MIT Libraries | DSpace@MIT

MIT Open Access Articles

Operational Strategies for Increasing Secondary Materials in Metals Production Under Uncertainty

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Noshadravan, Arash et al. "Operational Strategies for Increasing Secondary Materials in Metals Production Under Uncertainty." Journal of Sustainable Metallurgy 3.2 (2017): 350–361.

As Published: http://dx.doi.org/10.1007/s40831-016-0100-6

Publisher: Springer International Publishing

Persistent URL: http://hdl.handle.net/1721.1/109890

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Massachusetts Institute of Technology

Operational strategies for increasing secondary materials in metals production under uncertainty

*Arash Noshadravan[1**], G. Gaustad[3], R. Kirchain[1], E. Olivetti[2]**

[1]Engineering Systems Division, [2]Department of Materials Science & Engineering,

Massachusetts Institute of Technology, 77 Massachusetts Ave 8-403, Cambridge, MA 02139

[3]Golisano Institute for Sustainability, Rochester Institute of Technology 111 Lomb Memorial Drive, Rm. 81-2175, Rochester, NY

*elsao@mit.edu

**Current address: Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77802

Abstract

Increased use of secondary raw materials in metals production offers several benefits including reduced cost and lowered energy burden. The lower cost of secondary or scrap materials is accompanied by an increased uncertainty in elemental composition. This increased uncertainty for different scraps, if not managed well, results in increased risk that the elemental concentrations in the final products fall outside customer specifications. Previous results show that incorporating this uncertainty explicitly into batch planning can modify the potential use of scrap materials while managing risk. Chance constrained formulations provide one approach to uncertainty-aware batch planning; however typical formulations assume normal distributions to represent the compositional uncertainty of the materials. Compositional variation in scrap materials has been shown to have a skewed distribution and, therefore, the performance of these models, in terms of their ability to provide effective planning, may then be heavily influenced by the structure of the compositional data used. To address this issue, this work developed several approximations for skewed distributional forms within chance constrained formulations. We explored a lognormal approximation based on Fenton's method; a convex approximation based on Bernstein inequalities; and a linear approximation using fuzzy set theory. Each of these methods was formulated

and case studies executed using compositional data from an aluminum remelter. Results indicate that the relationship between the underlying structure/distribution of the compositional data and how these distributions are formulated in batch planning can modify the use of secondary raw materials.

Introduction

One key strategy that will likely play a role in addressing sustainability challenges in materials use is increasing reliance on secondary (i.e., recycled) materials. For recycled resources, energy benefits are well documented and in some cases exceptional [1]. In addition, making use of secondary raw materials typically provides economic benefits in the form of reduced materials cost, although this may not always be the case. Along with these benefits, increasing the volume of raw materials derived from secondary sources has challenges such as increased quality variability relative to primary due to contamination and co-mingling of diverse materials and alloys [2]. Socio-economic barriers offer challenges as well, such as low consumer participation [3, 4], market resistance [5, 6], transaction costs [7], and geographic discrepancies between sources and sinks [8]. This work offers operational strategies for dealing specifically with compositional variability in secondary streams compared to those derived from primary resources. Compositional uncertainty in alloying elements may lead to conservative use of these materials (in other words, these materials may be undervalued) because the risk of producing alloys out of specification may be costly. Operators make use of batch planning tools to assist in processing decisions [9]. Analytical techniques based on statistical methods can provide promising approaches to assess and improve the productivity of metal recycling process [10]. Previous work by some of the authors indicated that one approach to batch planning tools that explicitly considers raw material variability, a chance-constrained (CC) model formulation, increases secondary material use in the case of aluminum [11, 12]. Here we explore the consequences of underlying assumptions related to this formulation that may still underestimate the potential for scrap incorporation. In particular, most batch planning or blending models [13] and particularly those used for the metals industry, model the input quality distributions as normal, an assumption that has been challenged previously [14-16] and that we will show empirically here for the

case of aluminum. As an example, in the contribution by Van Schaik and Reuter, which focused (among other contributions) on the impact of distributions of product lifetime, composition, weight and demand on achievable recycling rate for automobiles (using an approach that would be extensible to other products) a series of Weibull distributions were used [16]. In the remainder of this introduction, we first provide background on blending models, then discuss the frequency of non-normal distributions, and finally we describe ways the literature has expanded the formulation beyond this assumption.

Optimal blending models have been a topic of study for several decades, including the use of linear programming to identify the lowest-cost mix of raw materials to yield an end product of given specifications [17, 18]. These linear formulations are used in industry and described in the published literature for cases such as blending of petrochemicals [19, 20], agricultural products [21], and recycling [11, 22]. Where relevant, these models have also included effective treatments of the thermodynamics of the system [23, 24]. Another extension of these models was to move beyond the deterministic treatment of raw material quality. Authors have called for strategies to address raw materials quality uncertainty as far back as 1957 [25]. This early work by Debeau led to the development of strategies that incorporate information about raw material quality variation within a linear performance constraint [26-29]. These approaches have been shown to improve the robustness of the batch plan given uncertain inputs; however, they do not completely characterize variation in the final batch. Another strategy described in the literature uses linear performance constraints (generally based on the **mean** quality of raw materials) but manages variability with a penalty function in the objective [30-32]. A CC formulation of the performance constraints more explicitly models the implications of feedstock variation. Such a formulation is a mechanism to embed a more rich set of statistical information into optimization based decision models. Many firms track the statistical variation in their feedstocks; this method relates the desired level of confidence to the underlying standard deviations observed in these sampled raw materials. With the understanding that the compositional constraints will not be satisfied always due to the inherent uncertainty, they can be rewritten as probabilistic expression and transformed into their

deterministic equivalents. CC variants were first formulated by Charnes and Cooper [33]. This technique has found several applications in problems such as feed mixing [34], materials production [11, 35, 36], and coal blending [37]. This previous research has demonstrated the potential benefits of the CC-formulation through numerical experiments around specific cases. Previous work by a subset of the current authors provided an analytic characterization of the benefit of using a CC formulation as compared to one with a linearized constraint [12]. However, given the implementation described above, it becomes a necessary part of the blending model to comprehend the statistical nature of the uncertainty in the feedstocks, for the metals industry, this means knowing the type of distribution followed by compositional variability.

In many of the cases cited above involving both linear and CC programming, normal distributions are assumed to represent the probabilistic description of uncertain parameters. From a theoretical point of view, this assumption may be justified in the view of the central limit theorem when the underlying factors influencing an uncertain parameter have additive effect. In CC formulations this assumption has practical advantages as the analytical solution leads to an equivalent deterministic form that facilitates handling the associated optimization problem as shown in the methods section below.

However, in many applications including, but much broader than inorganics recycling, the underlying uncertainty is not well captured by a normal distribution [38]. Some physical quantities are intrinsically non-negative and a distribution with positive support (in other words, bounded below by zero) can be a more appropriate choice. Moreover, in many applications the statistical distributions of random quantities can be asymmetric. For instance, in many problems involving chemical and material processes the statistical distributions can be highly skewed [39]. In material engineering applications such as metals and ceramics, the skewed and, in particular, lognormal distributions are very frequently observed in parameters describing the structure of materials, such as grain size and area distributions, which consequently influence material properties [40, 41]. In finance applications, particularly in the context of portfolio selection, asymmetric distributions are frequently observed in the quantities such as cash flows

[42, 43]. In these situations, the assumption of normality to represent the underlying uncertainty may lead to statistical bias in the simulation of quantities of interest. Goicoechea and Duckstein [44] used a change of variable technique to construct deterministic equivalents for some probabilistic programming with non-normal distributions involving mutually independent random variables. Biswal et al. [45] investigated the probabilistic linear programming with independent exponential random variables. Kataria et al.[46] compared the effect of normal, truncated normal and lognormal distributions in a chance-constrained formulation of water pollution. Pagnoncelli et al. [47] studied the chance constraint portfolio selection when the returns follow a multivariate lognormal distribution.

The problem of optimizing the secondary material use under uncertainty in raw material specifications can be challenging [48]. From a mechanistic perspective, one of the main challenges is finding a balance between robustness and efficiency in the representation of the underlying optimization problem. In many cases the use of a surrogate (approximate) model is unavoidable due to computational complexity and data availability. Normal distributions for approximating the uncertainty in raw material specifications are used in this context as described above mainly due to its efficiency and computational flexibility. However, the robustness of this assumption must be evaluated and compared with other approximations that maintain a comparable level of efficiency.

This paper examines the implied impact, in terms of cost and secondary raw material usage, of modeling the underlying incoming materials quality distribution as normal. First, we examine the statistical characteristics of a set of secondary aluminum raw materials to demonstrate the challenges of the normal assumption. We then develop three alternative formulations for non-normal distributions, focusing on the situation where the incoming composition distribution is actually lognormal, and compare the overall performance for the case of aluminum recycling. The choice of lognormal distribution is of particular interest since in many real life applications the underlying distributions are more appropriately approximated as lognormal distribution. Since many real life populations are multiplicative rather than additive and, in view of central limit theorem, the multiplicative effect results in lognormal distribution.

Furthermore, many physical quantities are positive resulting in skewed distributions and the lognormal distribution provides a more reasonable model to capture this behavior. As mentioned above, other distributions, such as Weibull, have also been used. The comparison is made using an abstracted statement of the problem, applicable to any blending algorithm where the quality specification blends linearly, as well as for a specific set of raw materials from an aluminum remelter. We also identify characteristics of raw materials for which the assumption of normality would be most problematic.

Method

The goal of the method development was twofold, first to examine the statistical distributions for specific secondary raw materials and second to modify the CC formulation to incorporate non-normal distributions.

First, goodness of fit statistical tests were performed. A large set of compositional data (approximately 500 data points) was gathered for aluminum scraps over the course of five years from three industrial aluminum remelters. This set of historical data represented more than 100 types of non-ferrous scrap types or grades; seven grades were chosen that each had over 80 distinct compositional samples and included used beverage cans, mixed wrought automotive body, siding, automotive castings, litho sheets, and general traded scrap. The chi-squared test determines how closely a real set of data follows a theoretical distribution with the metric χ^2 defined in Eq. 1, where O is an observed frequency and E is an expected frequency for the distribution of interest. The number of samples or observations is n.

$$\chi_o^2 = \sum_{i=1}^{n} \frac{\left(O_i - E_i\right)^2}{E_i} \tag{1}$$

Most generally, the chance constrained optimization method is used to determine an optimal composition of raw materials. The objective is to fulfill a set of demands for finished products at the lowest cost given available raw materials. For the case of aluminum these materials include primary material and alloying elements (both with a narrow uncertainty in quality) as well as scrap materials (with

a wide uncertainty in quality). Due to the uncertainty in the compositional specifications for different scrap types, the element concentrations in the final products will vary. As mentioned above, previous research shows that incorporating this uncertainty into the batch mixing optimization can increase the potential use of scrap materials. The CC formulation provides a computational setting to sufficiently account for this uncertainty. The objective is to ensure that with a certain confidence, the concentrations of different element in the finish goods remain within a prescribed threshold based on the products' specifications. The CC formulation can be described as follows:

$$Min.: \quad C(x) = \sum_i C_i X_i \tag{2}$$

Subject to:

$$\forall_i \sum_j x_{ij} = X_i \leq A_i \tag{3}$$

$$\forall_j \sum_i x_{ij} = B_j > M_j \tag{4}$$

$$\forall_{j,k} \quad \mathbf{Pr}\left\{\sum x_{ij}\epsilon_{ik} \leq B_j E_{jk}^{max} \right\} > \alpha^{max} \tag{5}$$

$$\forall_{j,k} \quad \mathbf{Pr}\left\{\sum x_{ij}\epsilon_{ik} \geq B_j E_{jk}^{min} \right\} > \alpha^{min} \tag{6}$$

Where the variables are defined below:

C_i: unit cost($/T) of raw material *i*

x_{ij}: mass of raw material *i* used in making finished good j

X_i: mass purchased raw material *i*

A_i: mass of raw material *i* available for purchasing

B_j: mass of finished good *j* produced

M_j: mass of finish good j demanded

ϵ_{ik}: mass element of k in raw material i (uncertainty quantity)

E_{jk}^{max} : maximum specification for mass element k in finished good j

E_{jk}^{min} : minimum specification for mass element k in finished good j

α^{max} : confidence level that the actual composition will fall below the upper limit of final alloy composition

α^{min} : confidence level that the actual composition will fall above the lower limit of final alloy composition

Pr$\{\cdot\}$ represents the probability or the likelihood and is defined with respect to the probability distribution of random variable ϵ_{ik}.

The above stochastic optimization is classically solved by transforming eqs. 5 and 6 to equivalent deterministic inequalities. When ϵ_{ik}'s are normally distributed this transformation is exact, that is, analytical forms for the equivalent deterministic constraints exist. However, such a closed form transformation is not generally available when ϵ_{ik}'s do not include normal distributions. Using some assumptions, however, the probabilistic constraints can be approximately transformed to deterministic inequalities, which facilitate solving the optimization problem. This approach assumes that most non-metal contamination in the scrap stream has been dealt with prior to the melting stage via sorting and segregation. Blending models typically only account for the compositions of metals within the scrap stream, although the overall yield which is captured does account for the fact that there may be some remaining contamination.

For the problem at hand, we assume that the compositions of the constituent raw materials ϵ_{ik}'s follow a lognormal distribution. We then demonstrate several approximation methods to solve the associated CC problem. The objective is then to study how this different approximation can influence the robustness of resulting batch. For this purpose, Monte Carlo simulations are performed to compute the error rate of resulting batch for each method, where the error rate is defined as the percentage of batches that fall

outside of the finished goods specifications. Here we describe reformulating the compositional constraints based on an approximation using normal distributions (the traditional approximation), lognormal distributions (the distribution closest to the actual uncertainty), a convex approximation, and a linear approximation using fuzzy chance constraints.

Normal approximation

Assuming that the compositional specifications can be sufficiently approximated with normal distributions, the equivalent deterministic constraints associated with eqs. 5 and 6 can be expressed as:

$$\forall_{j,k} \quad \sum x_{ij}\bar{\epsilon}_{ik} \quad + \Phi^{-1}(\alpha^{max})\left(\sum_i \sum_l \rho_{ilk}\sigma_{ik}\sigma_{lk}x_{ij}x_{lj}\right)^{1/2} \leq B_j E_{jk}^{max} \tag{7}$$

$$\forall_{j,k} \quad \sum x_{ij}\bar{\epsilon}_{ik} + \Phi^{-1}(1-\alpha^{min})\left(\sum_i \sum_l \rho_{ilk}\sigma_{ik}\sigma_{lk}x_{ij}x_{lj}\right)^{1/2} \geq B_j E_{jk}^{min}\,, \tag{8}$$

where:

$\bar{\epsilon}_{ik}$: average mass element *k* in raw material *i*

σ_{ik}: standard deviation of the composition of element *k* in raw material *i*

ρ_{ilk}: correlation coefficient between composition of element *k* in raw materials *i* and *l*

($\forall_k \ \rho_{ilk} = 1 \ for \ i = l$)

It should be emphasized that this transformation would be exact if ϵ_{ij}'s followed a normal distributions. For the problem under consideration, however, this representation involves approximating the lognormally distributed random variables, ϵ_{ik}'s ,with normal distributions.

Lognormal approximation

When the compositional specifications, ϵ_{ik}'s , follows a multivariate lognormal distribution, a closed form expression for the chance constraint problem is not available since the distribution of a sum of lognormal random variables, $\sum x_{ij}\epsilon_{ik}$, is not known. One way to circumvent this difficulty is to approximate the sum with a lognormal distribution. The problem of sum of lognormal random variables is

encountered in many applications including signal processing as well as financial theory. Different techniques have been suggested to identify the parameters of the resulting approximated lognormal distribution for the sum. One common and simple method is Fenton's method [49]. In this method a lognormal approximation is obtained for the sum of independent lognormally distributed random variables by matching the moments. The method was later extended for the case of dependent random variables in [50]. Using Fenton's approximation the equivalent deterministic constraints associated with eqs. 7 and 8 can be expressed as [47]:

$$\forall_{j,k} \quad \mathbf{2}\ln\left(\sum_i x_{ij}k_{ik}\right) - \frac{1}{2}\ln\left(\sum_i\sum_l m_{ilk}x_{ij}x_{lj}k_{ik}k_{lk}\right) + \Phi^{-1}(\alpha^{max})\left(\ln\sum_i\sum_l m_{ilk}x_{ij}x_{lj}k_{ik}k_{lk} - 2\ln\left(\sum_i x_{ij}k_{ik}\right)\right)^{\frac{1}{2}} \leq \ln\left(B_j E_{jk}^{max}\right) \tag{9}$$

$$\forall_{j,k} \quad \mathbf{2}\ln\left(\sum_i x_{ij}k_{ik}\right) - \frac{1}{2}\ln\left(\sum_i\sum_l m_{ilk}x_{ij}x_{lj}k_{ik}k_{lk}\right) + \Phi^{-1}(1-\alpha^{min})\left(\ln\sum_i\sum_l m_{ilk}x_{ij}x_{lj}k_{ik}k_{lk} - 2\ln\left(\sum_i x_{ij}k_{ik}\right)\right)^{\frac{1}{2}} \geq \ln\left(B_j E_{jk}^{min}\right) \tag{10}$$

where k_{ik} and m_{ilk} are defined as follows,

$$k_{ik} = \exp\left(\bar{\epsilon}_{ik} + \frac{1}{2}\sigma_{ik}^2\right) \tag{11}$$

$$m_{ilk} = \exp(\rho_{ilk}\,\sigma_{ik}\sigma_{lk}) \tag{12}$$

Bernstein's convex approximation

Another approach to make the chance constraint problem computationally tractable is to use a convex approximation of the probabilistic constraints. A class of analytical approximation can be obtained making use of the theoretical bounds on the probability of violating the constraints, and as a result converting the probabilistic constraints to the convex deterministic ones. Some of these inequalities include Chebyshev's inequality, Bernstein's inequality, Hoefding's inequality (see [51], [52]). These approximations are generally conservative, but they can provide a reasonable approximation with a low computational cost for the original problem. Ben-Tal derived a convex approximation of the constraints for unknown, but bounded uncertainty [52]. A very similar form was also proposed by Nemirovski and Shapiro based on Bernstein inequality for the so called "case of ambiguous chance constraints", where one does not rely on the exact knowledge of the underlying probability distributions of random variables [53]. For the problem under consideration, these convex approximations of eqs. 7 and 8 take the following forms:

$$\forall_{j,k} \quad \sum x_{ij}\bar{\epsilon}_{ik} \quad + \sqrt{2\log(1/\alpha_k^{max})}\left(\sum_i x_{ij}^2\left(\frac{u_{ik}-l_{ik}}{2}\right)^2\right)^{1/2} \leq B_j E_{jk}^{max}\,, \tag{13}$$

$$\forall_{j,k} \quad \sum x_{ij}\bar{\epsilon}_{ik} - \sqrt{2\log(1/\alpha_k^{min})}\left(\sum_i x_{ij}^2\left(\frac{u_{ik}-l_{ik}}{2}\right)^2\right)^{1/2} \geq B_j E_{jk}^{min}\,, \tag{14}$$

where α_k^{max} and α_k^{min} control the confidence level for the Max. and Min. constraints, respectively and we have : $\sum_k \alpha_k^{max} \leq \alpha^{max}$ and $\sum_k \alpha_k^{min} \leq \alpha^{min}$. This representation requires the knowledge of appropriate upper bound u_{ij} and lower bound l_{ij} corresponding to the compositional specifications ϵ_{ik}. In this work we set the upper and lower bounds to be, respectively, the 95th and 5th percentiles for ϵ_{ik}'s based on their lognormal distributions.

Linear approximation using fuzzy chance constraints

CC programming can be expressed in a fuzzy environment, which can provide an alternative approach to deal with uncertainty in the constraints [54]. Making use of fuzzy set theory, this method relies on

representing the uncertainty in the form of possibility measure, rather than probability measure. In this representation, the uncertain parameters are expressed as fuzzy numbers with prescribed membership functions (possibility distribution), analogous to probability distributions. Liu discussed different cases where the chance constraints with fuzzy numbers can be converted to their respective crisp equivalence [54]. Rong and Lahdelma applied this approach to the problem of scrap charge optimization in steel production [36]. The stochastic parameters were represented as triangular fuzzy numbers and the probability distributions were transformed into the respected triangular possibility distributions. The resulting fuzzy programming model can then be converted into a deterministic linear problem making use of the crisp equivalent representation of the fuzzy constraints.

Let $e_{ik} = (\hat{e}_{ik}, e_{ik}^{l}, e_{ik}^{u})$ be the fuzzy number representation of ϵ_{ik}, where $\hat{e}_{ik}$ is the mean and $(\hat{e}_{ik} - e_{ik}^{l})$ and $(e_{ik}^{u} - \hat{e}_{ik})$ are the left and right spread of the fuzzy number. A set of crisp equivalence for the chance constraints in eqs 7 and 8 can be represented as:

$$\forall_{j,k} \quad \sum x_{ij}(\hat{e}_{ik} + (2\lambda_{ik} - 1)(e_{ik}^{u} - \hat{e}_{ik})) \quad \leq B_j E_{jk}^{max}\,, \tag{15}$$

$$\forall_{j,k} \quad \sum x_{ij}\big(\hat{e}_{ik} - (2\lambda_{ik} - 1)(\hat{e}_{ik} - e_{ik}^{l})\big) \quad \geq B_j E_{jk}^{min}\,, \tag{16}$$

This formulation requires tuning parameters $\hat{e}_{ik}$, e_{ik}^{l} and e_{ik}^{u} appropriately based on the probability distribution of ϵ_{ik}. For a triangular fuzzy number these parameters can be expressed as $e_{ik}^{l} = \hat{e}_{ik} - n_1\sigma_{ik}$ and $e_{ik}^{u} = \hat{e}_{ik} + n_2\sigma_{ik}$, where n_1 and n_2 depend on the shape of probability distributions of ϵ_{ik}. The level of confidence is controlled by $(2\lambda_{ik} - 1)n_1$ and $(2\lambda_{ik} - 1)n_2$. In a practical setting these parameters can be tuned such that different failure rate is fulfilled.

Individual ways of representing the statistical characteristics of the lognormally distributed raw material compositions were needed for each the formulations described above. For the normal case, for each lognormally distributed composition of the scrap, a normal distribution is fitted and the

corresponding mean and standard deviation are computed. The optimization problem is then solved for different values of $\alpha_{ik}^{max} = \alpha^{max}$ and the associated error rate is computed using a Monte Carlo simulation. For the convex approximation the upper and lower bounds u_{ik}, l_{ik} are set to be the 95th and 5th percentiles of ϵ_{ik}'s based on their lognormal distributions. The optimization problem is solved with different confidence level by varying α_k^{max} and the associated error rate is computed using a Monte Carlo simulation. Finally for the fuzzy method the $\hat{e}_{ik}$ is set to the mean value of compositional specification: $\hat{e}_{ik} = \bar{\epsilon}_{ik}$. The optimization problem is solved for different set of values for parameters n_2 and $\lambda_{ik} = \lambda$ such that different failure rate is obtained.

Results

The goals of this analysis were to understand the limitations of formulating a CC batch planning algorithm based on normally distributed incoming raw materials quality when the actual uncertainty of the raw materials is lognormally, or otherwise, distributed. Specifically this work examined alternate ways to formulate the raw materials quality within CC algorithms for the particular case of aluminum remelting and compared these formulations to the traditional normal formulation to determine whether increased scrap or reduced cost could be realized.

We began the assessment with a quantitative examination of the compositional distributions of a large set of aluminum scraps gathered over five years from three industrial aluminum remelters. To illustrate these distributions, two example frequency histograms for iron and copper compositions in a scrap are shown in Figure 1a and b for one automotive scrap. In addition a set of box and whisker plots for a different scrap type over the same time period. Each box and whisker in Figure 1c are from at least 80 data points In all of the results shown, the values are clustered around minimum values with a tail at the high end of composition suggesting a skewed probability distribution.

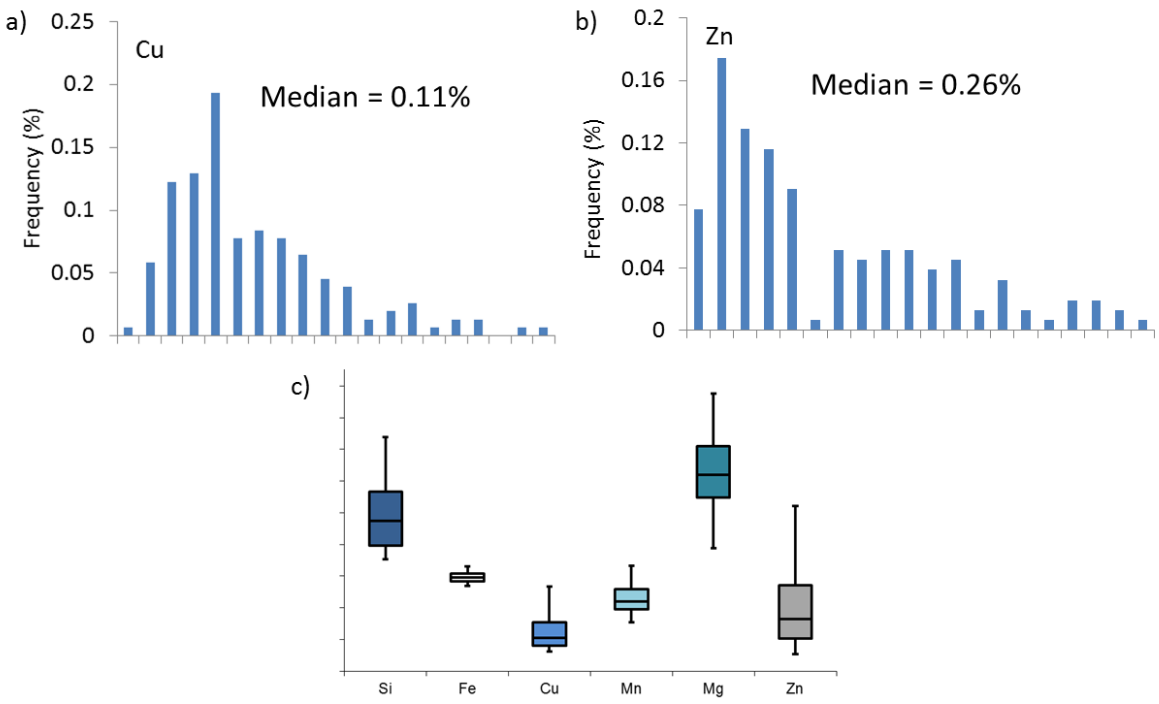

Figure 1. Compositional histograms from two elements, a) Cu and b) Fe within scrap aluminum alloys, c) box and whisker for six elements within a different scrap sample (median, 25th and 75th percentiles shown in the box, 5th and 95th in whisker)

In order to demonstrate a more quantitative understanding of the common distributions for alloying element composition within incoming raw materials , seven different scrap grades from an aluminum secondary processor were analyzed for best fit with results shown in Table 1. We determined the appropriateness of other probability distributions for the following distributions: normal, logistic, lognormal, maximum extreme, gamma, beta, Weibull, exponential, triangular, uniform, and Student's T according to the Anderson-Darling and chi-squared fit tests. A statistical fit to these data on scrap composition for a wide range of common major elements including lead, copper, silicon, iron, manganese, chromium, tin, zinc, and zirconium showed that there was a wide range in types of probability distributions and none were normally distributed. Of the seventy cases (seven scrap grades looking at ten elemental compositions), the most common distribution by a factor of two was the lognormal distribution (44%). The next most common type was the logistic (21%) followed by max extrema (16%). These distributions have higher weighting towards the lower end of the distribution. The

closest to a normal distribution was the Student's t fit for a few elements such as manganese, magnesium, and titanium.

Table 1. Best fit distributions by element for ten common alloying elements for seven scraps commonly used in aluminum remelting.

	Si	Fe	Cu	Mn	Mg	Cr	Zn	Ti	Zr	Pb	Sum
gamma	3	1				1					5
max extreme		3		2	1	1			4		11
lognormal	1		7	1	2	1	6	3	3	7	31
logistic	2	3		2	3	2		3			15
students				1	1			1			3
weibull						1	1				2
beta	1			1		1					3

Table 1 shows that for some elements the compositional distribution was almost always lognormal as is the case for copper, zinc and lead. Titanium, and zinc quantities are also more commonly lognormally or logistically distributed. Based on the evidence shown in Figure 1 and overall trends from the scraps investigated in Table 1, we found that the elemental compositions are not typically normally distributed. This result was also consistent with findings a subset of the authors found on a different data set from a different aluminum manufacturer[55]. We found that the coefficient of variation (standard deviation divided by the mean for the normal equivalent) for the compositional data examined varied quite widely -- as low as 6% to almost 70%.

The primary contribution of this work was to understand the implications of modifying the assumption of normality in the CC formulation, particularly as the first analysis showed very few compositions are normally distributed. Two analyses were made to illustrate the impact of varying distribution assumptions on overall cost and secondary materials use particular to the case of aluminum remelting. The importance of this application is that the compositions of elements are fractions of percent and are weighted towards the lower part of the distribution as described above. The first analysis was an abstract model developed to show the general trends behind each of the formulations described in the methods section. For this case, an abstract example was developed that incorporated only one

compositional constraint and the specifications for the scrap materials were all identical. In addition, the availability and price parameters for each of the scrap materials were also identical. The second analysis was for a case derived from an actual remelter. For this case, the number of tracked elements is six and the concentrations of these elements within either three or six scraps (depending on the scenario) were based on actual data from an aluminum remelter (normalized to protect sensitive information). Further detail is provided on the real case below. For both analyses, the batch plan was developed using a CC formulation for only one final product specification, although there is nothing in the formulation that limits increasing this number. This was done to simplify the presentation of the results below.

In order to evaluate the performance of different methods, a Monte Carlo simulation was performed to estimate the error rate of meeting the compositional specifications of the final product associated with the resulting batch for each method. The numerical values of scrap specifications are sampled from the corresponding lognormal distributions. The error rate is defined as the percentage of batches in N_{mc} (number of Monte Carlo runs) for which at least one of the compositional specifications falls outside of the final product specifications.

We begin with the results of the abstract case. Figure 2a shows the normalized production cost versus the error rate for different ways of formulating the compositional specifications (each plot is normalized by the maximum cost or scrap use in each case to facilitate interpretation by the reader). The production cost was determined from the unit material cost for each raw material and the quantity for each material determined from the CC batch plan. Error rate is varied as described above. The normalized amount of scrap usage versus the error rate is compared in Figure 2b (again normalized by the highest value for scrap in each case). These results assume the uncertainty in the composition is 30% of the mean with three scraps, based on the quantitative modeling above. This is a middle value of uncertainty in composition. Both figures show similar trends as would be expected for the abstract case. The batches resulting from solving the CC formulation with lognormal, normal, and Bernstein approximations have

similar lower cost and higher scrap usage than the fuzzy approximation for a given level of error rate. And as error rate increased, the amount of scrap used increases with a plateau at higher error rates.

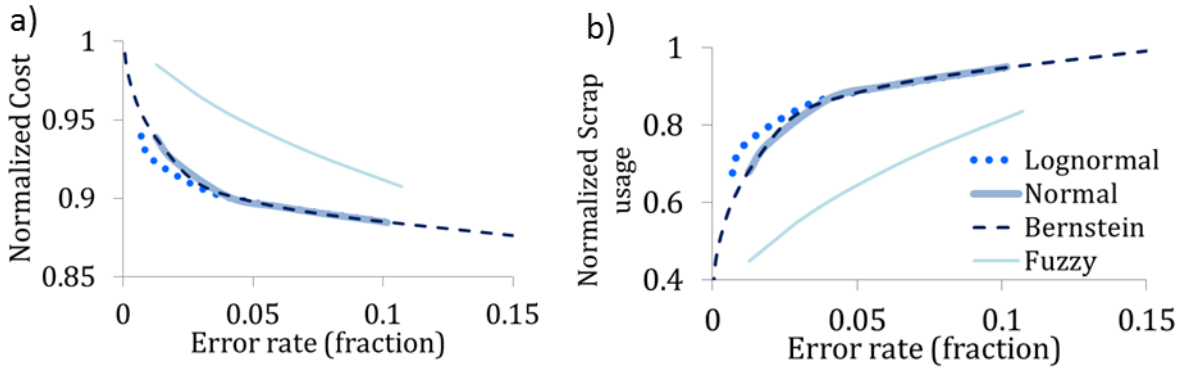

Figure 2: Abstract case of 30% coefficient of variation a) Normalized production cost versus error rate for different methods. b) Normalized scrap usage versus error rate for different methods.

If instead we have a level of uncertainty of 70%, as shown in Figure 3, the separation between the formulations is now more apparent where the lognormal exhibits ~10% decrease in cost and increase in scrap usage than the normal and Bernstein approximations. By approximating the middle range of scrap usage as a line (with R^2 of 0.9) we learn that as error rate decreases, the methods are more distinct: the lognormal exhibits closer to 16% more scrap use than the normal or Bernstein approximations for error rates of 1%. As shown previously [12], the benefit of the explicitly capturing uncertainty method increases as the uncertainty within raw materials quality increases so the difference between the approximations shows a similar trend between 30% and 70%. In other words, as the scraps become more uncertain, formulating the constraint closer to their actual behavior becomes more significant. Because the approximation scheme used for the case of fuzzy representation of chance constraint is linear in form, it is not surprising that it underperforms in terms of cost and scrap usage. Another observation comparing Figure 2 to Figure 3 is that the cost increases and scrap use decreases as would be expected because higher uncertainty in the lower quality raw materials means that increased primary is required to meet the final product specification. Scrap use for the 30% uncertain case reaches almost 800 kg versus only 600

for the 70% uncertain case. Figure 4 summarizes the scrap use at the two different coefficients of variation by distribution method at a 10% error rate (0.1 fraction).

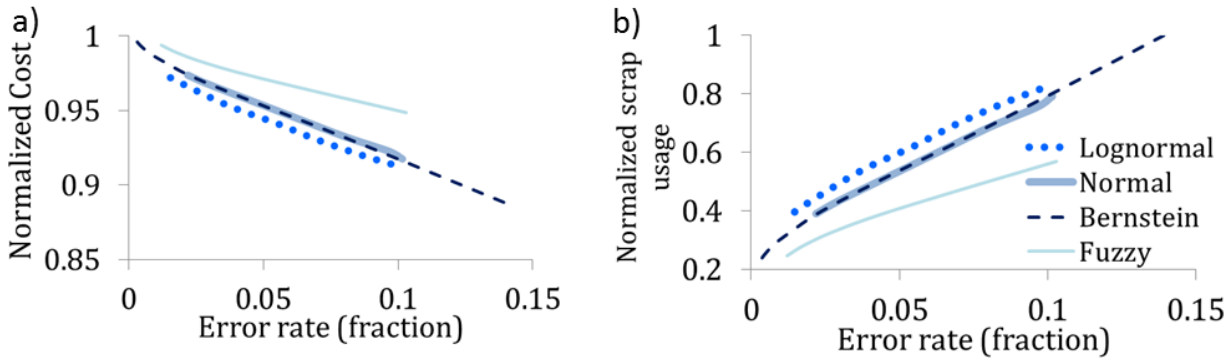

Figure 3. Abstract case of 70% coefficient of variation a) Normalized production cost versus error rate for different methods. b) Normalized scrap usage versus error rate for different methods.

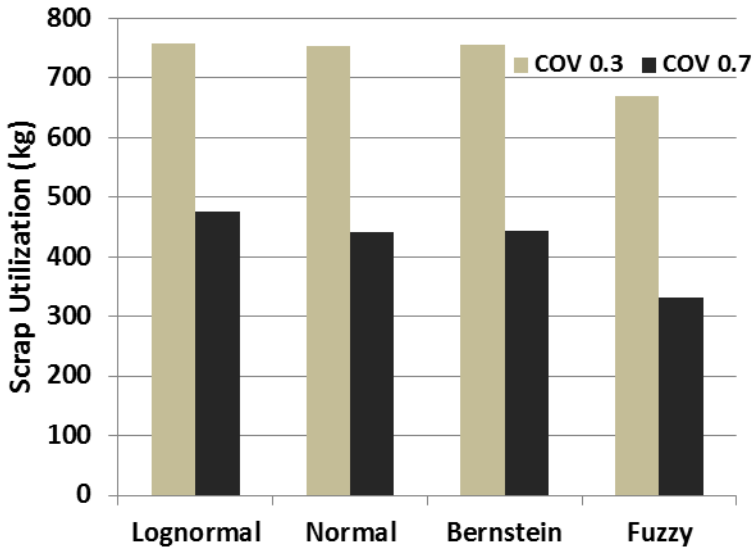

Figure 4. Scrap use by distribution approximation for an error rate of 0.1 (10%) for COV of 30% (tan) and 70% (black).

We also explored the impact of additional numbers of scraps on the problem. Figure 5 shows a similar trend between 30% COV and 70% COV for the case of 6 scraps, but just for the metric of cost as the trend in scrap usage is similar. Here the axes are the same scale so a more direct comparison of cost is possible. The difference between the lognormal and normal approximations for the 70% COV scrap usage is again around 10% averaged over the error rates. Normal and Bernstein approximations still behave very

similarly. Therefore, the number of uncertain raw materials used by a particular producer does not seem to have strong influence over the approximation used in the compositional constraint.

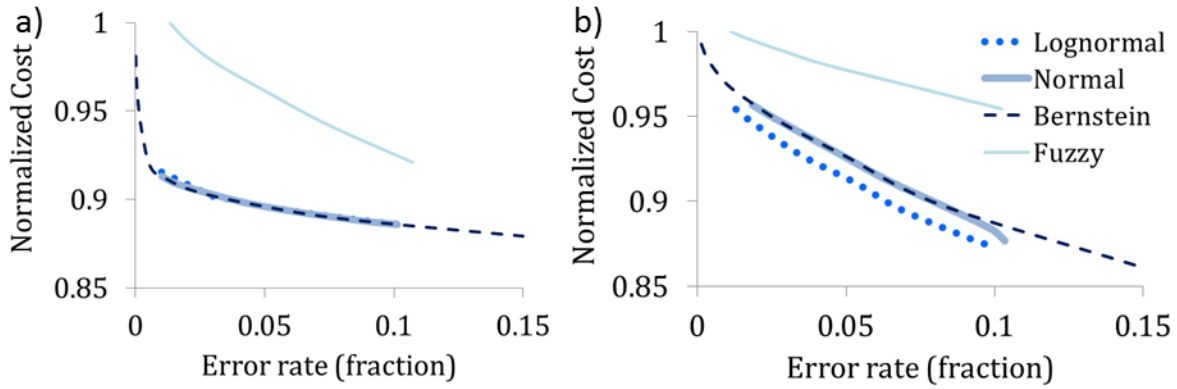

Figure 5: Abstract case of production cost for six secondary raw materials a) 30% coefficient of variation and b) 70% coefficient of variation.

Real production cases are more complex than the abstract analysis developed above. As such, the final analyses explored the performance of the various constraint formulations in the context of a real-world case of aluminum recycling. Because we did not see such a difference in the number of scraps in the analysis shown above, we only varied the compositional uncertainty derived from actual raw materials for the real case as shown in Figure 6. For this case, the models examine the problem of mixing quantities of up to six scraps with six relevant alloying elements available (Cu, Fe, Mg, Mn, Si, and Zn) and primary raw materials, all of unlimited availability, to produce an identical mass of one final product at lowest cost. The total demand for the product was based on one months' production at a typical remelter for the kinds of alloys under investigation, approximately 1.5 ktons. The elements chosen for this analysis have been identified by literature and industry as having a) the highest degree of uncertainty, combined with b) being the most problematic from a specification standpoint. Scrap streams are certainly dynamic and the problematic tramp elements of today will not be the problematic tramp elements of tomorrow. For example, as less and less cast is being produced, it in turn will end up as scrap less frequently and silicon may no longer become a problematic contaminant. The increasing use of advanced metals making use of specialty alloying elements such as chromium, nickel, vanadium, molybdenum, and zirconium, among

others, may indicate emerging contaminant elements for the future [56-58]. This approach is adaptable and expandable to accommodate whatever elements are of concern in the scrap stream of interest. As described in the introduction, this approach is also applicable beyond metal recycling and could be used for other blending models. However, these blending formulations cannot be used in the absence of fundamental considerations of thermodynamics of remelting. Changing and uncertain quality of recyclates has a large impact on the potential for recycling to offer the energy benefits suggested above [22, 24].

The case selected for this study represents the production decisions of a European aluminum recycler that produces a broad range of alloys including the final product being modeled. The scraps represent raw materials that were available and used by that producer during 2012 and normalized scrap composition data are provided in the appendix. The availability and price information are taken from data provided by industry. The elemental compositions are assumed to be uncorrelated. The baseline is to assume that the distributions are lognormally distributed and modeled as described above. For this analysis the linear approximation based on the fuzzy method was dropped as the performance was consistently lower for the abstract case.

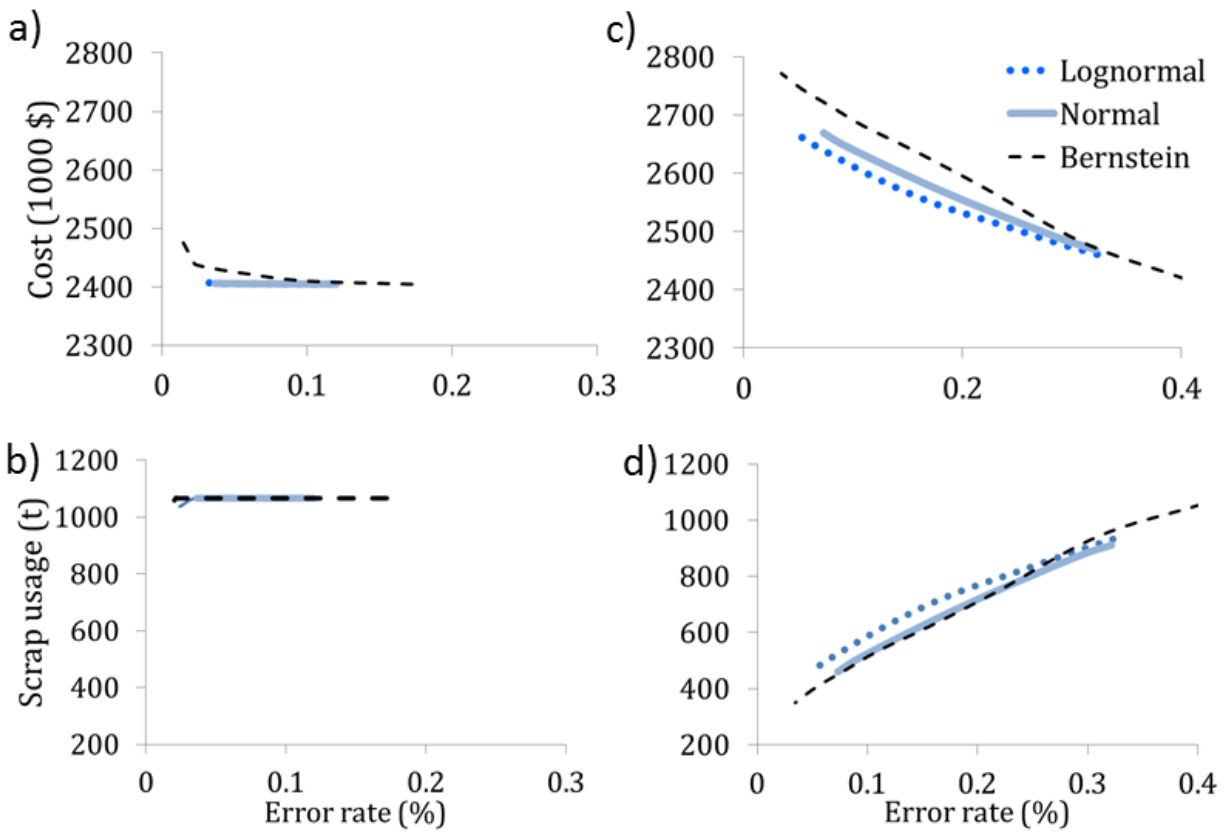

Figure 6. Actual case results a) Production cost versus error rate and b) scrap usage for different methods for 30% coefficient of variation assumption on scraps. c) Production cost versus error rate and d) scrap usage for different methods for 70% coefficient of variation assumption on scraps.

Figure 6 shows that similar trends were found for the real case as compared to the abstract analysis. First, for the case where the scraps had a 30% coefficient of variation, there was little difference in both cost and scrap usage among the CC formulations. However, for the 70% case, the lognormal distribution formulation shows an increased scrap use of 14%, relative to the normal and Bernstein approximations, and corresponding decrease in cost relative to the normal distribution. Because the best fit distribution found for the scrap data was lognormal, this could explain why this increased scrap use was found. In the abstract case it was assumed that the actual distributions of compositions are lognormal. The objective is see the error or performance of other approximation methods compared to the solution based on the lognormal. Because the actual distribution is lognormal, it is expected that the lognormal approximation would perform best (i.e. have a low error rate with high scrap utilization). In the case of real

compositional data, based on the goodness of fit results, the lognormal is in fact the best fit for more than 40% of the element-scrap type compositions. The best fit distributions for the rest also turned out to be skewed distributions, which can be sufficiently approximated as lognormal in most cases. These observations justify why we observe a better performance when we use a lognormal approximation for the chance constrained problem.

Results of the chance constrained formulation approach were calibrated with actual batch plans at an industrial partner; utilization rates and "first-on" error rates were found to be in excellent correlation with the predicted modeling values. Actual costs will obviously be much more complex than the simplified materials cost represented in this work. "First-on" error rates were used as in reality, the composition of the blend is tested to see if it is within specification and then modified via addition of alloying elements or fluxes or diluted with primary aluminum; in practice, there are a very small number of off specification batches that cannot be used for production. This aspect of dilution with primary has the largest impact on the environmental impacts of the process as described below.

Leveraging CC modeling to explicitly manage compositional uncertainty has been implemented previously to increase scrap utilization [12]. In this work, we have shown that incorporating statistical distributions that represent that uncertainty based on distributions that more match that of the actual composition also increases scrap use. This increase in scrap use may translate directly to energy savings. Readers are referred to previous work by a subset of the authors regarding energy savings relative to batch planning without explicit consideration of uncertainty (frequently leading to conservative scrap use) [11]. Here we offer an estimate of energy savings based on the incorporation of different distribution functions. Primary aluminum production uses on average 175MJ per kg while secondary production from old scrap uses on average a fraction of that amount or roughly 12 MJ per kg for a difference of approximately 160 MJ/kg although there is some variation in this number [59]. Comparing the results for the batch plans using the lognormal formulation with that of the normal and Bernstein (the latter two are quite similar to each other) shown in Figure 6d we see that approximately 66 more tonnes of scrap are

used (as a maximum) in a production plan of about 1.5 kt when these distributions are varied. This translates to approximately 11 million MJ of energy savings for the alternate distribution functions. Employing this approach to managing uncertainty in the scrap stream allows for 14% less dilution of finished alloys with primary aluminum in order to achieve compositional specifications. Given the 2014 secondary production within the United States of 3.64 million metric tons, this translates to savings as shown in Table 2. The US is not a major producer of aluminum and 2014 was a comparatively depressed commodity year so these savings have the potential, globally and in the future, to be much larger. However, this number is likely an upper bound for several reasons. First, we note that increased use of secondary may not directly lead to displacement of primary production, as found by several previous authors [60, 61], In addition, the energy impact will be highly dependent on the compositions (and types) of scrap, which influences their yield in a remelting setting. Future work could example these aspects in more detail as they are indirectly related to the distributions modeled in the batch planning algorithm.

Table 2. Potential energy savings from enhanced scrap utilization in industrial case

	Energy
Primary production	175 MJ/kg
Secondary production	12 MJ/kg
Savings	93%
Potential US Savings	83,064 million MJ/year

Conclusion

The major contribution of this work was to develop a formulation for non-normally distributed raw materials that could be incorporated into a chance constrained model for batch planning particular to the aluminum industry. To this end, several formulations were explored including one based on Fenton's method, Bernstein inequalities and an approximation using fuzzy set theory. These formulations were then compared to the more typical normally distributed approximation for an abstract case as well as one based on the conditions for an aluminum remelter. As shown previously, the quantitative assessments of the characteristic distributions for elemental composition within scraps from this remelter were found to rarely be normal. This makes sense based on the non-negative characteristic of scrap quality, i.e.

elemental compositions cannot be less than zero. The most common distribution for the industrial aluminum secondary materials was lognormal and the lognormal approximation formulation based on this had the highest scrap use (and lowest cost) batch plan. This would be a recommended approach for metals producers to incorporate in their blending models when they are working with scrap streams with well-characterized historical compositions. The benefit of the Bernstein inequality- based formulation is that no underlying distribution must be known, so for cases where compositional data has not been collected, this approach could be beneficial.

Acknowledgements

The authors would like to acknowledge funding from the National Science Foundation Award #1133422, Environmental Sustainability Program (ENG-CBET) that provided support to make this work possible.

References:

1. Keoleian, G., *Industrial Ecology of the Automobile*. 1997.
2. Vigeland, P., *Aluminium recycling: the commercial benefits, the technical issues and the sustainability imperative.* Hydro Aluminium Recycling, 2001.
3. Das, S.K. and M. Hughes, *Improving aluminum can recycling rates: a Six Sigma study in Kentucky.* JOM, 2006. **58**(8): p. 27-31.
4. Morgan, F.W. and M.V. Hughes, *Understanding recycling behavior in Kentucky: Who recycles and why.* JOM, 2006. **58**(8): p. 32-35.
5. Woodward, R., *Where next wrought aluminium alloys?* Aluminium International Today, 1997. **9**(6): p. 21.
6. Gesing, A., *Assuring the continued recycling of light metals in end-of-life vehicles: A global perspective.* JOM, 2004. **56**(8): p. 18-27.
7. Ehrenfeld, J. and N. Gertler, *Industrial ecology in practice.* Journal of industrial Ecology, 1997. **1**(1): p. 67-79.
8. Gregory, J.R., M.-C. Nadeau, and R.E. Kirchain, *Evaluating the economic viability of a material recovery system: the case of cathode ray tube glass.* Environmental science & technology, 2009. **43**(24): p. 9245-9251.
9. Xiao, Y. and M. Reuter, *Recycling of different aluminium scraps.* Minerals Engineering, 2002. **15**(2002): p. 763-970.
10. Khoei, A., I. Masters, and D. Gethin, *Design optimisation of aluminium recycling processes using Taguchi technique.* Journal of Materials Processing Technology, 2002. **127**(1): p. 96-106.
11. Gaustad, G., P. Li, and R. Kirchain, *Modeling methods for managing raw material compositional uncertainty in alloy production.* Resources, Conservation and Recycling, 2007. **52**(2): p. 180-207.
12. Olivetti, E.A., et al., *Increasing secondary and renewable material use: a chance constrained modeling approach to manage feedstock quality variation.* Environmental science & technology, 2011. **45**(9): p. 4118-4126.

13. Vasko, F., D. Newhart, and A. Strauss, *Coal blending models for optimum cokemaking and blast furnace operation.* Journal of the Operational Research Society, 2005. **56**(3): p. 235-243.
14. Peterson, R.D., *Scrap variability and its effects on producing alloys to specification.* LIGHT MET(WARRENDALE PA), 1999: p. 855-859.
15. Rusin, M., *The structure of nonlinear blending models.* Chemical Engineering Science, 1975. **30**(8): p. 937-944.
16. Van Schaik, A. and M. Reuter, *The time-varying factors influencing the recycling rate of products.* Resources, Conservation and Recycling, 2004. **40**(4): p. 301-328.
17. Kim, J. and R. Lewis, *A large scale linear programming application to least cost charging for foundry melting operations.* AFS Transactions, 1986: p. 735-744.
18. Metzger, R. and R. Schwarzbek, *A linear programming application to cupola charging.* Journal of Industrial Engineering, 1961. **12**(2): p. 87-93.
19. Baker, T.E. and L.S. Lasdon, *Successive linear programming at Exxon.* Management Science, 1985. **31**(3): p. 264-274.
20. Mendez, C.A., et al., *A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations.* Computers & Chemical Engineering, 2006. **30**(4): p. 614-634.
21. Hutton, R.F. and J.R. Allison, *A linear programming model for development of feed formulas under mill-operating conditions.* Journal of Farm Economics, 1957. **39**(1): p. 94-111.
22. Reuter, M.A., et al., *Fundamental limits for the recycling of end-of-life vehicles.* Minerals Engineering, 2006. **19**(5): p. 433-449.
23. Amos, F., M. Ronnqvist, and G. Gill, *Modeling the pooling problem at the New Zealand Refining Company.* Journal of the Operational Research Society, 1997. **48**(8): p. 767-778.
24. Reuter, M. and A. van Schaik, *Thermodynamic metrics for measuring the "sustainability" of design for recycling.* JOM Journal of the Minerals, Metals and Materials Society, 2008. **60**(8): p. 39-46.
25. Debeau, D., *Linear programming isn't always the answer.* Operations Research, 1957. **5**(3): p. 429-433.
26. Tintner, G., *A note on Stochastic Linear Programming.* Econometrica, 1960. **28**(2): p. 490-496.
27. Dupacova, J. and P. Popela, *Melt Control: Charge Optimization*, in *Applications of Stochastic Programming*, S. Wallace and W. Ziemba, Editors. 2005, Math. Program. Soc.: Philadelphia, PA.
28. Bliss, N., *Advances in scrap charge optimization.* Transactions of the American Foundrymen's Society, 1997. **105**: p. 27-30.
29. Hartley, A., P. Eastburn, and N. Leece, *Steelworks control of residuals.* Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1980. **295**(1413): p. 45-55.
30. Evers, W., *A new model for stochastic linear programming.* Management Science, 1967. **13**(9): p. 680-693.
31. Karmarkar, U.S. and K. Rajaram, *Grade selection and blending to optimize cost and quality.* Operations Research, 2001. **49**(2): p. 271-280.
32. Mihailidis, D. and K.R. Chelst, *Optimal blending of out-of-specifications substances.* Journal of the Operational Research Society, 1998. **49**(5): p. 458-466.
33. Charnes, A. and W. Cooper, *Chance constrained programming.* Management Science, 1959. **6**: p. 73-89.
34. Panne, C.v.d. and W. Popp, *Minimum-cost cattle feed under probabilistic protein constraints.* Management Science, 1963. **9**(3): p. 405-430.
35. Kumral, M., *Application of chance-constrained programming based on multi-objective simulated annealing to solve a mineral blending problem.* Engineering Optimization, 2003. **35**(6): p. 661-673.

36. Rong, A. and R. Lahdelma, *Fuzzy chance constrained linear programming model for optimizing the scrap charge in steel production.* European Journal of Operational Research, 2008. **186**(3): p. 953-964.
37. Shih, J. and H. Frey, *Coal blending optimization under uncertainty.* European Journal of Operational Research, 1995. **83**: p. 452-465.
38. Pyzde, T., *Why Normal Distributions Aren't [All That Normal].* Quality Engineering, 1995. **7**: p. 769-777.
39. Chang, Y.S., I.S. Choi, and D.S. Bai, *Process capability indices for skewed populations.* Quality and Reliability Engineering International, 2002. **18**: p. 383-393.
40. Kurtz, S. and F. Carpay, *Microstructure and normal grain growth in metals and ceramics. Part I. Theory.* Journal of Applied Physics, 1980. **51**(11): p. 5725-5744.
41. Papadakis, E.P., *Grain-Size Distribution in Metals and Its Influence on Ultrasonic Attenuation Measurements.* The Journal of the Acoustical Society of America, 1961. **33**(11): p. 1616-1621.
42. Gurgur, C.Z. and J.T. Luxhoj, *Application of Chance-Constrained Programming To Capital Rationing Problems With Asymmetrically Distributed Cash Flows and Available Budget.* The Engineering Economist, 2003. **48**: p. 241-258.
43. Sengupta, A.J.K., *Economics Department of the University of Pennsylvania Institute of Social and Economic Research -- Osaka University A Generalization of Some Distribution Aspects of Chance-Constrained Linear Programming A GENERALIZATION OF SOME DISTRIBUTION ASPECTS OF CH.* 2014. **11**: p. 287-304.
44. Goicoechea, A. and L. Duckstein, *Nonnormal deterministic equivalents and a transformation in stochastic mathematical programming.* Applied Mathematics and Computation, 1987. **21**: p. 51-72.
45. Biswal, M.P., N.P. Biswal, and D. Li, *Probabilistic linear programming problems with exponential random variables: A technical note.* European Journal of Operational Research, 1998. **111**: p. 589-597.
46. Kataria, M., K. Elofsson, and B. Hasler, *Distributional Assumptions in Chance-Constrained Programming Models of Stochastic Water Pollution.* Environmental Modeling & Assessment, 2009. **15**: p. 273-281.
47. Pagnoncelli, B., Ahmed, Shabbir and A. Shapiro, *Computational study of a chance constrained portfolio selection problem.* Optimization Online, 2008.
48. Schlesinger, M.E., *Aluminum recycling*. 2013: CRC Press.
49. Fenton, L., *The sum of log-normal probability distributions in scatter transmission systems.* Communications Systems, IRE Transactions 1960. **8.1**: p. 57-67.
50. Abu-Dayya, A. and N.C. Beaulieu, *Outage probabilities in the presence of correlated lognormal interferers.* Vehicular Technology, IEEE Transactions on, 1994. **43**(1): p. 164-173.
51. Pintér, J., *Deterministic approximations of probability inequalities.* Zeitschrift für Operations Research, 1989. **33**(4): p. 219-239.
52. Ben-Tal, A. and A. Nemirovski, *Robust optimization–methodology and applications.* Mathematical Programming, 2002. **92**(3): p. 453-480.
53. Nemirovski, A. and A. Shapiro, *Convex approximations of chance constrained programs.* SIAM Journal on Optimization, 2006. **17**(4): p. 969-996.
54. Liu, B. and K. Iwamura, *Chance constrained programming with fuzzy parameters.* Fuzzy sets and systems, 1998. **94**(2): p. 227-237.
55. Gaustad, G., R. Peterson, and R. Kirchain. *Modeling methods to guide recycling friendly alloy design: the impact of compositional data structure*. in *Proceedings of the annual meeting of TMS: 9th global innovations symposium–trends in integrated computational materials engineering for materials processing and manufacturing, New Orleans, LA*. 2008.

56. Sillekens, W., D. Verdoes, and W. Boender, *Refining aluminium scrap by means of fractional crystallisation: Status and prospects for development.* METALL-BERLIN-, 2002. **56**(7/8): p. 468-473.
57. Nakajima, K., et al., *Thermodynamic analysis of contamination by alloying elements in aluminum recycling.* Environmental science & technology, 2010. **44**(14): p. 5594-5600.
58. Mahfoud, M., A.P. Rao, and D. Emadi, *The role of thermal analysis in detecting impurity levels during aluminum recycling.* Journal of Thermal Analysis and Calorimetry, 2010. **100**(3): p. 847-851.
59. McMillan, C.A. and G.A. Keoleian, *Not all primary aluminum is created equal: life cycle greenhouse gas emissions from 1990 to 2005.* Environmental science & technology, 2009. **43**(5): p. 1571-1577.
60. Geyer, R., et al., *Common misconceptions about recycling.* Journal of Industrial Ecology, 2015.
61. Zink, T., R. Geyer, and R. Startz, *A Market-Based Framework for Quantifying Displaced Production from Recycling or Reuse.* Journal of Industrial Ecology, 2015.

Appendix

Table S.1. Normalized mean composition relative to mean composition of the finished good.

	Si	Fe	Cu	Mn	Mg	Zn
Scrap 1	0.9	0.6	0.4	0.5	1.2	0.1
Scrap 2	0.6	1.1	0.7	2.1	0.2	0.0
Scrap 3	1.3	0.6	0.7	0.5	1.2	1.3