Optimization of Composition and Processing Parameters for Alloy Development: A Statistical Model-Based Approach

Alexandr Golodnikov†, Yevgeny Macheret‡, A Alexandre Trindade§,
Stan Uryasev†, and Grigoriy Zrazhevsky†

† Department of Industrial and Systems Engineering, University of Florida
‡ Institute for Defense Analysis
§ Department of Statistics, University of Florida

Abstract

We describe the second step in a two-step approach for the development of new and improved alloys. The first step, proposed by Golodnikov *et al* [3], entails using experimental data to statistically model tensile yield strength and the 20th percentile of the impact toughness, as a function of alloy composition and processing variables. We demonstrate how the models can be used in the second step to search for combinations of the variables in small neighborhoods of the data space, that result in alloys having optimal levels of the properties modeled. The optimization is performed via the efficient frontier methodology. Such an approach, based on validated statistical models, can lead to a substantial reduction in the experimental effort and cost associated with alloy development. The procedure can also be used at various stages of the experimental program, to indicate what changes should be made in the composition and processing variables in order to shift the alloy development process toward the efficient frontier. Data from these more refined experiments can then be used to adjust the model and improve the second step, in an iterative search for superior alloys.

Keywords: materials science, steel fabrication, regression model, Charpy V-Notch, efficient frontier

1 Introduction

In recent work, Golodnikov *et al* [3] developed statistical models to predict the tensile yield strength and toughness behavior of high strength low alloy (HSLA-100) steel. The yield strength was shown to be well approximated by a liner regression model. The alloy toughness (as evaluated by a Charpy V-notch, CVN, at -84°C test), was modeled by fitting separate quantile regressions to the 20th, 50th, and 80th percentiles of its probability distributions. The toughness model was shown to be reasonably accurate, and ranking of the alloys and selection of the best compositions and processing parameters based on the strength and toughness regression models were similar to that of the experimental program

reported by Goldren and Cox [2]. This heralded a new era in alloy development, since the current practice embodied by the experimental program involves several expensive rounds and phases of steel pouring and testing.

Models with the capability to estimate the effect of processing parameters and chemical composition on toughness, are particularly important for alloy design. While the tensile strength can be modeled with reasonable accuracy by, for example, Neural Networks (Metzbower and Czyryca [7]), the prediction of CVN values remains a difficult problem. One of the reasons for this is that experimental CVN data often exhibit substantial scatter. The Charpy test provides a measure of an invariant material property, and CVN values depend on many parameters, including specimen geometry, stress distribution around the notch, and microstructural inhomogeneities around the notch tip. More on the CVN test, including the reasons behind the scatter and statistical aspects of this type of data analysis, can be found in McClintock and Argon [6], Corowin and Houghland [1], Lucon *et al* [4], and Todinov [9]. Developing alloys with minimum allowable CVN values, therefore, results in multiple specimens for each experimental condition, leading to complex and expensive experimental programs. In addition, it is possible that optimum combinations of processing parameters and alloy compositions will be missed due to the practical limitations on the number of experimental alloys and processing conditions.

In this paper we utilize the regression models developed by Golodnikov *et al* [3], to perform computer simulations for optimization of chemical composition and processing, in order to obtain the best combination of alloy strength and toughness. Applying the steel selection criteria described in Goldren and Cox [2], we determine the compositions and processing parameters that result in the best steels. To perform the optimization, we employ the *efficient frontier* methodology, a much-used tool in financial risk management. For a hypothetical specimen of steel on the efficient frontier with given strength and toughness values, we will see that it is not possible to attain the same or a higher strength with a higher value of toughness, nor the same or a higher toughness with a higher value of strength.

Such an endeavor has been thus far practically unattainable in the industry, due to the time and cost involved in the physical manufacture and subsequent testing of specimens with a new set of parameters (the experimental program). Our approach demonstrates that using simulations based on validated statistical models in conjunction with data from experiments, can lead to a significant reduction in the experimental effort and cost associated with the development of new and improved alloys. This paper therefore continues the stream of innovation begun by Golodnikov *et al* [3], where the incorporation of mathematical modeling into the experimental program has the potential to lead to an increase in the rate of discovery of optimal alloys.

The rest of the paper is organized as follows. Section 2 revisits the key results of Golodnikov *et al* [3] which provides the statistical modeling framework on which the current work is based. The optimization problems, the solution of which will identify the improved alloys, are detailed in Section 3. Section 4 presents the results of this optimization by illustrating the various possibilities for the respective efficient frontiers. We conclude with an in-depth discussion outlining the relevance of our proposed methodology to the materials

science community.

2 Statistical Models for Tensile Yield Strength and Charpy V-Notch

In Golodnikov *et al* [3], we developed statistical regression models to predict tensile yield strength (Yield), and fracture toughness (as measured by Charpy V-Notch, CVN, at $-84°$C) of High Strength Low Alloy (HSLA-100) steel. These predictions are based on a particular specimen's chemical composition (C, Mn, Si, Cr, Ni, Mo, Cu) measured in weight percent, and the three alloy processing parameters: plate thickness in mm (Thick), solution treating (Aust), and aging temperature (Aging). As described in the analysis of Golodnikov *et al* [3], the yield strength, chemical composition, and temperature data, have been normalized by their average values with respect to the original source, Goldren and Cox [2].

Generically speaking, statistical regression models attempt to parsimoniously describe the relationship between several predictor (or explanatory) variables, $X_1, \ldots, X_k$, and a particular dependent (or response) variable of interest, Y, measured contemporaneously. The hope is that observations made on the response, $Y_1 \ldots, Y_n$, can be modeled as linear combinations of corresponding observations from the predictors,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

for some unknown coefficients $\beta_0, \beta_1, \ldots, \beta_k$, to be estimated from the available data. The ϵ_i are random variables with common probability distribution function $F_\epsilon(\cdot)$, that account for the surplus variability or scatter in Y that cannot be explained by $X_1, \ldots, X_k$. Many phenomena in science can be usefully modeled in this simple way; the cumulative effects of unmeasured and/or unforeseen variables being lumped into the stochastic ϵ term.

For a given real number $0 \leq \theta \leq 1$, the predictions on Y derived from model (1) take on the generic form

$$h(Y) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + F_\epsilon^{-1}(\theta).$$

In *ordinary* (least squares) regression, the quantity predicted at the given $X_1, \ldots, X_k$ is the mean value of Y, i.e. $h(Y) = E(Y)$, $\theta = 0.5$, and ϵ is assumed to have a symmetric distribution about zero so that $F_\epsilon^{-1}(0.5) = 0$. (See for example Neter *et al* [8] for a thorough treatment of ordinary least squares regression.) In *quantile* regression, with θ the quantile of interest, $h(Y) = F_Y^{-1}(\theta)$, so that the quantity predicted at the given $X_1, \ldots, X_k$, is the θth quantile (or 100θth percentile) of Y.

One way to quantify the success of a regression model, is to measure the percentage of the variability in Y that is explained by the X's. In ordinary regression, this quantity is the R^2; e.g. $R^2 = 90\%$ meaning that only 10% of the variation in Y cannot be accounted for by the X's. In θth quantile regression, this quantity is the $R^1(\theta)$, with analogous interpretation. An important issue when many candidate predictor variables are involved, is how to choose an appropriate subset so that a large percentage of the variability in Y is explained with few X variables. For reasons of efficiency, it is desirable to exclude from selection variables

that contribute little to the explanation of Y. Also desirable, is the exclusion of predictors that may be highly correlated with other predictors.

Let us denote the i-th alloy predictor variable (chemical composition or processing parameter), by x_i, $i = 1, \ldots, 16$. The particular variable represented by the index i is given in Tables 1 and 2. In Golodnikov *et al* [3], we found the ordinary regression model

$$\begin{aligned}\text{Yield} = 0.7 + 0.173x_1 + 0.045x_2 + 0.13x_4 + 0.123x_6 + 0.358x_7 \\ + 0.004x_{13} + 0.285x_{14} - 0.759x_{15} - 0.002x_{16}, \quad (2)\end{aligned}$$

to be fairly successful in predicting Yield, as judged by a value of $R^2 = 94\%$ and other standard measures of assessing model adequacy. The CVN data however, had considerably more scatter, so that ordinary regression models did not provide a suitable fit. The scatter was dealt with by fitting separate quantile regression models, each targeting a specific percentile of the CVN distribution. In particular, the 20%th percentile of the distribution function of CVN is of interest in this analysis, since it plausibly models the smallest value of the three values of CVN associated with each data point. The following model was obtained,

$$\begin{aligned}20\% \text{ log CVN} = 0.000 - 0.1x_2 + 0.04x_4 - 0.419x_6 + 0.608x_7 - 0.144x_{10} \\ - 0.035x_{13} - 0.693x_{14} + 1.692x_{15} - 0.004x_{16}. \quad (3)\end{aligned}$$

(CVN was modeled on the logarithmic scale in order to guarantee that predicted values would always be positive.) Although not quite as successful as the model for Yield, the $R^1(0.2) = 52\%$ value and further goodness of fit analyses showed this model to be sufficiently useful for its intended purpose, the selection and ranking of good candidate steels (Golodnikov *et al* [3]).

Note that Ni and Cb do not appear in the above models. They were excluded from consideration due to their high degree of collinearity (linear correlation) with the remaining predictors. However, since they have an important effect on the properties of steel, their composition in the alloys was accounted for by developing a separate linear regression model for each. These models are as follows,

$$\begin{aligned}\hat{x}_5 = 0.688 - 0.148x_1 - 0.390x_2 + 0.367x_4 + 0.061x_6 \\ + 0.496x_7 + 0.111x_{10} - 0.015x_{13} - 0.222x_{14}, \quad (4)\end{aligned}$$

and,

$$\begin{aligned}\hat{x}_8 = 0.189 + 1.158x_1 - 0.239x_2 + 0.346x_4 - 0.126x_6 \\ - 0.284x_{10} + 0.013x_{13} + 0.431x_{14} + 0.287x_{15}. \quad (5)\end{aligned}$$

(All estimated coefficients are statistically significant at a confidence level of 95%.) Si, Al, P, and S, were also excluded from Yield and CVN modeling. These are contaminants, their

appearance in the composition of the alloys strictly controlled by maintaining them at the lowest possible levels.

Since the primary aim of this article is to search for combinations of chemical compositions and processing parameters that result in steels with optimal Yield and CVN characteristics (the higher the better), feasibility bounds must be imposed on these variables. The actual bounds we will use are adapted from Goldren and Cox [2], and appear in Tables 1 and 2.

Table 1: Feasibility bounds on 10 normalized chemical compositions considered in the search for the optimal steels. To each chemical there corresponds also an index which will be used to reference it in subsequent notation.

	C	Mn	Si	Cr	Ni	Mo	Cu	Cb	Al	N
index i	1	2	3	4	5	6	7	8	9	10
minimum	0.61	0.47	0.17	0.40	0.28	0.46	0.51	0.60	0.43	0.49
maximum	1.36	2.01	1.89	1.40	1.21	1.60	1.26	1.68	1.57	1.64

Table 2: Feasibility bounds on P, S, and V chemical compositions and normalized processing parameters, considered in the search for the optimal steels. To each variable there corresponds also an index which will be used to reference it in subsequent notation.

	P	S	V	Aust	Aging	Thick (mm)
index i	11	12	13	14	15	16
minimum	0.74	0.63	0	0.96	0.842	6.4
maximum	2.32	2.079	30.83	1.05	1.072	51

3 Efficient Frontier Approach for Optimization of Steel Processing and Composition

Theoretically, regression models (2) and (3) could be used for predicting the Yield and 20% log CVN of new data points, points that were not used in fitting these models (the *training* set). While the prediction error for the training data points is known, the prediction error for a new data point is not. Nevertheless, from continuity considerations, it follows that the closer a new data point is to the training set, the better the prediction accuracy for that point. In other words, predictions made in the neighborhood of the training set are generally more reliable than predictions for more distant points.

Using this reasoning we will search for the best data points in a small neighborhood of the training set. By definition, this neighborhood is the union of all neighborhoods

corresponding to individual points in the training set. However, since this mathematical construction is rather complex from a computational point of view, we will consider instead the neighborhood defined by the convex hull of the training set. To further reduce computational complexity, we will consider only specimens with a thickness of 51 mm, restricting our attention to the convex hull of such points. This involves the following specimens of steel: # 21 (1 point), # 23 (1 point), # 28 (8 points), # 30 (4 points), # 35 (4 points), # 37 (1 point); a total of 19 data points.

In order to define the aforementioned convex hull, we introduce some notation. Let z_{ij} denote the value of the i-th alloy predictor variable (chemical composition or processing parameter) for the j-th data point, $i = 1, \ldots, 15$ (as indexed by the first row of Tables 1 and 2), $j = 1, \ldots, 19$. Let $\boldsymbol{z}_j = [z_{1,j}, \ldots, z_{15,j}]'$ denote the vector of these values for the j-th data point.

Definition 1 (Convex Hull of Training Set) *The convex hull of the training set is described by the vector* $\boldsymbol{y} = [y_1, \ldots, y_{15}]'$*, whose i-th component is*

$$y_i = z_{i,1}\alpha_1 + z_{i,2}\alpha_2 + \cdots + z_{i,19}\alpha_{19}, \tag{6}$$

for a set of scalars $\{\alpha_j\}$*,* $j = 1, \ldots, 19$*, satisfying,*

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{19} = 1, \tag{7}$$

and

$$\alpha_1 \geqslant 0,\ \alpha_2 \geqslant 0,\ \cdots,\ \alpha_{19} \geqslant 0. \tag{8}$$

Let $l_i \leq u_i$ denote the lower and upper bounds on the i-th variable presented in Tables 1 and 2, $i = 1, \ldots, 15$. For $0 \leq \varepsilon \leq 100$, the $\varepsilon\%$ neighborhood of the convex hull defined by (6)-(8) is the set of points $\{x_i\}$, $i = 1, \ldots, 15$, represented by the vector $\boldsymbol{x} = [x_1, \ldots, x_{15}]'$, that satisfy the constraints,

$$\mid x_i - (z_{i,1}\alpha_1 + z_{i,2}\alpha_2 + \cdots + z_{i,19}\alpha_{19}) \mid < (u_i - l_i)\varepsilon/100, \tag{9}$$

and (7)–(8). In (9), $\boldsymbol{x}$ is the analogue of $\boldsymbol{y}$ in (6).

Consider now the search for the best points, $\boldsymbol{x}$, in a small $\varepsilon\%$ neighborhood of the convex hull of the training set, with ε in the range 1-3%. By definition, these best points will be in the $\varepsilon\%$ neighborhood of the convex hull, and will correspond to specimens of steel having a maximum possible value of Yield, under the constraint that the 20% of log CVN is not less than some fixed value a. As a varies, these best points define a curve on the Yield-CVN diagram, called the *efficient frontier*. For a point on this curve with given Yield and CVN values, one cannot therefore attain the same or a higher Yield with a higher value of CVN, nor the same or a higher CVN with a higher value of Yield.

The concept of an efficient frontier was first introduced by Markowitz (1952) in connection with financial portfolio optimization, where he desired to choose portfolio weights so as to maximize its return while simultaneously minimizing its variance. The concept can

be generalized to any situation where one seeks to simultaneously optimize two attributes that are controlled by a finite set of variables. For a fixed value of one attribute, the efficient frontier defines the optimality boundary beyond which it is not possible to achieve a more optimal level of the other attribute. With the incorporation of linear constraints, the resulting set of optimization tools has become standard in financial risk management, and is also finding widespread usage in other fields.

In our case, construction of the efficient frontier involves solving two optimization problems.

Optimization Problem 1 (Maximize Yield) *Find the values of* $\{x_1, \ldots, x_{16}\}$ *that maximize Yield, as given by model (2), subject to the constraints:*

(i) 20% log CVN $\geqslant a$*, with 20% log CVN given by model (3);*

(ii) $\mid x_5 - \hat{x}_5 \mid < \varepsilon_{\mathrm{Ni}}$*, where* $\hat{x}_5$ *is the predicted Ni value from model (4);*

(iii) $\mid x_8 - \hat{x}_8 \mid < \varepsilon_{\mathrm{Cb}}$*, where* $\hat{x}_8$ *is the predicted Cb value from model (5);*

(iv) $\mid x_i - (z_{i,1}\alpha_1 + z_{i,2}\alpha_2 + \cdots + z_{i,19}\alpha_{19}) \mid < (u_i - l_i)\varepsilon/100$*,* $i = 1, \ldots, 15$*;*

(v) $\alpha_1 + \alpha_2 + \cdots + \alpha_{19} = 1$*;*

(vi) $\alpha_1 \geqslant 0,\ \alpha_2 \geqslant 0,\ \cdots,\ \alpha_{19} \geqslant 0$*;*

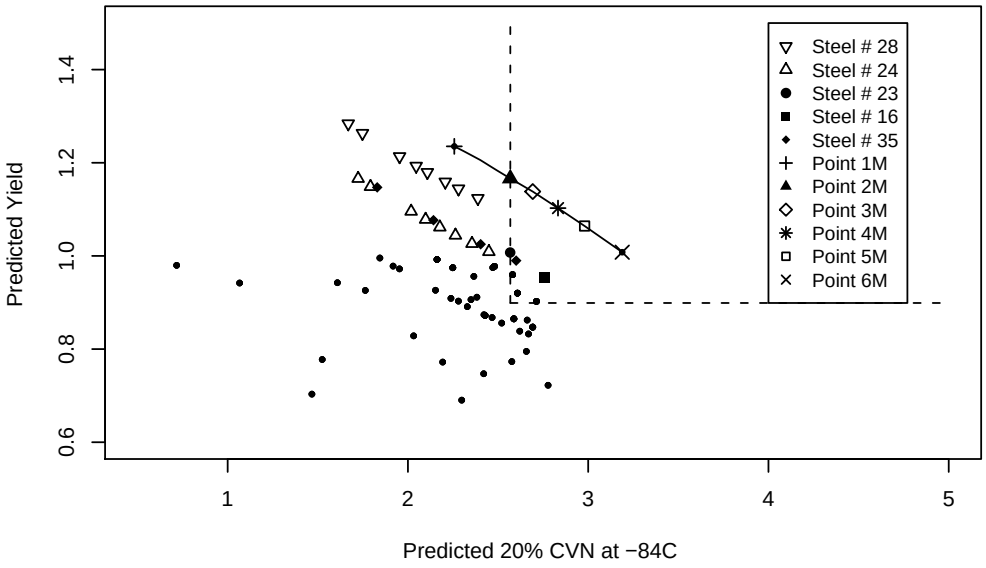
(vii) bounds on $x_1, \ldots, x_{15}$ *as specified in Tables 1 and 2.*

Restriction (i) embodies one of the acceptability screening rules, which stipulates that a specimen of steel be disqualified if the smallest of its three values of CVN at $-84°$C is less then a. Restrictions (ii)-(iii) ensure that Ni and Cb levels are approximately equal to that predicted by their respective regression models, within a tolerance of $\varepsilon_{\mathrm{Ni}}$ and $\varepsilon_{\mathrm{Cb}}$. These restrictions tie together variables that were not used in the regression models for Yield and 20%-percentile of log of CVN, with those that were. Together with (vii), this ensures that Ni and Cb affect the optimization. Finally, (iv)-(vi) restrict the search for the best points to the ε%-neighborhood of the convex hull of the training set.

Optimization Problem 1 maximizes Yield with a restriction on the 20%-percentile of the logarithm of CVN at $-84°$C. By symmetry, we should also maximize the 20%-percentile of log CVN under the condition that Yield belong to some feasible range of values.

Optimization Problem 2 (Maximize CVN) *Find the values of* $\{x_1, \ldots, x_{16}\}$ *that maximize CVN, as given by model (3), subject to the constraint that the predicted value of Yield from model (2) be in the range* $0.9 \leq Yield \leq 1.079$*, and restrictions (ii)-(vii) of Optimization Problem 1.*

Figure 1: Strength-Toughness efficient frontier for the 1% neighborhood of the convex hull of the training data set.



4 Optimization Results

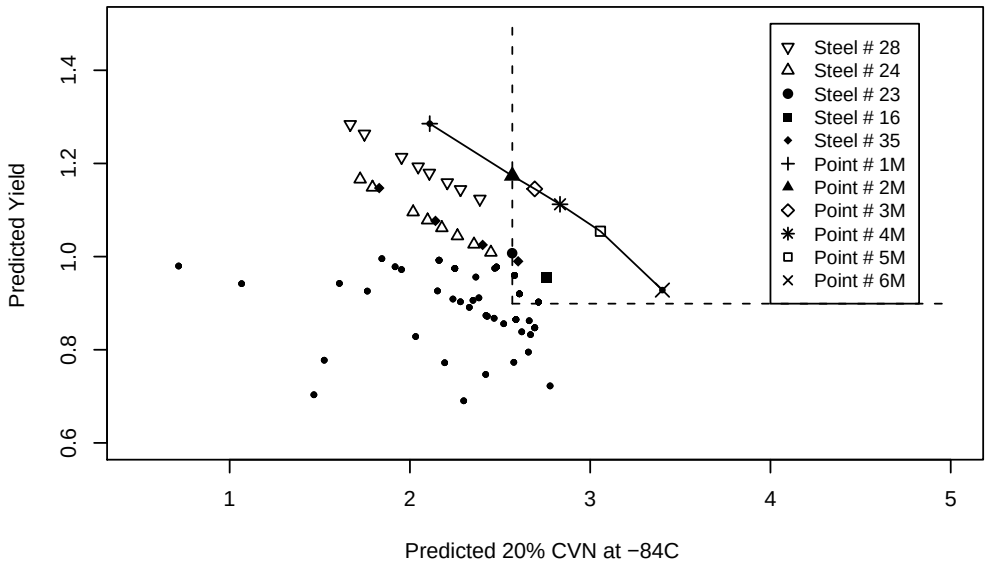
In this section we present and discuss the solutions to Optimization Problems 1 and 2, the efficient frontiers, for different ε% neighborhoods of the convex hull of the training set ($\varepsilon = 1, 2, 3$). The efficient frontier for $\varepsilon = 1$ is plotted as a solid line in Figure 1. The acceptable bounds on the 20%-percentile of CVN (2.568) and Yield (0.899), are displayed as vertical and horizontal dashed lines, respectively. The *feasible* region is therefore defined by the top right quadrant delimited by these lines. The vertical dashed line intercepts the efficient frontier at point 2M. The left end of the efficient frontier, point 1M, is outside the feasible region. The training points closest to the efficient frontier, correspond to specimens of steel # 28, but all of them lie outside the feasible region. The point marked by a black square (steel # 16) designates a simulated specimen of steel # 16, since the test data for this steel were obtained for specimens of thickness 19 mm, and not 51 mm as this one has. In fact, only two points in the feasible region correspond to actual specimens: steels # 23 and # 35, with an aging temperature of 1.05, are closest to the efficient frontier.

Values of the composition and processing variables $x_1, \ldots, x_{16}$ for 5 feasible points on the efficient frontier obtained by considering the 1% neighborhood of the convex hull of the

training set, are presented in Table 3 of the Appendix. The best points are 2M and 6M. Point 2M is characterized by the largest value of Yield among all points that satisfy the restriction on the 20% of CVN. It is a solution of Optimization Problem 1 with $a = \log(2.568)$. Point 6M is characterized by the largest value of 20% of CVN among all points that satisfy the restriction on Yield, and is a solution of Optimization Problem 2.

The efficient frontiers for $\varepsilon = 2\%$ and $\varepsilon = 3\%$ are plotted in Figure 2 and 3, respectively. Corresponding composition and processing variable values for selected points on the efficient frontier and within the feasible region are presented in Tables 4 and 5 of the Appendix.

Figure 2: Strength-Toughness efficient frontier for the 2% neighborhood of the convex hull of the training data set.

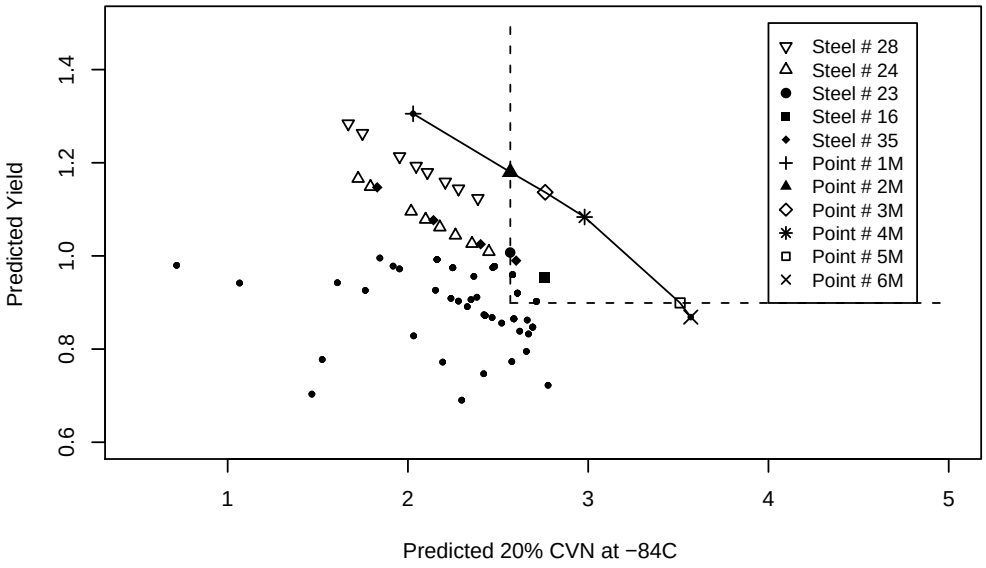


As is evident from Figure 3, and differing from the previous cases of 1% and 2% neighborhoods, point 6M is infeasible in the $\varepsilon = 3$ case since it is below the horizontal dashed line (the lower feasibility bound on Yield). Although points on the efficient frontier for the 3% neighborhood exhibit better strength-toughness combinations than those for the 1% and 2% neighborhoods, they suffer from larger prediction errors.

Finally, we considered the efficient frontier when no restrictions are imposed on the neighborhood of the convex hull of the training set, i.e. the neighborhood is permitted to be arbitrarily large. The efficient frontier for this case is plotted in Figure 4.

As is to be expected, the most striking feature is that the distance between the training

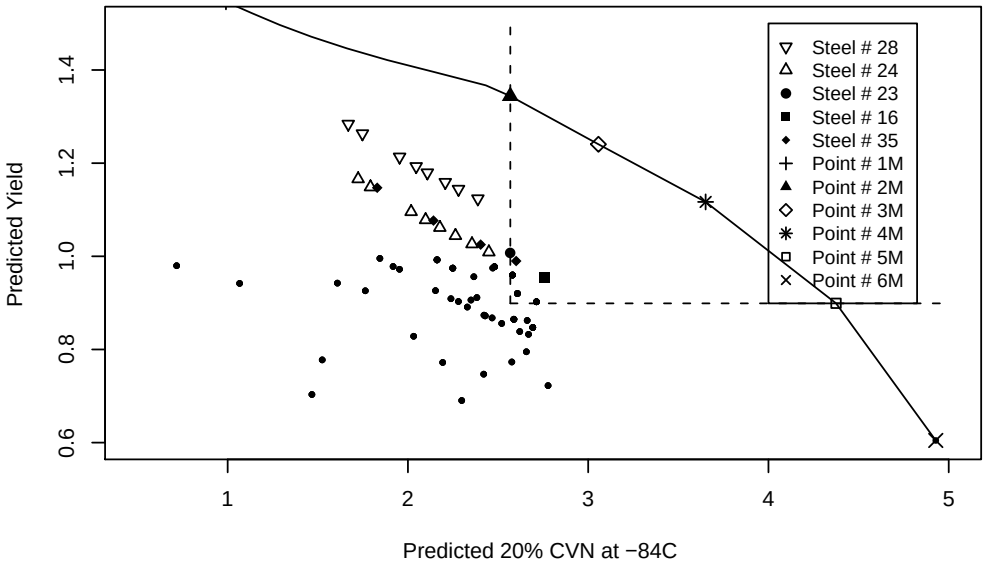
Figure 3: Strength-Toughness efficient Frontier for the 3% neighborhood of the convex hull of the training data set.



set points and the efficient frontier is much larger in comparison to the $\varepsilon = 1\%$, 2%, and 3% cases. We also note that the maximum feasible value of 20% CVN at $-84°$C is very large. Although points on the efficient frontier exhibit much better strength-toughness combinations than before, they also have much larger prediction errors. Corresponding composition and processing variable values for selected points on the efficient frontier and within the feasible region are presented in Table 6 of the Appendix.

Tables 3-5 also show that the composition of the contaminants Si, Al, P, and S varies among points on the efficient frontier. This is a result of imposing optimization restriction (iv) on the ε% neighborhood of the convex hull of the training set, which covers all composition and processing variables. Note that the contaminants do not affect Yield, CVN at $-84°$C, and the remaining explanatory variables. That is, the optimization results are unaffected if the restrictions for $i = 3, 9, 11, 12$ are omitted from optimization constraint (iv). Indeed, this is borne out by Table 6, which displays the result of Optimization Problems 1 and 2 with constraint (iv) omitted, where we see that the levels of Si, Al, P, and S do not vary across points on the efficient frontier.

Figure 4: Strength-Toughness efficient frontier when no restrictions are imposed on the neighborhood of the convex hull of the training set.



5 Discussion

In this paper we have detailed the second step of a two-step approach for the development of new and improved alloys. In the first step explored by Golodnikov *et al* [3], experimental data was used to statistically model yield strength and the 20% percentile of toughness (measured by Charpy V-Notch at $-84°$C), as a function of 16 alloy composition and processing variables. We then used these models in the second step to search for combinations of the variables in small neighborhoods of the data space, that result in alloys having optimal levels of the properties modeled. We have shown how this step can be cast in the framework of a linear optimization with constraints problem, the optimal points lying on the efficient frontier of the solution. By considering larger neighborhoods of the data space, the efficient frontier was seen to lead to alloys with increasingly better strength-toughness characteristics. However, there is a trade-off in that the consideration of larger neighborhoods also results in larger model-prediction errors, with a consequent decrease in certainty in the optimal solutions thus identified. Guided by engineering considerations, a decision maker can nevertheless select an appropriately small neighborhood of the data space in the 1-3% range considered above.

The proposed approach has important implications for the design of physical experi-

ments targeting the discovery of new and improved alloys. Its primary merit is the incorporation of information from prior experimental data into mathematical models, which are subsequently used in refining the fabrication process. Computer simulations based on these models establish the efficient frontier, thereby determining a region of optimal alloy composition and processing parameters. This can lead to a substantial reduction in the experimental effort and its associated cost. The simulations can also be used successively at various stages of the experimental program, to indicate what changes should be made in the parameters (composition and processing) in order to shift the next test/experiment in the direction of the efficient frontier. Data obtained from these more focused experiments could then be used to adjust the model and make recommendations on subsequent steps, in an iterative search for superior alloys.

Finally, it may be useful to monitor the discrepancy in strength and toughness between the optimization-predicted steels on the efficient frontier, and actual observed values upon manufacture of specimens with the predicted optimal composition and processing levels. One would expect to see a gradual decrease in the magnitude of these errors as the proposed methodology, manufacture of new specimens followed by mathematical modeling/optimization, is iterated. Convergence to a fairly optimal set of parameters would then be indicated by errors that change little from one iteration to the next.

References

[1] W.R. Corowin and A.M. Houghland, *Effect of specimen size and material condition on the Charpy impact properties of 9Cr-1Mo-V-Nb steel*, in: The Use of Small-Scale Specimens for Testing Irradiated Material, ASTM STP 888, (Philadelphia, PA, 1986) 325-338.

[2] A.P. Goldren and T.B. Cox, *Development of 100 KSI Yield Strength HSLA Steel*, AMAX Report, CPR-2, AMAX Materials Research Center, (Ann Arbor, MI, July 1986).

[3] A. Golodnikov, Y. Macheret, A. Trindade, S. Uryasev and G. Zrazhevsky, *Statistical Modelling of Composition and Processing Parameters for Alloy Development*, Modelling and Simulation in Materials Science and Engineering, 13, (2005), 633-644.

[4] E. Lucon *et al Characterizing Material Properties by the Use of Full-Size and Sub-Size Charpy Tests*, in: Pendulum Impact Testing: A Century of Progress, ASTM STP 1380, T. Siewert and M.P. Manahan, Sr. eds., American Society for Testing and Materials, (West Conshohocken, PA, 1999) 146-163.

[5] H.M. Markowitz, *Portfolio selection*, Journal of Finance, 7, (1952), 77-91.

[6] F.A. McClintock and A.S. Argon, Mechanical Behavior of Materials, Reading, Massachusetts (U.S.A.), Addison-Wesley Publishing Company, Inc., and Ontario (Canada), Addison-Wesley Ltd., Don Mills, 1966.

[7] E.A. Metzbower and E.J. Czyryca, *Neural Network Analysis of HSLA Steels*, in: T.S. Srivatsan, D.R. Lesuer, and E.M. Taleff eds., Modeling the Performance of Engineering Structural Materials, TMS, 2002.

[8] J. Neter, M.H. Kutner, C.J. Nachtsheim, and W. Wasserman, Applied Linear Statistical Models , Chicago, Irwin, 1996.

[9] M.T. Todinov, *Uncertainty and risk associated with the Charpy impact energy of multi-run welds*, Nuclear Engineering and Design, 231, (2004) 27-38.

A Tables of Composition and Processing Variables for Selected Points on the Efficient Frontiers

Table 3: Values of variables for 5 points on the efficient frontier of 1% neighborhoods of the convex hull of the training data set. All points are in the feasible region, with V=0 and Thick=51.

	Point				
Variable	2M	3M	4M	5M	6M
C	1.183	1.183	1.149	1.097	1.006
Mn	1.095	1.095	1.064	1.016	0.949
Si	1.361	1.361	1.196	0.943	0.593
Cr	1.317	1.317	1.286	1.237	1.167
Ni	1.209	1.209	1.192	1.168	1.142
Mo	1.288	1.288	1.268	1.238	1.195
Cu	1.257	1.257	1.235	1.201	1.152
Cb	1.064	1.064	1.030	0.988	0.925
Al	1.067	1.067	0.990	0.871	0.708
N	1.423	1.423	1.339	1.210	1.053
P	1.313	1.313	1.253	1.161	1.035
S	1.464	1.464	1.348	1.172	0.930
Aust	0.976	0.976	0.979	0.982	0.987
Aging	0.977	1.014	1.033	1.040	1.05

Table 4: Values of variables for 5 points on the efficient frontier of 2% neighborhoods of the convex hull of the training data set. All points are in the feasible region, with V=0 and Thick=51.

	Point				
Variable	2M	3M	4M	5M	6M
C	1.191	1.191	1.164	1.095	0 .781
Mn	1.078	1.078	1.056	0.982	0.819
Si	1.346	1.346	1.216	0.876	0.353
Cr	1.326	1.326	1.303	1.225	0.994
Ni	1.211	1.211	1.208	1.173	1.095
Mo	1.298	1.298	1.284	1.235	1.131
Cu	1.264	1.264	1.247	1.194	1.077
Cb	1.055	1.096	1.036	1.008	0.768
Al	1.078	1.078	1.019	0.85	0.695
N	1.408	1.408	1.346	1.153	0.813
P	1.331	1.331	1.282	1.167	1.017
S	1.48	1.48	1.389	1.15	0.979
Aust	1.013	1.013	0.977	0.983	0.986
Aging	0.989	1.025	1.034	1.045	1.021

Table 5: Values of variables for 4 points on the efficient frontier of 3% neighborhoods of the convex hull of the training data set. All points are in the feasible region, with V=0 and Thick=51.

	Point			
Variable	2M	3M	4M	5M
C	1.204	1.207	1.127	0.976
Mn	1.053	1.048	0.999	0.778
Si	1.35	1.361	0.983	0.561
Cr	1.324	1.317	1.271	0.934
Ni	1.211	1.211	1.198	1.064
Mo	1.301	1.296	1.269	1.078
Cu	1.264	1.259	1.226	1.055
Cb	1.061	1.133	1.033	0.907
Al	1.087	1.085	0.929	0.775
N	1.379	1.368	1.225	0.97
P	1.365	1.376	1.22	1.127
S	1.507	1.514	1.253	1.101
Aust	1.031	1.011	0.979	0.985
Aging	0.988	1.034	1.043	1.071

Table 6: Values of variables for 4 points on the efficient frontier when no restrictions on the neighborhood of the convex hull of the training data set are imposed. All points are in the feasible region, with V=0 and Thick=51.

	Point			
Variable	2M	3M	4M	5M
C	1.357	1.357	1.357	1.247
Mn	0.853	0.853	0.661	0.469
Si	0.168	0.168	0.168	0.168
Cr	1.402	1.402	1.143	1.402
Ni	1.211	1.211	1.211	1.211
Mo	1.599	1.599	1.599	0.461
Cu	1.264	1.264	1.264	1.028
Cb	1.528	1.567	1.583	1.675
Al	0.427	0.427	0.427	0.427
N	0.492	0.492	0.492	0.492
P	0.741	0.741	0.741	0.741
S	0.63	0.63	0.63	0.63
Aust	1.049	1.049	0.957	0.957
Aging	0.864	1	1.072	1.072