LCA Methodology

Screening Stochastic Life Cycle Assessment Inventory Models

Kelly G. Canter, Dale J. Kennedy, Douglas C. Montgomery, J. Bert Keats and W. Matthew Carlyle

Arizona State University, Tempe AZ., 85287-5906, USA

Corresponding author: Kelly G. Canter, 2227 W. Carmen St, Chicago, IL 60625, USA; e-mail: kcanter3@excite.com

DOI: http://dx.doi.org/10.1065/lca2001.08.063

Abstract. A screening methodology is presented that utilizes the linear structure within the deterministic life cycle inventory (LCI) model. The methodology ranks each input data element based upon the amount it contributes toward the final output. The identified data elements along with their position in the deterministic model are then sorted into descending order based upon their individual contributions. This enables practitioners and model users to identify those input data elements that contribute the most in the inventory stage. Percentages of the top ranked data elements are then selected, and their corresponding data quality index (DQI) value is upgraded in the stochastic LCI model. Monte Carlo computer simulations are obtained and used to compare the output variance of the original stochastic model with modified stochastic model. The methodology is applied to four real-world beverage delivery system LCA inventory models for verification. This research assists LCA practitioners by streamlining the conversion process when converting a deterministic LCI model to a stochastic model form. Model users and decision-makers can benefit from the reduction in output variance and the increase in ability to discriminate between product system alternatives.

Keywords: Data quality indicator (DQI); input data quality; LCI; life cycle inventory (LCI); monte carlo simulation; ranking; screening; stochastic modeling; streamlining

Introduction

Environmental life cycle inventory (LCI) models provide useful information regarding the environmental consequences associated with a number of systems under study. These models tend to be deterministic, which limits the model results by failing to provide a qualitative understanding of the uncertainty associated with the input data. Only more recently were approaches developed to build summary or data quality indicators to account for data uncertainty in a quantitative fashion. The incorporation of these techniques has lead the U.S.-EPA's Science Policy Council to approve a new agency policy on probabilistic risk assessment using Monte Carlo methods (Risk Policy Report [1]).

Converting deterministic LCI models to a stochastic form can be more informative by providing a level of confidence when assessing system alternatives, Kennedy et al. [2]. Overall, stochastic LCI modeling provides a method for dealing with issues of varying input data quality and its effect on the model's output. Uncertainty or variability in the input data transfers or propagates into the end results, thus directly affecting the reliability and comparability of system alternatives. The ability to account for data uncertainty, via data quality indicators (DQI), strengthens the inventory-modeling framework. However, the large quantity of input data and the sequential mathematical combinations of the data that is required to support this type of modeling present two challenges. First, the uncertainty associated with the large quantity of input data and their corresponding sequential mathematical combinations confound the propagation of variability in the end results. Second, the required qualitative assessments needed for each input data element's DQI make converting deterministic LCI models to a stochastic form tedious and laborious. Therefore, research efforts need to focus on developing a screening methodology to identify those input data elements that contribute the most toward the final output. If the uncertainty or data quality of an identified input data element is poor, LCI practitioners can resort to better data collection efforts to improve its uncertainty or distributional accuracy. These efforts will reduce the required number of qualitative assessments when converting deterministic LCI's to a stochastic form and reduce the model output variance.

In regards to above, recent literature has recognized the need for screening model results and streamlining the LCA methodology. Weitz et al. [3], provides a thorough discussion about the importance of, and techniques being used today, for simplifying the LCA methodology. Franklin and Hunt [4], introduces a screening methodology that identifies certain product components in the inventory model that may be eliminated. The results of the methodology suggest that the majority of the product components, those making up at least 95% of the mass of the product, need to be included in the LCI study. This methodology is helpful since it would inherently reduce some of the required data quality assessments needed when converting deterministic LCI models to a stochastic form. Heijungs [5] expresses the importance of having operational methods and criterion in the inventory phase to tell LCA practitioners where to invest in further qualitative research with high priority. The article also states that in life cycle screening it is necessary to distinguish data that is uncertain from data that contributes grossly to the final result and for which the final result is quite sensitive.

The methodology presented in this paper responds to the need for streamlining and screening LCI models to identify issues where a more detailed assessment is necessary. The rank-based approach incorporates the linear structure of the LCI modeling framework and demonstrates how improving the data quality of the identified data elements enhances model performance. The approach advances the stochastic LCA inventory modeling methodology two-fold: first by reducing the workload in the conversion stage, and second by improving the accuracy when distinguishing between system alternatives. The screening methodology for the input data elements is presented in section two. Verification of the methodology is presented section three, it also highlight the benefits of reducing the stochastic model's output variance. Conclusions, recommendations, and future research are discussed in section four.

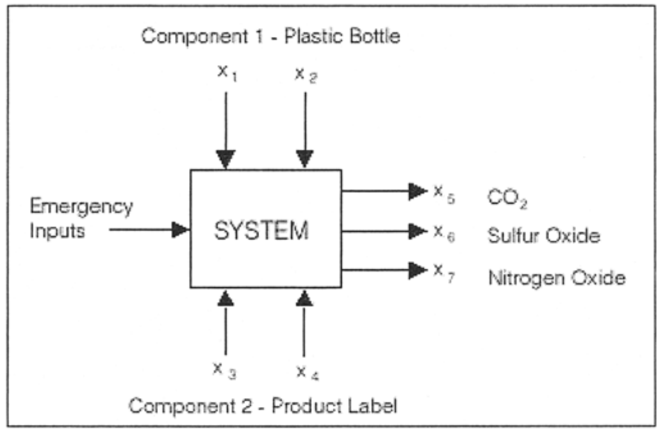

Fig. 1: Beverage product system with multiple inputs and outputs

1 Methodology

The methodology for screening the input data elements in the inventory stage is practical, general, and therefore is applicable to any LCI model. The basis of the methodology requires ranking each input data element in the deterministic LCI model. This ranking is based upon the amount each data element contributes toward the final output of an environmental burden (e.g. total energy, atmospheric emissions, etc.). The data elements along with their position in the deterministic model are then sorted into descending order based upon their individual contributions. This enables practitioners and model users to identify those data elements that contribute the most in the inventory stage. Finally, percentages of the top ranked data elements are selected, and their corresponding data quality indicator (DQI) value is upgraded in the stochastic LCI model. The modified stochastic model is then simulated and evaluated for improvements.

To better illustrate the methodology, structure, and mathematical dependencies involved with both the deterministic and stochastic LCI model, a simplified overview and example is presented. Consider one environmental burden (atmospheric emissions) for a beverage product system. Assume the product has four input materials (e.g. two that make up the plastic container component and two that make up the product label component), and has three environmental emissions (e.g., CO_2, nitrogen oxide, and sulfur oxide). Let the environmental burden be denoted by B, the input materials into the system by $\mathbf{x}_1$ to $\mathbf{x}_4$, and the output emissions by $\mathbf{x}_5$ to $\mathbf{x}_7$, see **Fig. 1**. It will be assumed that the input data estimates and the material, energy, chemical, and mass balances for this product system's inventory phase are representative and satisfy the SETAC guidelines and principles. For more information on the formulation and principles of LCI modeling readers are encouraged to reference Curran [6].

The deterministic LCI models used for this research resemble the structure of an environmental input-output model. This is often the case since LCI is an extension of the input-output form of economic analysis. This structure provides a useful framework for tracing energy use and other activities such as environmental pollution associated with a product's system, see Heijungs [7], or Miller [8] for references. The objective of the LCI model is to combine and compile the input and output data for each step of the system. The majority of the information and computations involved with a LCI model can typically be represented in vector and matrix notation:

$$A = m'E; \tag{1}$$

where E represents a matrix that makes up the 'input data' for a common specified unit, 100 kg for example, of each input material across each step in the system. That is, E represents a matrix of input data point estimates for each of the four materials and their corresponding three output emissions aggregated across the whole product life cycle. It is these estimates whose input data quality is assessed when converting from a deterministic model form to a stochastic form. The matrix **A** can be termed as a 'weighted requirement matrix', because it provides coefficients of contribution toward the burden of each input across all outputs. These coefficients are based upon the amount of weight/mass each input material requires to deliver a functional unit of product. The vector of corresponding mass estimates, weight factor **m'** , is made up of linear combinations of multiple factors found in the 'Input System Specification' (ISS). The ISS is a user specified part of the LCI model that provides estimates for factors that make up the required weight for each component of the product based on the functional unit being delivered. It is common for a product system to have alternative formulations. When this is the case the ISS adjusts the components' required weight estimates according to the alternative being modeled. **Table 1** provides a visual representation of a deterministic LCI model in spreadsheet format.

The individual cell values within the **A** matrix can be validated by using equation (1) above, that is, multiply each cell of **m'** vector by the corresponding row of **E**. The environmental burden, B, for this example system is 2.678 kg assuming 1000 units of product. This value is obtained from summing each individual column within the **A** matrix:

$$B = \Sigma A = \Sigma a_j = \Sigma \Sigma m_{ij} \varepsilon_{ij} \tag{2}$$

Table 1: Spreadsheet format for a deterministic inventory model

Input System Specification (ISS) *per 1000 units of product delivered*					
	Required Weight (kg)	*Recycling Rec. Rate*		*Req'd Wt*	*Poly Mfg*
C1 Container	*13.25*	*0.2464*	*Input x_1*	*8.25*	*1.52*
C2 Label	*7.75*		*Input x_2*	*5.00*	*0.55*
DETERMINISTIC FORM			*Input x_3*	*3.25*	*0.34*
			Input x_4	*4.50*	*0.72*

Input Data By System Steps – Matrix [E]					
			Total Pollutants		
Components		*Weight (kg)*	*CO_2 x_5*	*Nitrogen Oxide x_6*	*Sulfur Oxide x_7*
C1 Petrol - x_1		*100 kg*	***3.85***	***2.48***	***0.87***
C1 ChemA - x_2		*100 kg*	***12.2***	***8.54***	***0.54***
C2 ChemB - x_3		*100 kg*	***0.95***	***1.32***	***1.25***
C2 ChemC - x_4		*100 kg*	***1.98***	***4.58***	***2.47***

Weighted Requirements - Matrix [A] and Vector m					
			Total Pollutants		
Components	*Wt Factor **[m]***		*CO_2 x_5*	*Nitrogen Oxide x_6*	*Sulfur Oxide x_7*
C1 Petrol - x_1	***0.151***		***0.581***	***0.374***	***0.131***
C1 ChemA - x_2	***0.054***		***0.658***	***0.461***	***0.029***
C2 ChemB - x_3	***0.019***		***0.018***	***0.025***	***0.023***
C2 ChemC - x_4	***0.042***		***0.083***	***0.192***	***0.103***
		Output Subtotals a_j =	*1.340*	*1.052*	*0.286*
		Atmospheric Waste (kg) or B =	*2.678*		

where a_j is the coefficient of contribution for each atmospheric emission output, and $m_{ij}e_{ij}$ is the product of the individual cells in both **m'** and **A**, respectively. Hence, the sum of each individual column divided by the grand total output yields the proportion that each atmospheric emission contributes toward the product's final burden. The individual data elements within each column are then scaled by their corresponding column proportion value:

$$\frac{m_{ij}\varepsilon_{ij}}{(a_{\bullet}/B)} \tag{3}$$

This results in an input/output matrix of data elements that are ranked based on their individual contributions. These values are then sorted into descending order and placed into a vector containing its contribution percentage and position in the original deterministic model. The original position of the cells is needed because its corresponding position in the stochastic model will be subject to a DQI upgrade.

The ranking should be applied to the original deterministic model, even if a stochastic model already exists. This assures that the individual contributions obtained in the ranking methodology are correct, since the data elements in the stochastic model are subject to change based on the random draws each receives during a simulation. However, if only a deterministic model exists then the procedure is highlighting those input data elements whose uncertainty should be assessed with the highest quality. Therefore, as shown in the following section, assessing only the important data elements and ignoring all remaining data elements streamlines the conversion from deterministic to stochastic form.

Once the original deterministic model has been screened, LCA practitioners should focus their efforts on robust data collection methods for the identified data elements. These efforts will assist in trying to determine each input data element's corresponding distribution or assist in improving the uncertainty in its estimate value. If actual data exist, practitioners can use graphical methods such as boxplots or histograms to help characterize the distribution. However, if data doesn' t exist expert judgment in the form of DQIs can be applied to quantify the distribution of the input data, for references see Coulon [9], Weidema [10] or Kennedy [11]. Each input data element estimate in the deterministic LCI model can then be transformed into a random variable that draws from a probabilistic distribution. This in turn provides the means for stochastic modeling.

Returning to our example, all the estimates within the **ISS** and E matrices of Table 1 require obtaining a corresponding DQI value. See **Table 2** for the stochastic form for the LCI model. Notice all estimates in the **ISS** table and all input data element estimates found in the **E** matrix have their corresponding DQI assessment value shaded and shown in the upper right-hand corner of each cell. For this example, we are assuming that knowledge of the actual probability distribution and its associated parameters is not available. When this is the case the appropriate probability distribution to use is the beta distribution.

The DQI values for this research use a single rating to measure the overall quality of each data element. This rating is based upon a Likert-type [12] sliding scale of one to five, with a one representing the worst quality (maximum uncertainty), and a five representing the best quality (minimum uncertainty). These qualitative assessments are then used to parameterize the probability density function of a beta random variable x:

$$f(x;\alpha,\beta,a,b)=\left[\frac{1}{b-a}\right]\cdot\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\right]\cdot\left[\frac{x-a}{b-a}\right]^{\alpha-1}\cdot\left[\frac{b-x}{b-a}\right]^{\beta-1}$$

$$for\begin{Bmatrix} a\le x\le b \\ 0, otherwise\end{Bmatrix}; \tag{4}$$

where α and β are the distribution shape parameters, and a and b are the selectable range endpoints. This distribution is appropriate primarily because the shape parameters and range end points allow virtually any shape probability distribution to be represented. The shape parameters establish the shape of the distribution and thus the location of the probability mass, whereas the endpoints limit the range of possible values. The range end points are based on a +/- percentage of the original input data element estimate. The

Table 2: Spreadsheet format for a stochastic inventory model

Input System Requirements (ISS) per 1000 units of product delivered					
	Required Weight (kg)	Recycling Rec. Rate		Req'd Wt	Poly Mfg
C1 Container	13.25	5 0.2464	Input x_1	5 8.25	5 1.52
C2 Label	7.75		Input x_2	5 5.0	5 0.55
STOCHASTIC FORM			Input x_3	5 3.25	5 0.34
			Input x_4	5 4.50	5 0.72

Input Data By System Steps – [E] Matrix					
			Total Pollutants		
Component		Weight (kg)	CO_2 x_5	Nitrogen Oxide x_6	Sulfur Oxide x_7
C1 Petrol – x_1		100 kg	5 **3.85**	5 **2.48**	5 **0.87**
C1 Chem A – x_2		100 kg	2 **12.2**	2 **8.54**	2 **0.54**
C2 Chem B – x_3		100 kg	5 **0.95**	5 **1.32**	5 **1.25**
C2 Chem C – x_4		100 kg	5 **1.98**	5 **4.58**	5 **2.47**

Weighted Requirements – Matrix [A] and Vector m					
			Total Pollutants		
Component	Wt Factor **[m]**		CO_2 x_5	Nitrogen Oxide X_6	Sulfur Oxide X_7
C1 Petrol – x_1	**0.151**		**0.581**	**0.374**	**0.131**
C1 ChemA – x_2	**0.054**		**0.658**	**0.461**	**0.029**
C2 ChemB – x_3	**0.019**		**0.018**	**0.025**	**0.023**
C2 ChemC – x_4	**0.042**		**0.083**	**0.192**	**0.103**
		Output Subtotals =	1.340	1.052	0.286
		Atmospheric Waste (kg) =	2.678		

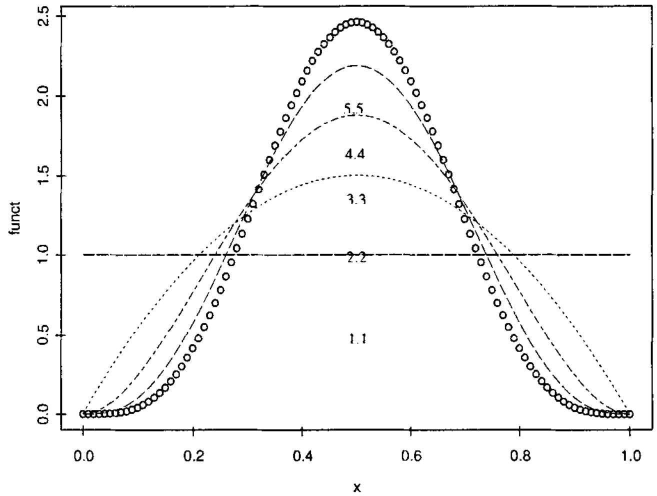

Fig. 2: Beta probability density function (pdf) for various shape parameters (a and ß) listed in Table 1

beta distribution shape parameters and range endpoints corresponding to each DQI are presented in **Table 3**. The beta probability distribution functions shown in **Fig. 2** provide an indication for how likely a particular value of x will occur, where x can be thought of as the value taken on by an input data element having range end points of 0 and 1. Referencing Table 2 and Fig. 2 respectively, we see that all the DQI values for the input materials are of highest quality except for input material x_2, therefore the random draws for material x_2's input data elements will vary about their original estimate more than the other material input data estimates. Hence, if one of the input data elements for x_2 had the highest rank or contribution, increasing its DQI value can reduce the LCI model output variation since its draws will be from a tighter symmetric distribution.

The shape parameters and range endpoints specified in Table 3 were established and used if LCA practitioners were not able to provide adequate corresponding information. Unless otherwise specified by the LCA practitioners, the variability of the input data is symmetrically distributed about its original point estimate utilizing the beta distribution shape parameters.

Table 3: DQI values and their corresponding beta distribution parameters

	Beta Distribution	
Data Quality Indicator	**Shape Parameters** (α, β)	**Range Endpoints** (+/-%)
5	5,5	10
4.5	4,4	15
4.0	3,3	20
3.5	2,2	25
3.0	1,1	30
2.5	1,1	35
2.0	1,1	40
1.5	1,1	45
1.0	1,1	50

This format is appropriate because with no additional information, it ensures an equal probability of selecting values that are the same amount above and below the given the data element value. The range endpoints are coded at +/- 10% of the input value from a DQI of five (highest quality) to +/- 50% for a DQI of one (lowest quality). These percentages were selected such that the range of random variable's values the input data element could assume continuously increase with incremental decreases in data quality.

The code that ranks all the data elements in the input/output matrix can be found in the **Appendix**. The language selected was Matlab, version 5.0, since it is a matrix manipulation-based language. However, any computer language can be used to provide the basic operations. If the LCI models have a spreadsheet format to manage all the mathematical linear combinations, then all that is required is copying the correct input/output matrix and assigning it to a corresponding variable. However, care must be taken to remove any rows in the matrix that may subtotal a stage of the resource components or any columns that may subtotal a stage of the environmental emissions. Failure to do so will result in an incorrect ranked vector highlighting certain data elements that do not exist.

2 Application

This methodology was applied to an existing stochastic LCI model developed for a beverage delivery system and presented in Kennedy [1]. The model computes the total energy in Gigajoules (GJ) required to deliver 10,000 liters of beverage. Four system alternatives were initially studied to validate and verify any reduction in their output variance. The choice of the total energy model was arbitrary. The methodology could have easily been applied to a variety of other existing environmental burdens such as waterborne waste or atmospheric emissions.

Prior to verifying any variance reduction, the output from the original deterministic and stochastic (baseline) LCI model presented in Kennedy [1] was compared to the output from the modified stochastic model having five-percent of the top ranked data elements receiving a DQI upgrade. The choice of five-percent was selected just to test and validate the methodology. The deterministic model's output point estimate and the stochastic model's (baseline and modified) output mean values for each system alternative were computed and reported in **Table 4**. The mean values are based upon fifty independent Monte Carlo simulation runs. The mean output associated with each system alternative and each LCI model type highlights that the methodology has not altered or shifted the output drastically from its original deterministic and stochastic form.

To measure the amount of variance reduction from the methodology, an incremental selection of one-percent was selected. That is, one-percent increments of the top ranked data elements in the LCI model are selected to receive an upgrade in data quality. A small percentage increment was selected in order to highlight the behavior of the DQI upgrades and any resulting variance reduction. Based on the number of input data elements in the LCI model, this resulted in a cumulative sequence of twenty-seven cells receiving the maximum upgrade (DQI value of 5) at each incremental simulation run. The incremental sequence of upgrades continues until all the active cells in the model have been updated. The total number of active cells for each system alternative is highlighted in the last column of Table 4. Therefore, one-percent increments results in fourteen total incremental simulation runs, no more simulations were needed since the remainder of input data elements have a zero value – non active cells. Each simulation run obtained fifty independent output values for every one-percent increment, a sufficient sample size for acquiring a variance estimate and performing graphical analysis. These output values were then compared to a single output variance estimate - based on five hundred independent simulation runs - from the original corresponding stochastic model. The whole process was repeated for all four-beverage system alternatives. Once the output data was obtained, the results were analyzed using basic summary statistics and plots. The variances from the summary statistics for each modified model at each one-percent increment were used to obtain the sequence of plots shown in **Fig. 3**. These estimates were compared to the corresponding baseline output variance estimate to determine the percentage amount of variance reduction:

Table 4: Output from each system alternative for both model types

	Deterministic Model Point Estimate	Baseline Stochastic Mean	Modified Stochastic Mean	Number of Active Cells
System Alternative 1	10.7	10.93	11.2	391
System Alternative 2	28.2	29.55	30.75	386
System Alternative 3	30.2	33.48	33.15	386
System Alternative 4	25.2	27.34	27.17	374

$$\%Var\_Red. = \frac{(\sigma^2_{Baseline} - \sigma^2_{Modified})}{\sigma^2_{Baseline}}.$$

Fig. 3 highlights how much the output variance is reduced as the percentage of data elements selected for upgrade increases. The remainder of data elements not selected for an upgrade remained at their existing DQI value. At two percent, three out of the four system alternatives obtained a variance reduction greater than twenty-five percent. Alternative one shows the highest gain in variance reduction out of all four-system alternatives, at one-percent a forty-eight percent reduction was obtained. Reductions of this amount will increase practitioners' and model users' ability to statistically discriminate between resembling system alternatives. Overall, each alternative obtained an average variance reduction of 51.66, 28.46, 27.05, and 27.45 percent, respectively.

The benefit of applying this methodology to existing stochastic models is that LCA practitioners will know where to focus their resources to improve the uncertainty associated with those

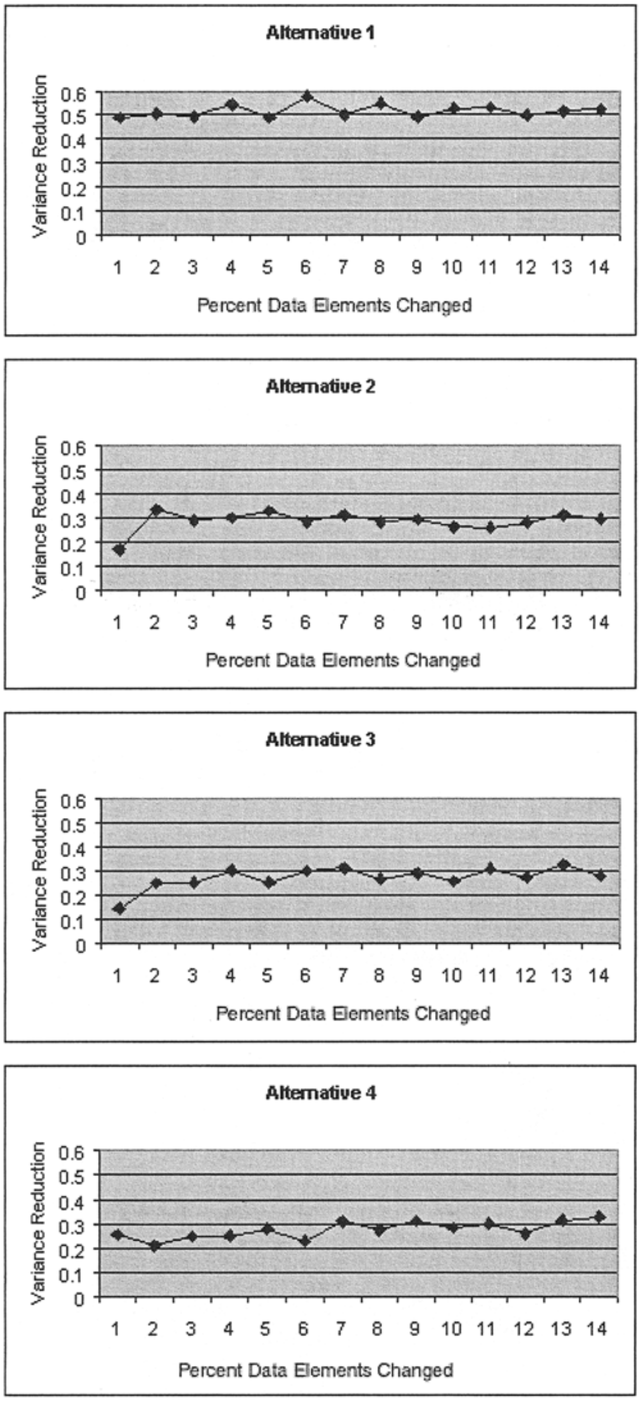

Fig. 3: Variance reduction plots for each system alternative when methodology applied to existing stochastic models

data elements deemed important. Based upon Fig. 3, improving two percent of the data elements can substantially improve the output variance. Two percent in this case requires that fifty-four data elements need to have their uncertainty be re-evaluated. Even if this amount of re-evaluation is too costly or time consuming, it is highly likely that fifty-four re-evaluations will not be needed. Generally, some of these data elements will already be at their highest index value. Data elements that have marginally high DQI values should be left

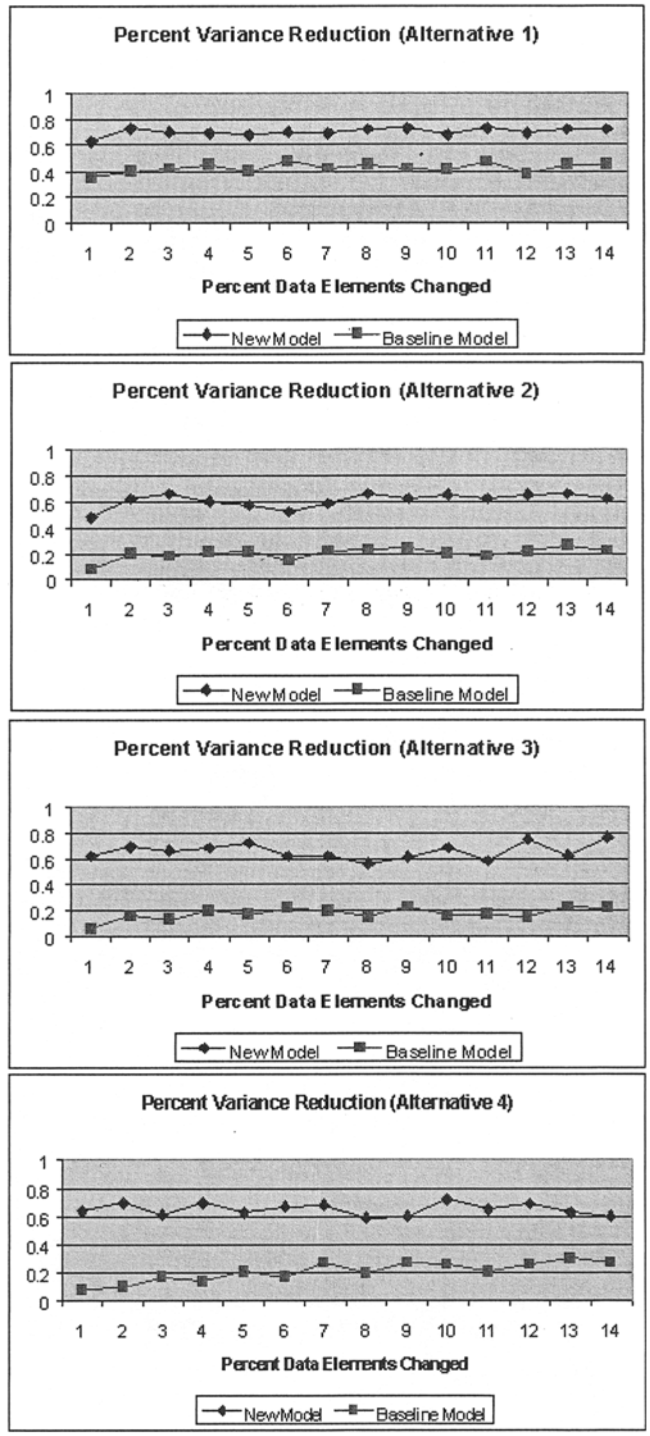

Fig. 4: Variance reduction plots for each system alternative when the non-selected data elements were set at worse case scenario

alone, since the time and energy to increase them from a four to a five, for example, will not be that beneficial. Simply re-evaluating the data elements with lower DQIs will be sufficient enough to obtain a reduction in the output variance.

The methodology's ability to reduce the workload during the conversion stage was investigated next. This required setting all the remaining data elements not selected for an initial assessment (upgrade) to a worse case scenario - DQI values of one. Therefore, the only DQI assessments required are for the percentage of data element deemed as important, all remaining data elements are ignored and set at maximum uncertainty (worse case). In doing so, we assume that the uncertainty for each of the selected data elements will be capable of receiving the highest assessment value – DQI value of five. Once again a one-percent selection criterion was used and the same simulation process was repeated. The output variances were analyzed and first compared to the output variance estimates used in the previous simulations for the original stochastic baseline models. The second comparison was with the new converted stochastic model, which was run with each data element set at worse case scenario. The results for each system alternative are presented in **Fig. 4**.

The plots in Fig. 4 indicate that practitioners need only focus on a certain percentage of the data elements identified in the screening process. This is based upon the variance reductions obtained when compared to the existing stochastic or baseline model output. Once again, by two-percent, each system alternative experiences some amount of variance reduction. Therefore practitioners could have assessed just two-percent of the data elements, instead of all of them, to obtain model results with less variability. These results suggest that the methodology presented within is capable of streamlining the conversion process of going from a deterministic to stochastic form. The second comparison with the new converted model highlights the overall gain in variance reduction that can be achieved prior to any DQI assessments. Even though large gains are expected, the plots emphasize that DQI upgrades beyond five-percent do not increase the gain in variance reduction significantly.

It should be noted that this report does not imply or suggest that selecting two or five-percent for assessment be the rule-of-thumb since results will vary from model to model. Practitioners will need to exhibit great care when selecting the initial percentage of data elements for assessment. Often plotting the amount each data element contributes toward the total against its corresponding ranking order will help in making an educated selection. **Fig. 5** illustrates this helpful plot for system alternative 2. The plot highlights that two-percent would be sufficient for re-assessment since the contribution of the data elements beyond 54 is minimal. This corresponds

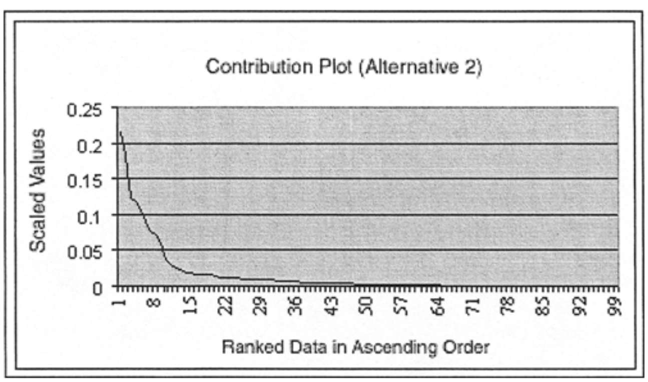

Fig. 5: Data element contribution versus ranking order

with the results shown in Fig. 3, that is, the maximum gain in variance reduction was achieved at two-percent.

Lastly, the methodology was tested to see if the stochastic LCI model's reduction in output variance would improve the ability to discriminate between system alternatives.

To more adequately show the benefits of the variance reduction an additional beverage system alternative was created. The fifth alternative created is a feasible derivative of alternative four, with mean and variance characteristics similar to alternative two.

Performing an F-test for assessing the equality of means is the first step when making multiple comparisons of alternatives. Analysis of variance (ANOVA) makes use of such a test. An ANOVA determines if there exists a statistically significant difference between the means of the system alternatives. A single factor ANOVA was computed on two separate data sets using $\alpha = 0.05$ to test for equality of means between the k treatments (i.e., $k = 5$ system alternatives). The two separate data sets are comprised of running fifty independent runs for each system alternative from the existing (baseline) stochastic model, used in Kennedy [1], and the modified stochastic model having the top two-percent ranked data elements receiving DQIs of five. It should be noted that fifty new observations for the new alternative five were required and simulated - utilizing the two-percent upgrade. Next a 95% Tukey's comparison of means test was performed using $\alpha = 0.05$ to test for statistical significant differences between the mean output values of each possible pair of system alternatives. This test is appropriate to further determine which pairs of alternative means are significantly different if the results from the ANOVA and F-test indicate a significant difference between means exists. Tukey's test was selected because it is very conservative and exact when the sample size is equal for all alternative means under comparison. For detailed information on the analytical methods mentioned above readers should reference Montgomery [13].

Table 5 contains output from SAS, a statistical software package, highlighting the ANOVA table and 95% Tukey's confidence intervals for all pairwise differences between means. Based upon the F-test and small P-values for equality of means in the ANOVA table, there is strong evidence that at least one system alternative has a mean value significantly different than the others. Referencing the Tukey confidence intervals in Table 5 one can determine where the significant pairwise differences exist. If the confidence interval contains zero or no stars are present in the last column, then no significant differences exist between the pair of alternatives. Observe the pairwise confidence intervals for alternatives two and five. The interval contains zero in the baseline analysis, whereas the modified analysis does not contain zero. This demonstrates that the reduction in variance improves the ability to discern whether a significant difference in means exists. This information is very beneficial for model users and decision-makers, because they can be assured when discriminating between candidate system alternatives with similar characteristics.

Table 5: ANOVA and 95% Tukey confidence intervals for both baseline and modified models

Analysis of Variance Table

Modified model:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	15666.49	3916.62	7041.94	0.0001
Error	245	136.26	0.5561		
Total	249	15802.75			

Baseline model:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	15915.78	3978.94	4926.50	0.0001
Error	245	197.87	0.80766		
Total	249	16113.65			

Tukey Comparison of Means Test

Analysis for baseline model:

Alternative Comparison	Lower Confidence Limit	Difference Between Means	Upper Confidence Limit
1 – 2	–20.0816	–19.5877	–19.0937 ***
1 – 3	–22.6918	–22.1979	–21.7039 ***
1 – 4	–16.3788	–15.8848	–15.3909 ***
1 – 5	–20.0391	–19.5452	–19.0512 ***
2 – 3	–3.1042	–2.6102	–2.1163 ***
2 – 4	3.2089	3.7028	4.1968 ***
2 – 5	–0.4515	0.0425	0.5364
3 – 4	5.8191	6.3130	6.8070 ***
3 – 5	2.1587	2.6527	3.1467 ***
4 – 5	–4.1543	–3.6603	–3.1664 ***

Analysis for Modified Model:

Alternative Comparison	Lower Confidence Limit	Difference Between Means	Upper Confidence Limit
1 – 2	–19.9968	–19.5869	–19.1770 ***
1 – 3	–22.6446	–22.2347	–21.8248 ***
1 – 4	–16.3814	–15.9715	–15.5616 ***
1 – 5	–19.2205	–18.8106	–18.4007 ***
2 – 3	–3.0577	–2.6478	–2.2379 ***
2 – 4	3.2055	3.6154	4.0253 ***
2 – 5	0.3664	0.7763	1.1862 ***
3 – 4	5.8533	6.2632	6.6731 ***
3 – 5	3.0142	3.4241	3.8340 ***
4 – 5	–3.2490	–2.8391	–2.4292 ***

3 Conclusions

The methodology presented within has proven to be effective at streamlining and improving the stochastic LCI modeling process. It is general and applicable to any environmental inventory model. The application makes the conversion process of going from a deterministic to a stochastic model form more practicable. Lastly, the methodology improves the output variance of existing stochastic models by showing practitioners where energy should be focused to improve the data uncertainty for those important input data elements.

Even if converting a deterministic model is not feasible, the methodology can still be used on existing LCI models to identify the important data elements. Once identified, measures can be taken to assure that proper attention is always allocated to their input data point estimates. However, if the deterministic LCI model's total output is due to a small number of input data elements, the conversion to a stochastic form should be considered. The amount of work in the conversion is negligible compared to the benefits and information gained from stochastic LCI modeling. The ability to statistically discriminate between system alternatives gives LCA practitioners and decision-makers an edge in policy making.

The results from the application of the methodology illustrated that a reduction in the output variance is possible after only a small percentage of input data elements received a DQI upgrade. The plot of the data element contributions versus the ranking order also illustrated that small percentages of the data elements generally contribute the most toward the total inventory output. This phenomenon suggests that LCI input data elements may tend to follow a Pareto principle. In other words, a small proportion of the variables, input data elements in this case, make up the majority of the variability. When this characteristic is prevalent, the application of a screening experiment or methodology proves to be beneficial.

The development of this research has identified the necessity of future research. First, work involving the optimization of the ranked vector would prove to be beneficial. Create an algorithm-based screening methodology capable of evaluating the combinatorial explosion of possible DQI upgrades for each input data element in the stochastic model. The algorithm should select those data elements constrained on the amount each contributes toward the total output, allocation, costs, and the value of its current DQI value. The results should highlight lower ranked data elements that may have been overlooked utilizing the methodology presented within. Second, efforts to improve the quality of the existing parameterization format for the beta distribution. That is, research should focus on robust parameter representation with the focus of developing a standard. Advances such as these will make stochastic modeling itself a standard when modeling the inventory stage.

References

[1] Risk Policy Report – February 21, 1997, 3 (2) p 6

[2] Kennedy DJ, Montgomery DC, Quay BH (1996): A Probabilistic Approach to Incorporating Variable Input Data Quality. Int J LCA 1 (4) 299–207

[3] Weitz KA, Todd JA, Curran MA, Malkin MJ (1996): Streamlining LCA: Considerations and a Report On the State of Practice. Int J LCA 1 (2) 79–85

[4] Franklin WE, Hunt RG (1998): Can a Universal Screening Methodology be Devised for LCAs Complex Products. SAE Int Congress & Exposition. February 23–26, p 183–189

[5] Heijungs R (1996): Identification of Key Issues for Further Investigation in Improving the Reliability of Life-Cycle Assessments. Journal of Cleaner Production 4 (3–4) 159–166

[6] Curran M (1996): Environmental Life-Cycle Assessment. McGraw-Hill, NY, USA

[7] Miller B, Blair P (1985): Input-Output Analysis. Prentice-Hall, New Jersey, USA

[8] Coulon R, Camobreco V; Teulon H, Besnainou J (1997): Int J LCA 2 (3) 178–182

[9] Weidema BP, Wesnoes MS (1995): Data Quality Management for Life Cycle Inventories – An example of Using Data Quality Indicators. Presented at the 1995 SETAC World Congress

[10] Kennedy DJ, Montgomery D, Rollier DA, Keats JB (1997): Assessing Input Data Uncertainty in Life Cycle Assessment Inventory Models. Int J LCA 2 (4) 229–39

[11] Likert RA (1932): Technique for the Measurement of Attitudes. Archives of Psychology 60, p 140

[12] Montgomery DC (1997): Design and Analysis of Experiments. 4th ed. John Wiley & Sons, New York, USA

Received: May 2nd, 2000
Accepted: August 3rd, 2001
OnlineFirst: August 31st, 2001

Appendix: Matlab code

```
%    Reads in Matrix A
function manip=rank_matrix(A)
pct=.1
[m,n]=size(A);
te_col=sum(A);
te=sum(te_col);
col_pct=te_col/te;
%
for i=1:n
  B(:,i)=(A(:,i)*col_pct(i));
end;
%
%    Puts ranked vector into ascending order!!!
%[a,b]=size(B);
%B_col=reshape(B,m*n,1);
%[y,p]=sort(B_col);
%rv=[y,p];
%
%    Puts ranked vector into descending order!!!
neg_B=(-1)*B;
[a,b]=size(neg_B);
B_col=reshape(neg_B,m*n,1);
[y,p]=sort(B_col);
p=(-1)*p;
rv=(-1)*[y,p];
%
%    Returning the specified percent of the ranked vec-
tor!!!
[r,s]=size(rv);
pct_rv=rv([1:pct*r],:);
manip=pct_rv(:,2)
%
%    Storing output to a file!!
%diary store
%diary off
%
%===-====================================
```