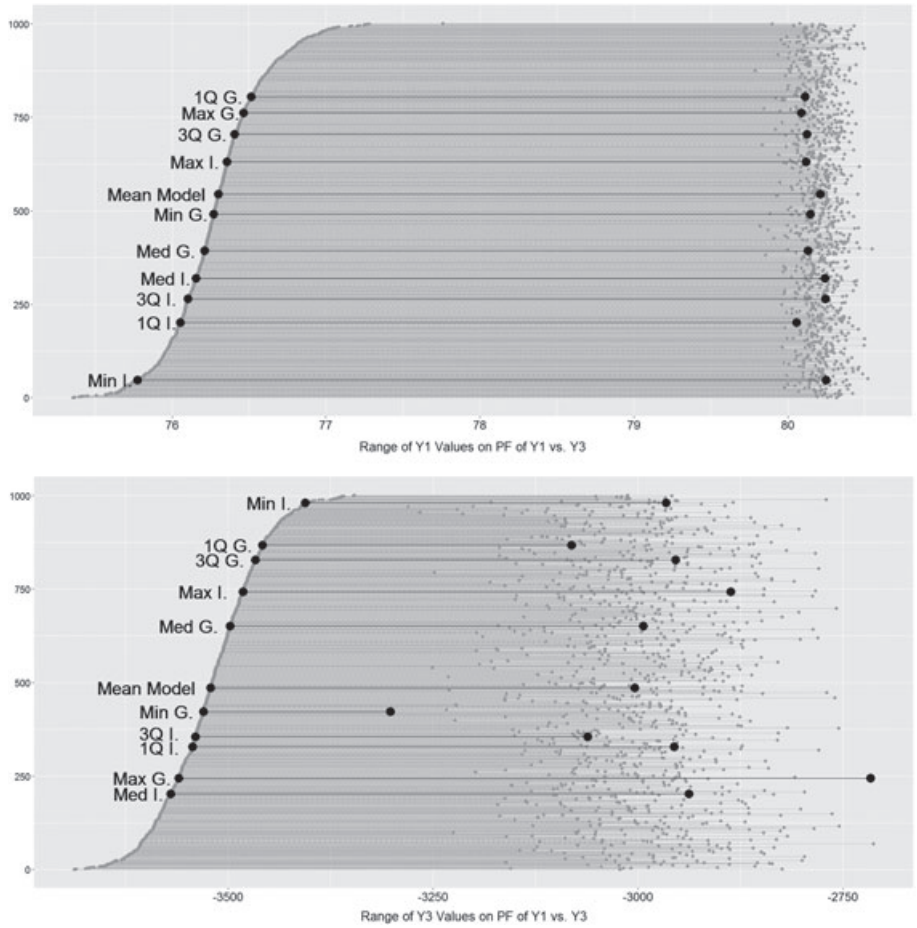

FIGURE 6 Ranges of yield (y_1) and molecular weight (y_3) for the Pareto fronts across the simulations

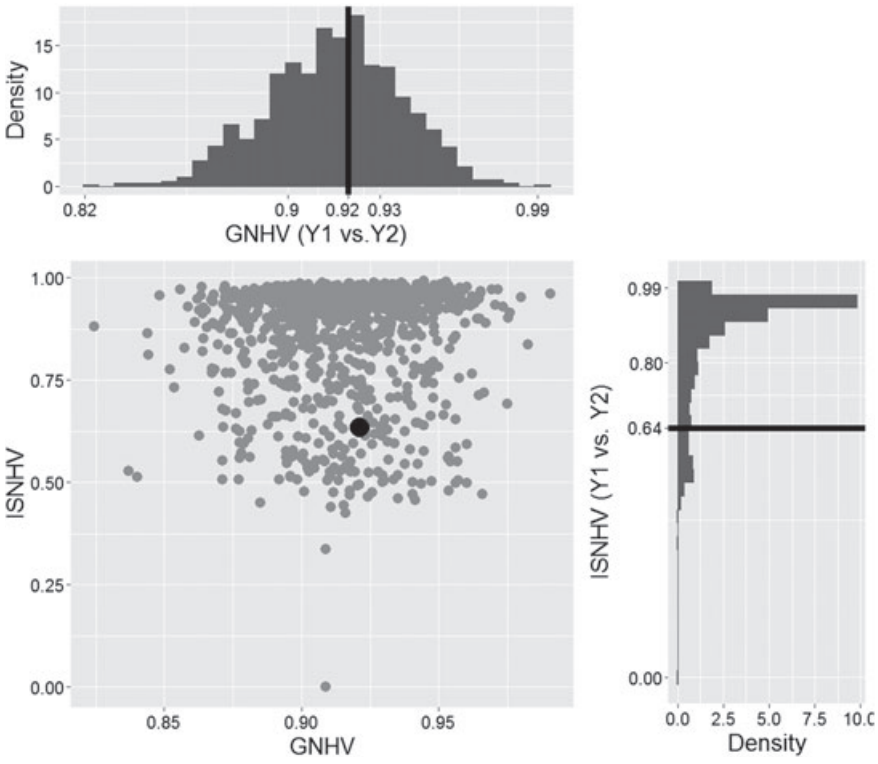

FIGURE 7 Scatterplot and distribution of global and instance-specific hypervolume values for the PFs for yield (Y_1) and viscosity (Y_2)

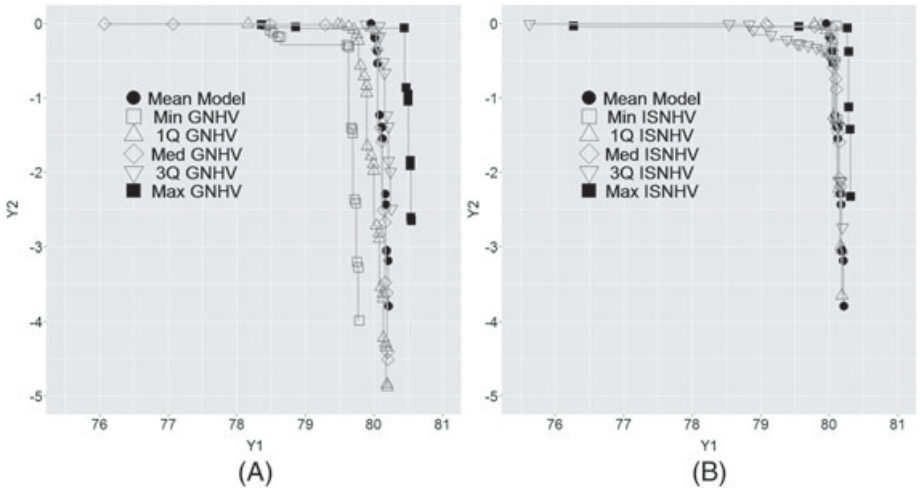

FIGURE 8 Samples of Pareto fronts from the simulations for yield (Y_1) and viscosity (Y_2) based on A, global scaling and B, instance-specific scaling

the Y_1 and Y_2 responses. This is strongly connected to one of the optimization goals being to hit a target and hence results in more discontinuity of the solution sets. This can be also observed from Figure 8, where the sampled PFs contain fewer solution points (compared with Figure 4 for Y_1 and Y_3) especially towards the ends of the PFs. A change in a single value on the ends of the PF can lead to big changes in the instance-specific scaling and hence the large variation for the area under the PF across the simulations.

Figure 9 shows the PSs of optimal solutions for the mean model and for the minimum and maximum NHVs for the 2 scaling choices. The PS for the mean model forms a contiguous triangle-shaped region connecting the optimal location for maximizing Y_1 and the upper part of the ellipse around the target value of Y_2. Note that the locations around the lower part of the ellipse close to target value for Y_2 are not included in the PS, as they are generally associated with poorer Y_1 values for the mean model (and hence are dominated by locations around the upper ellipse with similar distance to the target of Y_2). However, as estimation uncertainty is incorporated, the ellipse where the target value is achieved can shift up and down. This can result in different ranges of Y_1 values to be included on the PF. For example, when the optimal regions for the simulated response surfaces for both Y_1 and Y_2 are very close, the PS contains very few solutions, such as the PS shown in Figure 9C which contains only 2 solutions on the PF and thus results in maximum trade-off (ie, minimum NHV of zero based on the instance-specific scaling). As the optimal regions for both responses are separated further, more solutions are included on the PF, even including the locations around the lower part of the ellipse such as in the PS shown in Figure 9B, 9D, and 9E.

From Figure 9, the PS for the mean model contains solutions closer to the Utopia point that are more likely to be included consistently across simulations. However, completely relying on the mean model approximation might miss some possible regions containing optimal locations for some simulations, especially when unequal weighting between the responses is preferred. In addition, when optimization involves hitting a target, there is often considerable discontinuity in the possible PFs from the simulations and large variation in the identified PS regions. In this case, the ISNHV can have much more variation than the GNHV.

Figure 10 shows the range of response values across all simulations. Despite the fact that Y_1 has the largest signal-to-noise ratio (and smallest MSE) among all 3 responses, it is associated with a larger range of optimal response values than Y_2 for which the goal is set to hit a target value. The optimal or near-optimal values (close to zero distance to the target) are almost always included on the PF, unless the target is unattainable in the design space.

As a summary, there is generally less trade-off required between the responses Y_1 and Y_2, given the frequent proximity of some of the optimal regions for the 2 responses. However, when achieving a target is an objective, the uncertainty in the response surfaces can lead to the PS having some discontinuity, with isolated regions in the input space. We also see that the PFs have fewer points and larger differences in response values between adjacent PF points. Hence, there is substantial variation across simulations of anticipated PFs, particularly towards the ends of the PF (with solutions doing well on one response but more poorly on the other). This translates into large variation in the ISNHV, since there are large differences between the chosen instance-specific scalings. On the other hand, the global scaling is more robust to variation at the ends of the PFs by using a consistent scaling including the most extreme cases.

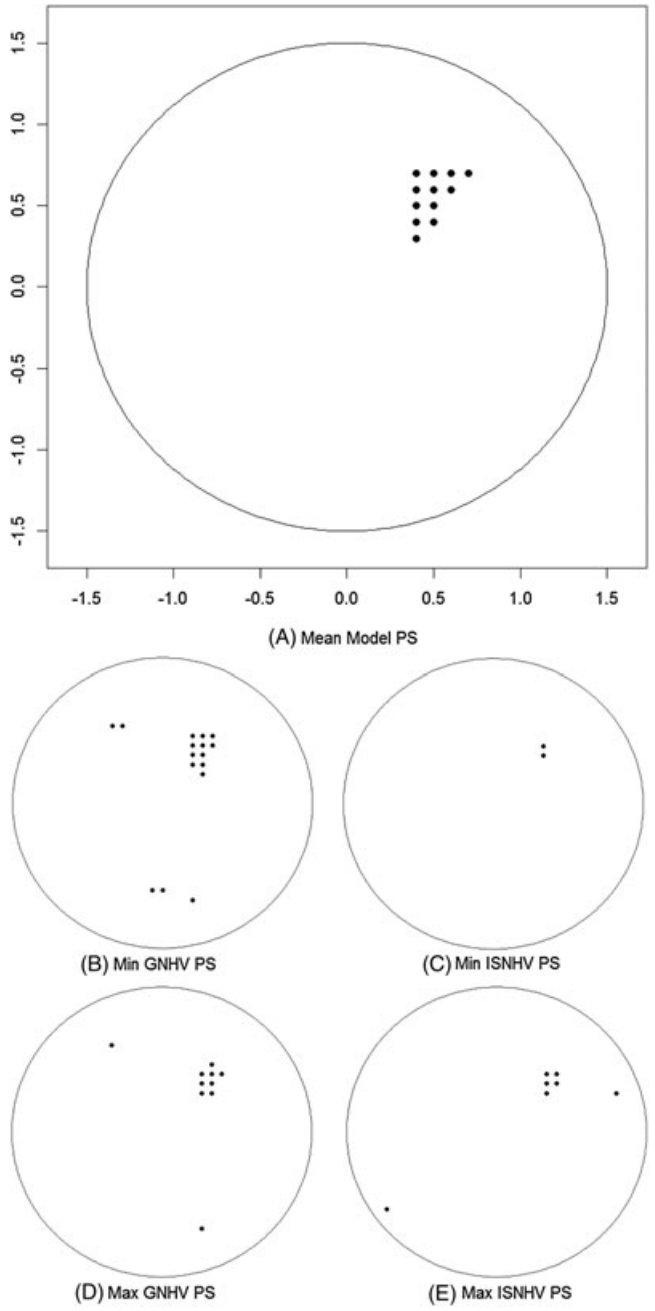

FIGURE 9 Pareto sets for yield (Y_1) and viscosity (Y_2) for the mean model and the minimum and maximum for each of the GNHV and ISNHV from the simulations

4.3 | Optimizing Y_2 = viscosity and Y_3 = number − average molecular weight

The last two-response optimization considers Y_2 and Y_3, which again considers a target combined with a minimization. Similar figures as were used in Sections 4.1 and 4.2 are included in Appendix A. For this optimization, we observe patterns consistent with our understanding from the earlier scenarios. For example, the GNHV values between 0.5 and 0.99 are symmetrically distributed around the mean model value of 0.75. The ISNHV values cover the entire range of possible values and are skewed towards smaller values. The mean model ISNHV is 0.96, which indicates the mean model PF contains solutions with response values close to the ideal values on that PF. As before, there is little correlation (−0.036) between the GNHV and ISNHV (observed by the scattered gray points in Figure A1).

Similar to the optimization of Y_1 and Y_2, where achieving the target for Y_2 is generally achievable in the design space, the best values of Y_2 for all of the simulated PFs are at or near 0, which bounds one end of the PF (the bottom right corner in Figure A2). Large uncertainty for Y_3 results in a large range of best Y_3 values across the simulations. This is seen with large variation in upper part of the simulated PFs (at the top left corner of Figure A2(A) and a considerable range for the maximum $-Y_3$ values in Figure A4).

The sparsity of the solutions on the PF leads to large variation in the estimated Y_3 values and results in a range of values at the top left end of the PF (corresponding to the best solution on Y_3). Differences in the number of solutions on the PF result in variability for the instance-specific scalings across the simulations. From Figures A2 and A3, we see that the ISNHV covers the extreme scenarios of trade-off relationship between a pair of criteria. The minimum ISNHV is obtained when the simulated PF has only 2 points (open squares in Figure A2B) and 2 input locations (Figure A3C) in the PS.

The maximum ISNHV is obtained when the simulation PF is comprised of a single point (black solid square in Figure A2B) and a single location in the PS. Here, both responses can be optimized simultaneously and no trade-off is needed. Despite the largest possible difference in the ISNHV values, the actual PS locations in Figure A3 are similar. For the extreme GNHV values (Figure A3B and A3D), the difference in the PS locations reflects the range of the Y_3 values.

5 | SIMULTANEOUSLY OPTIMIZING 3 RESPONSES

In this section, we consider optimizing all 3 responses simultaneously. Recall the general result that when we are balancing more criteria with a higher dimensional PF, usually, more trade-off between the responses is required. Figure 11 shows the pairwise scatter plot of the GNHV and ISNHV values with their individual marginal distributions. The GNHV for the 3 response PF ranges between 0.42 and 0.74, with the mean model GNHV near 0.54. The ISNHV has a narrower range between 0.53 and 0.77, with the mean model ISNHV takes value at 0.65. Both GNHV and ISNHV distributions are close to symmetric. However, the mean model NHVs are not located at the center of the distributions but are closer to the smaller values. This indicates that the mean model

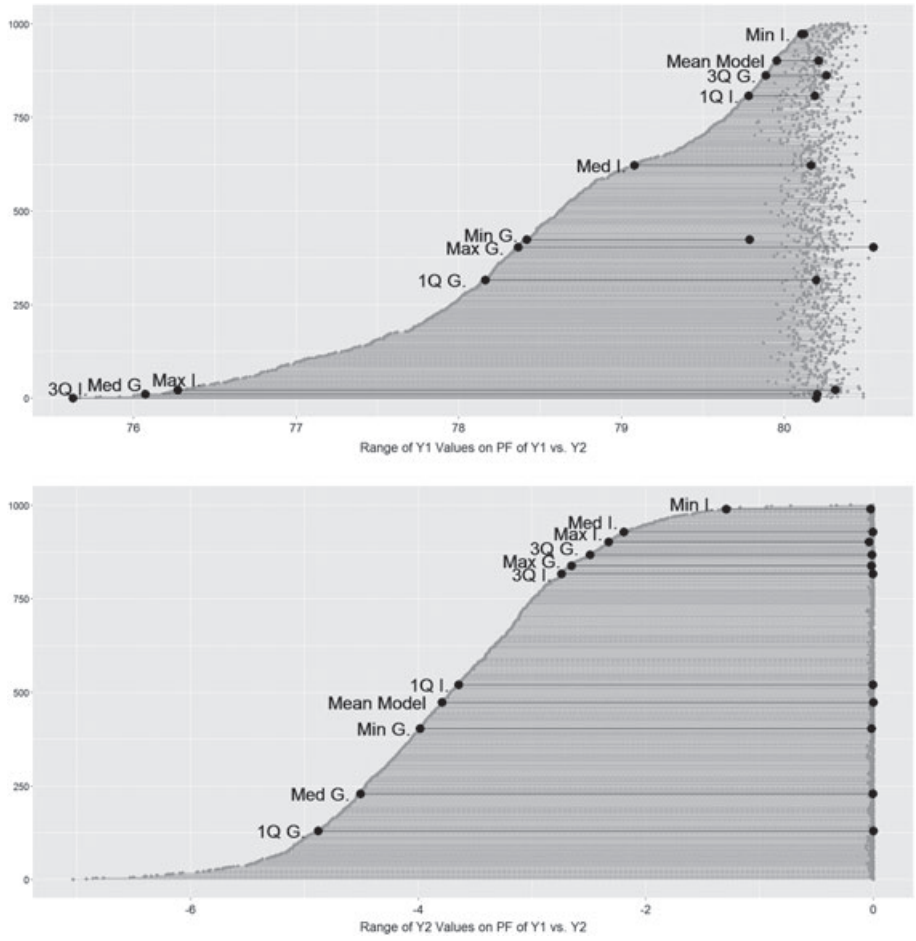

FIGURE 10 Ranges of yield (Y_1) and viscosity (Y_2) for the Pareto front across the simulations

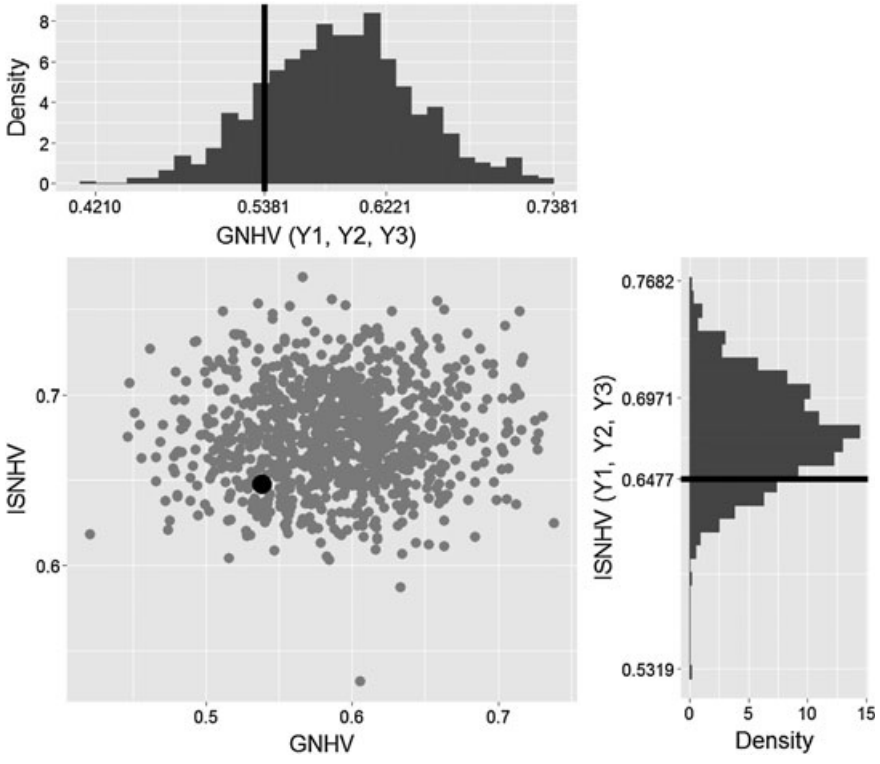

FIGURE 11 Scatterplot and distribution of global and instance-specific hypervolume values for the PFs for all 3 responses

response surface location is slightly farther than typical from the global scaling Utopia point from all simulations.

The GNHV has a slightly wider range of values because of the larger variation of the relative location of the simulated PFs within the larger unit square used for the global scaling than the variation of richness of the PFs reflected in the ISNHVs. Since the 3 response PFs are points in a 3-dimensional space, there are more opportunities for trade-off between all three of the responses. This leads to a richer front with adjacent solutions being more similar.

The differences between the GNHV and ISNHV for 3 responses are smaller than the two-response scenarios described in Section 4, because the global scaling range is defined by the most extreme values across all simulations. Since the 3-dimensional PF must contain the best value for each response, this potentially leads to more mediocre values of the other 2 responses at that location, which extends the ranges for the global scaling. Table 1 contains a summary of the worst values for each of the responses for the scenarios considered. The lower worst case values have an impact on the GNHV by adding some area at the bottom of the HV but do not impact the values for the ISNHV. In addition to the extending of the ranges of worst values considered, the PF for the three-response optimization is also notably richer, since trade-offs between all responses need to be considered. As noted with some of the 2 response systems, the higher cardinality or richer fronts have more stable ISNHV values. The sample correlation between GNHV and ISNHV is 0.065, which matches the pattern of observations in the scatterplot in Figure 11.

Figure 12 shows the PS corresponding to the mean model as well as the minimum and maximum for the 2 types of NHVs. The PS for the mean model response surfaces forms a bow tie–shaped region. The bottom of the bow tie has more solutions located near the bottom left corner of the design region with near optimal values for Y_3 and Y_2 but worse values for Y_1. When the global scaling is used, the PSs corresponding to the extreme GNHV values are relatively similar in shape to the mean model PS with some local changes (eg, the PS for minimum GNHV shifts towards smaller values on the horizontal axis). However, when the instance-specific scaling is used, the PS locations for the 2 extreme cases show larger differences from the mean model PS with different regions of the input space and the overall shape.

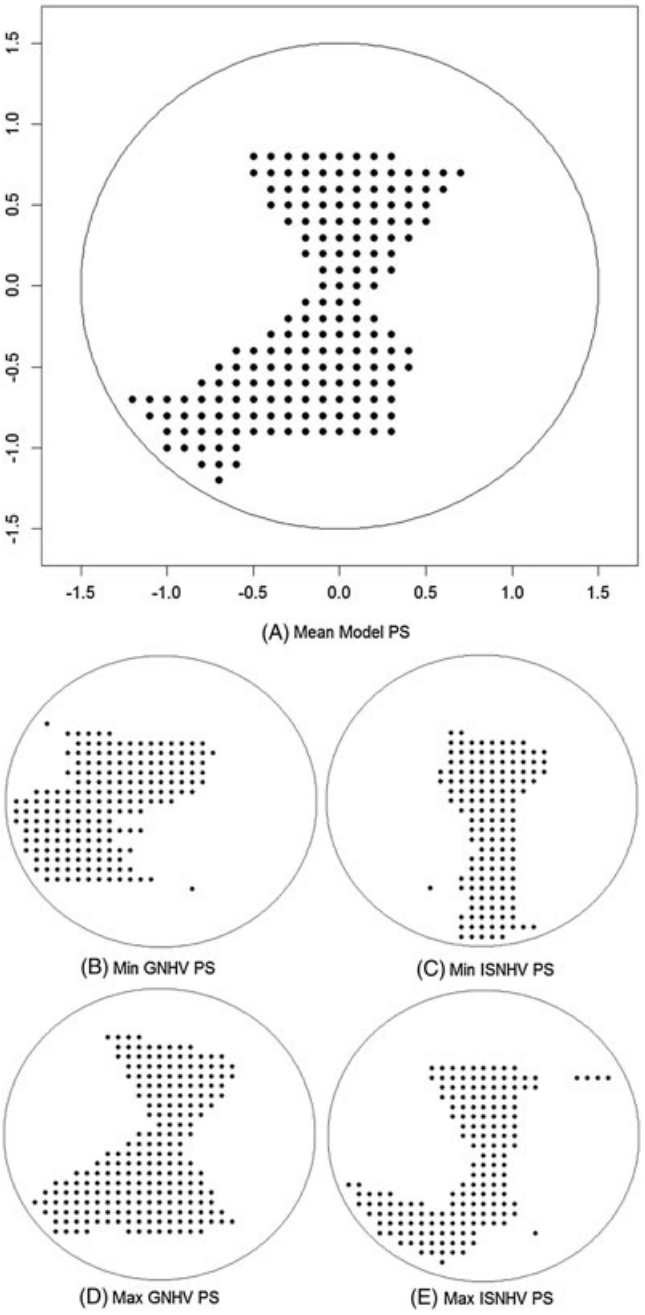

FIGURE 12 Pareto sets for optimizing all 3 responses for the mean model and the minimum and maximum for each of the GNHV and ISNHV from the simulation

TABLE 1 Worst case values on the PF for each optimization scenario, where values were set to 0 for the global scaling

Scenario	Worst Y_1	Worst Y_2	Worst Y_3
Y_1 & Y_2	75.628	−7.035	---
Y_1 & Y_3	75.356	---	−3687.97
Y_2 & Y_3	---	−9.799	−3898.25
Y_1, Y_2, & Y_3	75.305	−10.857	−3926.55

Figure 13 shows the range of the response values on the PF for all 3 responses across the 1000 simulations. Similar patterns to the 2 response at a time cases are observed. There is little variation in the best Y_2 values as the input space almost always includes a solution that achieves the target value of Y_2. The best values for Y_1 show smaller variation than the best values for Y_3 since it has much smaller MSE for the estimated model (ie, a

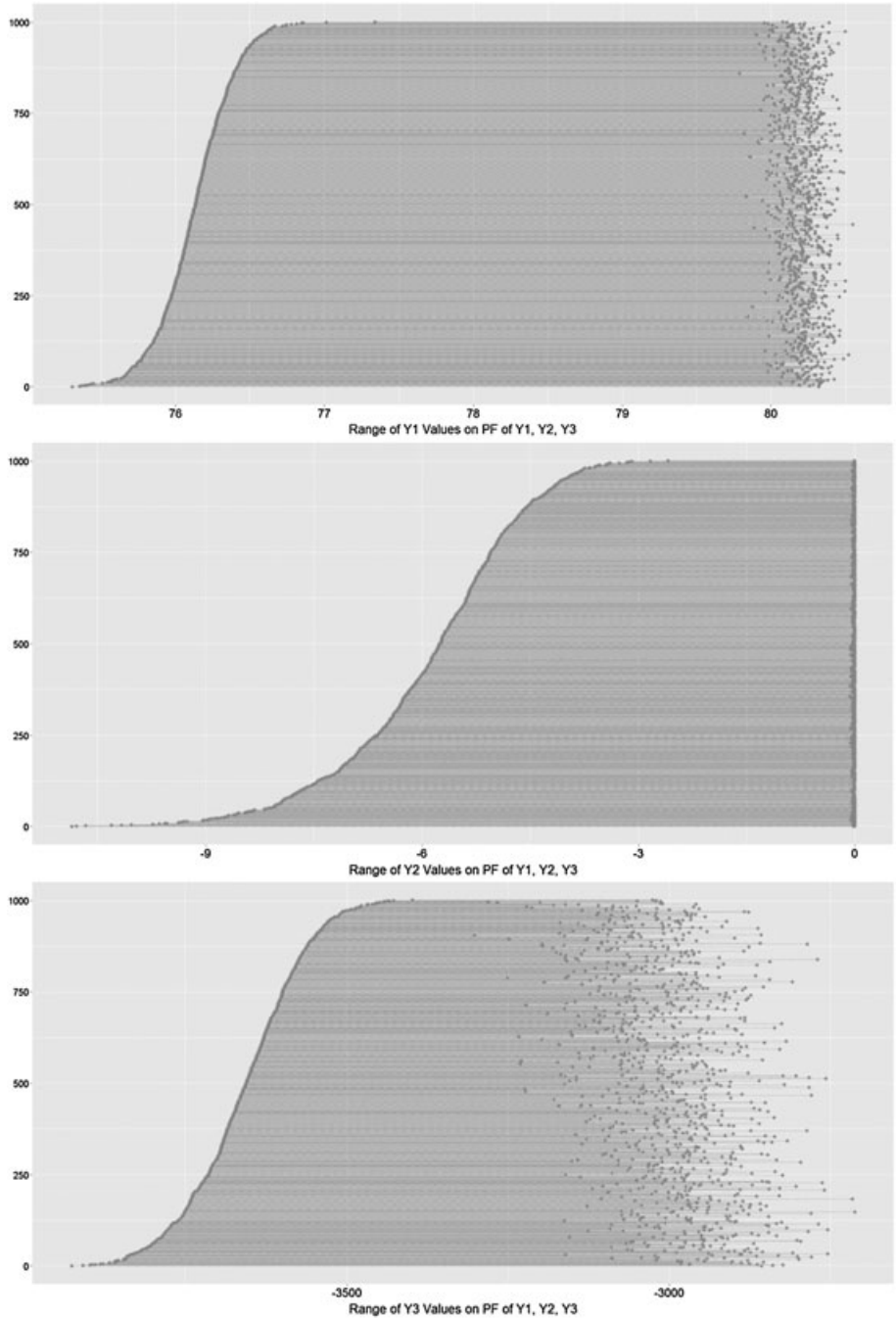

FIGURE 13 Ranges of all 3 responses for the Pareto fronts across the simulations

larger signal-to-noise ratio). The best values for each response appear more consistent than in the 2 response optimization scenarios relative to the entire range of the possible values on the PF across simulations, especially for Y_3 as well as Y_1. Some of this is due to the worst case values extending the set of values that are considered, and some is due to the richness of the front providing a larger collection of values to be included in the plots. This matches with our expectation of having less variable PFs across simulations for all 3 responses than considering only a pair of responses.

In summary, for the optimization of all 3 responses simultaneously, due to having a consistently richer and less variable PF across all the simulations, there are less differences between GNHV and ISNHV in general. The resulting PS are generally more similar with more common input regions across the simulations. Despite the GNHV having a slighter wider range of values due to the larger range of values included in the global scaling, there is more consistency in the identified PS of locations between the extreme cases and the mean model when using the global scaling than the instance-specific scaling.

This again indicates larger variation in the values of the simulated response surfaces, which results in larger variation of simulated PFs on a unified scale than the variation of the richness of the PF across simulations.

6 | CONCLUSIONS

Pareto optimization is a useful multiple response optimization method to generate solutions for problems involving compromise and trade-offs. Typically, the PF is built using mean response values. Therefore, it is important to assess the consistency of the optimal operating conditions from this "mean" PF. To assess the consistency of results, an effective method is to use simulations based on the model parameter uncertainty (or the posterior distribution, if a Bayesian analysis was used). To complement the work exploring changes in the simulated PFs, we propose 2 types of NHV, GNHV, and ISNHV, which consider different aspect of changes in the PFs across the variability in the response surfaces. The GNHV is scaled based on the best and worst values observed across all of the simulations and provides information about the consistency of the PF relative to a fixed unit area of all possible values across simulations. The ISNHV is scaled separately for each of the simulations and characterizes the richness of the front and the severity of the trade-off between responses.

These 2 measures are minimally correlated and measure different aspects of changes in the PFs. They complement each other and when used together, allow good exploration. Some key findings of our exploration are that (1) both GNHV and ISNHV are large when the PF is close to the utopia point and (2) the distributions of both NHVs are sensitive to the type of the optimization. When the goal is achieving a target and the response surface is expected to contain this value, the GNHV values are generally larger. (3) ISNHV is sensitive to the cardinality the PF. One aspect that introduces large ranges in the ISNHV is that a PF with a single solution has ISNHV of 1, while when it has 2 solutions, the ISNHV is 0. Hence, for very sparse PFs, obtaining a wide range of values for the ISNHV is likely. This is not true for GNHV, which is more stable regardless of the richness of the front. (4) As more responses are considered, the richness of the PF increases, which leads to more stable ISNHV values and generally increased ranges used for the GNHV scaling. The distributions of the 2 summaries characterize different aspects of the uncertainty in the trade-offs and hence examining both GNHV and ISNHV can improve understanding.

In this paper, we have considered the nature and variability in the PF, as characterized by the trade-offs between responses. The GNHV and ISNHV provide quantitative information about the interrelationship of the responses. Another important aspect of the optimization process beyond the scope of this paper is how to further select the solution from the PF that best matches the study goals. For the chemical process, it is likely that only one input combination will be selected at which to run the process, and this needs to be decided based on the priorities of the experimenters. For this example, Chapman et al[17] describe a structured process to incorporate subjective priorities into the selection process, while taking into account the variability of the responses. Figures 4, 8, and A2 show how the large differences in the shape and location of the PFs across the simulations will impact the changes in the optimal combination obtained and the uncertainty in the input locations.

REFERENCES

1. Lu L, Anderson-Cook CM, Robinson TJ. Optimization of designed experiments based on multiple criteria utilizing Pareto frontier. *Technometrics*. 2011;53:353-365.
2. Marler RT, Arora JS. Survey of multi-objective optimization methods for engineering. *Structured Multidisciplinary Optimization*. 2004;26:369-395.
3. Borrotti M, Sambo F, Mylona K, Gilmour S. A multi-objective coordinate-exchange two-phase local search algorithm for multi-stratum experiments. *Statistics and Computing*. 2017;27:469-481.
4. Chapman JL, Lu L, Anderson-Cook CM. Process optimization for multiple responses utilizing the Pareto front approach. *Quality Engineering*. 2014;26:253-268.
5. Chapman JL, Lu L, Anderson-Cook CM. Incorporating response variability into Pareto front optimization. *Computers and Industrial Engineering*. 2014;76:253-267.
6. Chankong V, Haimes YY. *Multiobjective Decision Making: Theory and Methodology*. North Holland; 1983.
7. Zitzler, E., Thiele, L. Multiobjective optimization using evolutionary algorithms – a comparative case study. *Parallel Problem Solving from Nature PPSN V: 5th International Conference*. Germany: Springer; 1998:292–301.
8. Lu L, Anderson-Cook CM. Adapting the hypervolume quality indicator to quantify trade-offs and search efficiency for multiple criteria decision-making using Pareto fronts. *Quality and Reliability Engineering International*. 2013;29:1117-1133.
9. Cao Y, Smucker BJ, Robinson TJ. On using the hypervolume indicator to compare Pareto fronts: applications to multi-criteria optimal experimental design. *Journal of Statistical Planning and Inference*. 2015;160:60-74.
10. Asadzadeh M, Tolson DB, Burn DH. A new selection metric for multiobjective hydrologic model calibration. *Water Resour Res*. 2014;50:7082-7099.
11. Myers RH, Montgomery DC, Anderson-Cook CM. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 4th ed. Wiley; 2016.

12. Derringer G, Suich R. Simultaneous optimization of several response variables. *Journal of Quality Technology*. 1980;12:214-219.
13. Anderson-Cook CM, Cao Y, Lu L. Maximize, minimize or target. *Quality Progress*. 2016;49:52-55.
14. Anderson-Cook CM, Cao Y, Lu L. Understanding aspects influencing different types of multiple response optimization. *Quality Engineering*. 2017;29:329-341.
15. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
16. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27:861-874.
17. Chapman JL, Lu L, Anderson-Cook CM. Impact of response variability on Pareto front optimization. *Statistical Analysis and Data Mining*. 2015;8:314-328.

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

Lu Lu is an assistant professor in the Department of Mathematics and Statistics at the University of South Florida. She earned her PhD in Statistics at Iowa State University and was a postdoctoral research associate at Los Alamos National Laboratory. Her research areas include reliability, multiple response optimization, design of experiments, and sampling.

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

How to cite this article: Cao Y, Lu L, Anderson-Cook CM. Uncertainty analysis of trade-offs between multiple responses using hypervolume. *Qual Reliab Engng Int*. 2017;33:2343–2360. https://doi.org/10.1002/qre.2193

APPENDIX A.

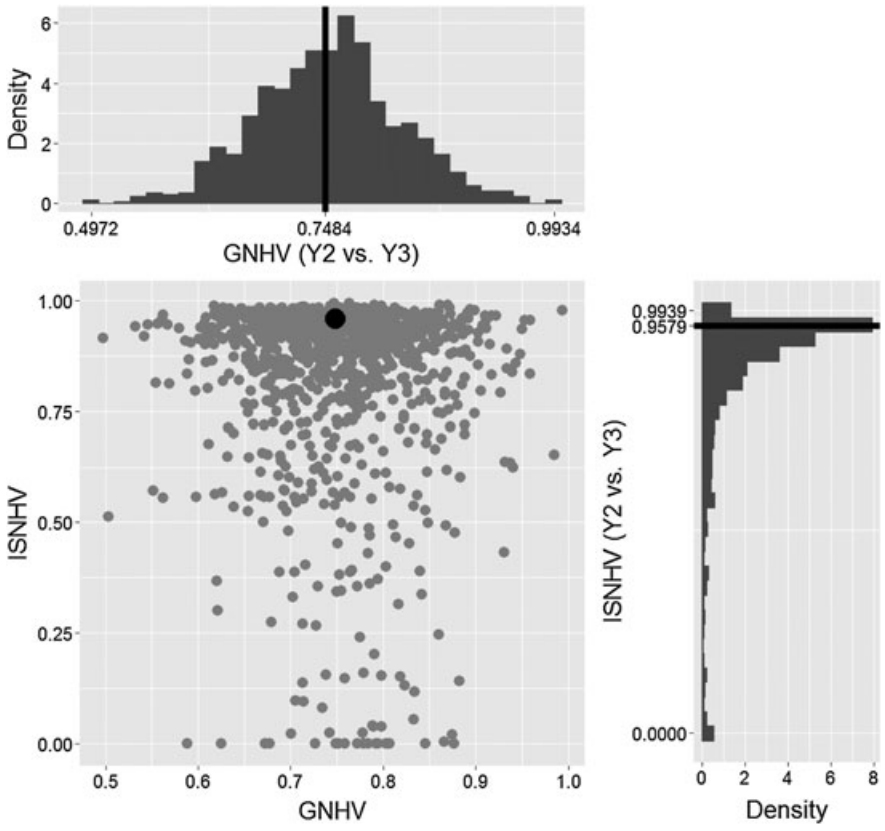

FIGURE A1 Scatterplot and histograms of hypervolume values for viscosity (Y_2) and molecular weight (Y_3) from global and instance-specific scaling

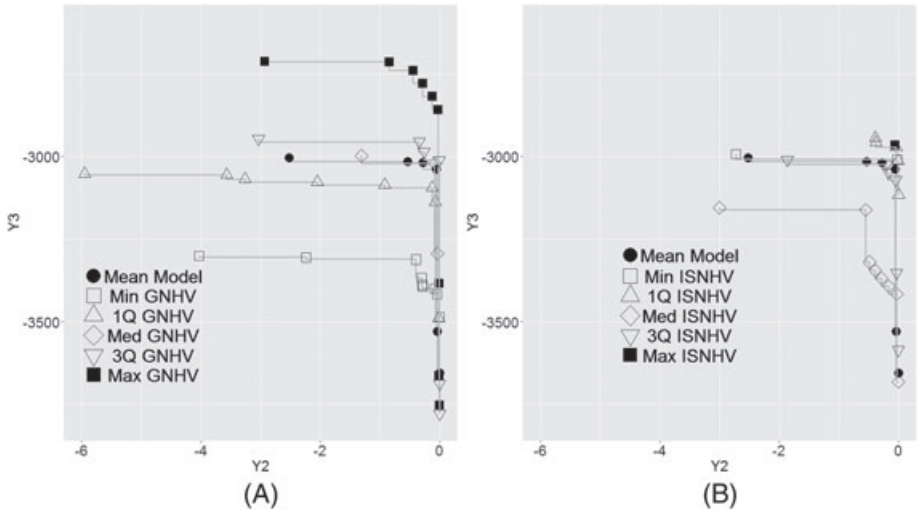

FIGURE A2 Samples of Pareto fronts from the simulations for optimizing viscosity (Y_2) and molecular weight (Y_3) based on A, global scaling and B, instance-specific scaling

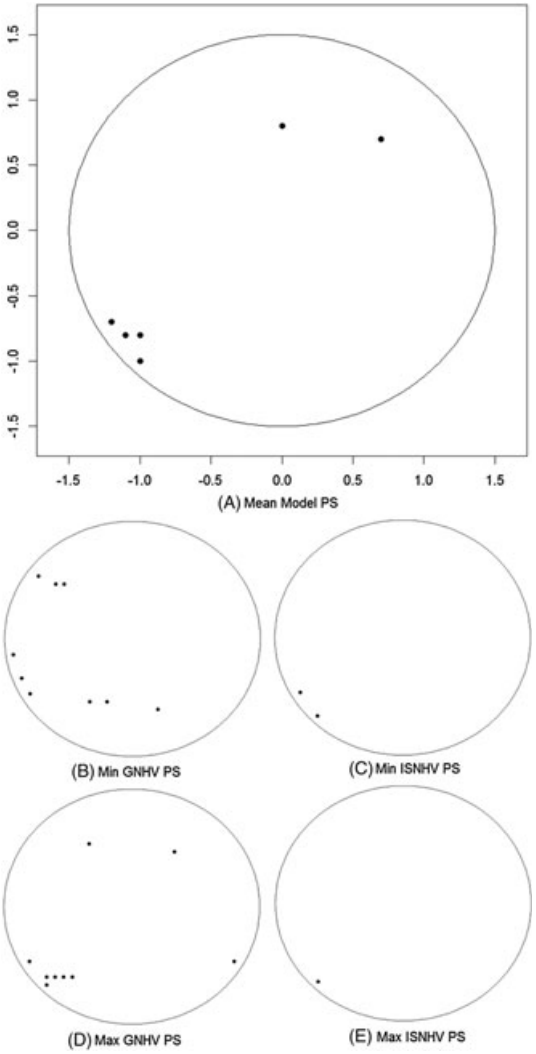

FIGURE A3 Pareto sets for optimizing viscosity (Y_2) and molecular weight (Y_3) for the mean model, and the minimum and maximum for each of the GNHV and ISNHV from the simulations

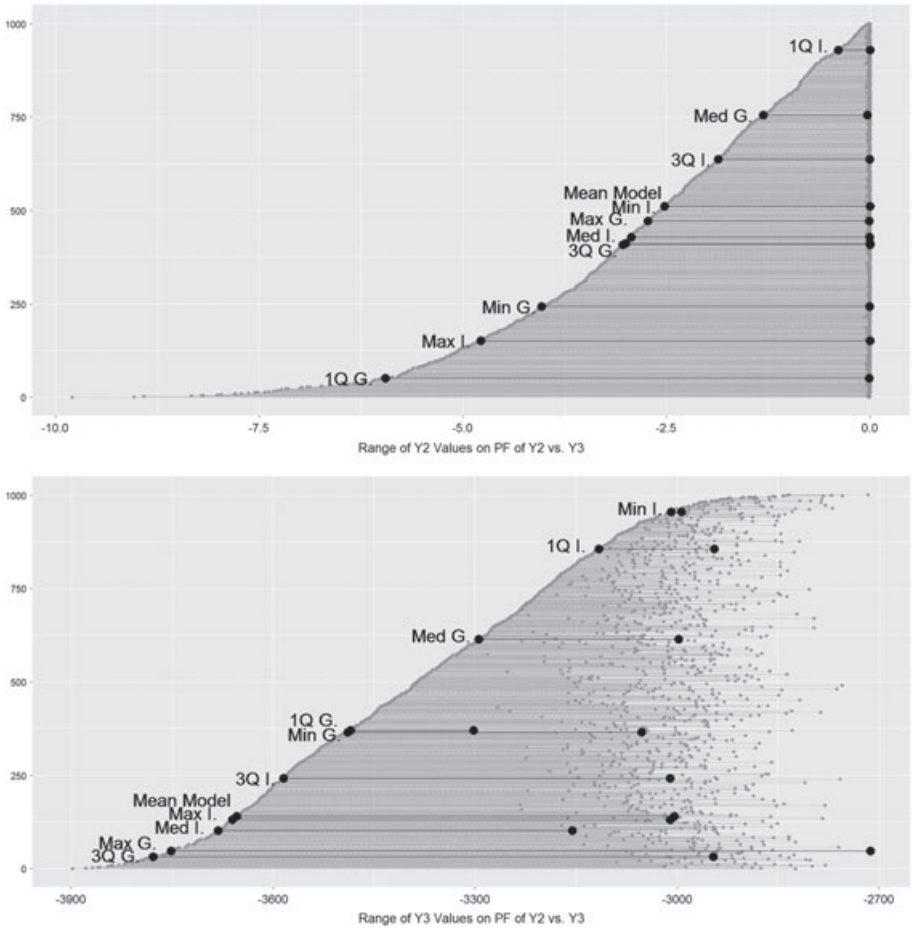

FIGURE A4 Ranges of viscosity (Y_2) and molecular weight (Y_3) for the Pareto fronts across the simulations