AIChE Wiley InterScience DISCOVER SOMETHING GREAT

Rectification of Multiscale Data with Application to Life Cycle Inventories

Heui-Seok Yi and Bhavik R. Bakshi
Dept. of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210

DOI 10.1002/aic.11119
Published online February 26, 2007 in Wiley InterScience (www.interscience.wiley.com).

The quality of life cycle inventory (LCI) data is crucial to the reliability of decisions made via life cycle analysis (LCA). However, many LCI data, be they from commercial software or public domain databases, violate the laws of thermodynamics due to errors, missing data, and other inconsistencies. The process engineering method of data rectification is appealing for improving the quality of LCI data, but applying this approach poses many new challenges that are not addressed in traditional rectification. One such challenge is due to the availability of life cycle data at multiple overlapping scales. This includes engineering data at the equipment and process scales, process LCA data at the value chain scale, and economic input–output and toxic release inventory data at the economy scale. This article develops a method for rectification of such multiscale data. Rectification is accomplished via a mixed integer optimization-based approach, and it is integrated with the existing computational methods for different types of traditional and hybrid LCA. The developed method may also be used for multiscale data from process engineering. It is applied to data for a caustic soda process from a public domain LCI database to illustrate the benefits of the proposed approach and identify further challenges. © 2007 American Institute of Chemical Engineers *AIChE J,* 53: 876–890, 2007

Keywords: life cycle assessment, data rectification, multiscale rectification, sustainable engineering

Introduction

Life cycle assessment (LCA) is emerging as an essential tool for evaluating the broader environmental aspects of engineering products and processes.[1,2] Such analysis is increasingly common in many corporations and research laboratories for strategic decision-making, based on the sustainability potential of alternate decisions and engineering activities. LCA aims to account for inputs and outputs of processes in the supply and demand chains (value chain) of a selected product or process, which requires extensive knowledge about the inputs of raw materials and energy, and outputs including products, byproducts, and emissions of several industrial processes and economic activities. This makes development of the life cycle inventory (LCI) an extremely important but labor intensive component of a high quality and reliable LCA. Many LCIs have been compiled by governments and private entities,[3] and are available for free[4,5] or as part of software packages.[6,7] These inventories are usually compiled from diverse sources such as industrial data and public domain inventories. Some companies may also wish to use their own private process inventories for LCA study.

Since LCI data are the result of combining a variety of sources, they usually contain errors due to effects of aggregation, disparity in data from different sources, missing values, and random and gross errors in measurements.[8,9] It is common for such data to violate the conservation laws of thermodynamics.[10,11] Nevertheless, developers and some users of LCI data may argue that the emissions data, which are the main focus of LCA, are accurate enough making satisfaction of conservation laws unimportant for most users. While this argument may be reasonable in the narrow domain of a

Correspondence concerning this article should be addressed to B. R. Bakshi at bakshi.2@osu.edu.

traditional LCA, LCI data are increasingly combined with data at different scales and are used by other methods such as material, energy and exergy analysis. Consequently, the quality of input data is important for these methods, and satisfaction of conservation may improve overall data quality. Furthermore, most commercial LCA packages are starting to quantify the extent of errors in emission data. If the emission data are accurate, the proposed rectification approach will have less effect on these data and will only improve data with more errors or missing data. Without rectification, LCA based on inconsistent LCI data may result in misleading or incorrect decisions, which reduce the reliability of LCA and user confidence in this method.

The problems of inconsistent and error-prone data are commonly encountered in process systems engineering, and have been the focus of research for many decades. Methods for data rectification have been developed to improve data quality and consistency by rectifying them with the conservation laws of thermodynamics, and using statistical and optimization techniques for detection and compensation of gross errors.[12,13] The data reconciliation problem assumes that no gross error exists in measurements. In the absence of inequality and nonlinear constraints, it is usually formulated as a quadratic programming problem. The detection and identification methods of gross errors are usually derived from statistical significance tests. Identified gross errors are removed by compensation methods that utilize a nonlinear program. A mixed integer program can also be used to estimate gross error sizes and reconciled values simultaneously. The methods of gross error compensation can be categorized into serial elimination, serial compensation, and simultaneous compensation. The serial elimination method[14,15] identifies one gross error at a time by using a significance test, and considers a measured variable where the gross error is identified as an unmeasured variable. These procedures continue until no gross error is detected. The serial compensation method[16] identifies a gross error at a time, and estimates the size of the gross error until no gross error is detected. The simultaneous compensation method[17–20] estimates the sizes of all gross errors simultaneously for keeping a degree of redundancy. The degree of redundancy is defined as the number of constraints that are composed of only measured variables, which affect the accuracy of rectified values. The serial elimination method considers a measured variable that is identified as containing a gross error as an unmeasured variable so that the degree of redundancy decreases with iterations. In contrast, the simultaneous compensation method does not eliminate measured variables, and the degree of redundancy does not decrease.

Data rectification can be accomplished only if an objective function and constraints are formulated with measured variables. Since measuring instruments are usually installed only on some streams, unmeasured variables, if present in the rectification problem, must be removed. The methods of matrix projection[21] and QR factorization[22] have been developed to accomplish this task. Observable unmeasured variables can be estimated from the measured values and constraints after rectification. A complete survey of these topics is available elsewhere.[23]

Existing methods for data rectification usually focus on measurement data at only one scale, which are directly measured by flow meters at the equipment scale. Such an approach has already been used for enhancing the quality of LCI data at a single scale.[10] However, LCI data are usually gathered from diverse sources such as plant operation and inventory databases, aggregate process LCA databases that represent average data for typical processes in a selected region, and data about economic sectors that aggregate all processes in a sector. These data represent different levels of spatial resolution and aggregation. The finest scale data at the *equipment scale* are detailed and relatively accurate, but maintaining such a level of detail for all processes in the life cycle is computationally intractable. Process LCA, standardized in ISO 14000, assumes that the amounts of resource consumption and waste emission have a linear relationship with the amount of a final product, and constructs the life cycle network of processes to calculate the total amounts of resource consumption and waste emission.[24] LCI used in process LCA is usually at the level of the *value chain*, but relies on a finite boundary, since it is practically impossible to collect information about all the interconnected processes in the life cycle. Thus, process LCA suffers from errors due to incompleteness of the system boundary. Economic input–output LCA (EIOLCA) also assumes that the amounts of resource consumption and waste emission have a linear relationship with the household demand of each economic sector, and utilizes the aggregated information of entire economic sectors.[25] These data at the *economy scale* are highly aggregated, but also the most complete in capturing the life cycle network albeit at a coarse scale. Hybrid LCA combines process LCA and EIOLCA to overcome the use of an incomplete system boundary in process LCA and the aggregation error in EIOLCA.[26] Most LCI data are prone to have errors, and do not satisfy the conservation laws of thermodynamics. Consequently, data rectification may be applied to LCI data for enhancing their quality.[11] Data rectification integrated with hybrid LCA fuses LCI data of process LCA and EIOLCA, while ensuring satisfaction of conservation laws. Because of the data being at multiple scales, this is a *multiscale* data rectification problem.

The multiscale nature of data described in the previous paragraph may also be encountered in a traditional chemical process.[27,28] For example, measurement data for high-level tasks such as planning and scheduling are at a coarse scale, while those for low-level tasks such as supervisory and regulatory controls are at a fine scale. The models and data at the coarse scale may be inconsistent with those at the fine scale. Consequently, the methods developed in this article can also be useful for rectification of traditional process data.

Multiscale methods do exist for data rectification, but are inadequate for solving the LCI rectification problem. Existing multiscale rectification methods use a Bayesian approach and wavelet basis functions for dynamic and steady state rectification.[29–31] Noisy data are decomposed into measurement sets at multiple scales, and Bayesian estimation is applied to calculate noise-free wavelet coefficients. However, these multiscale rectification methods decompose single-scale data into multiple scales for rectification or can combine data at dyadic scales when sufficient measurements or stochastic models are available.[32] The approach proposed in this article is inspired by existing multiscale methods, but significant new challenges exist in LCI rectification, which prohibit direct use of existing multiscale rectification methods. These challenges include the nested overlapping scales, presence of constraints at each scale and across scales, need to detect and compensate gross errors, few measurements and stochastic models.

The main contributions of this article include a rectification method that combines data at multiple spatial scales while satisfying the conservation laws of thermodynamics at and across coarse and fine scales, and its application to LCI data using an integrated formulation of data rectification with existing computational methods of LCA. Specifically, the rectification method is extended to combine process LCI data and economic LCI data, thus it is integrated with the computational structure of process LCA, EIOLCA, and hybrid LCA. The developed method is applied to the life cycle of a caustic soda process for demonstrating the benefits and challenges of the developed approach. The proposed approach is expected to be an important step toward providing much needed scientific rigor in LCA, and increasing the reliability and credibility of its results.

The rest of this article is organized as follows. In the next section, the multiscale characteristic of LCI data and existing methods for combining such data are explained in more detail. This is followed by a description of the developed rectification method integrated with the existing computational methods of hybrid LCA. Finally, the proposed method is applied to the life cycle of a simulated set of data and to LCI data for a caustic soda process being compiled by the National Renewable Energy Laboratory.

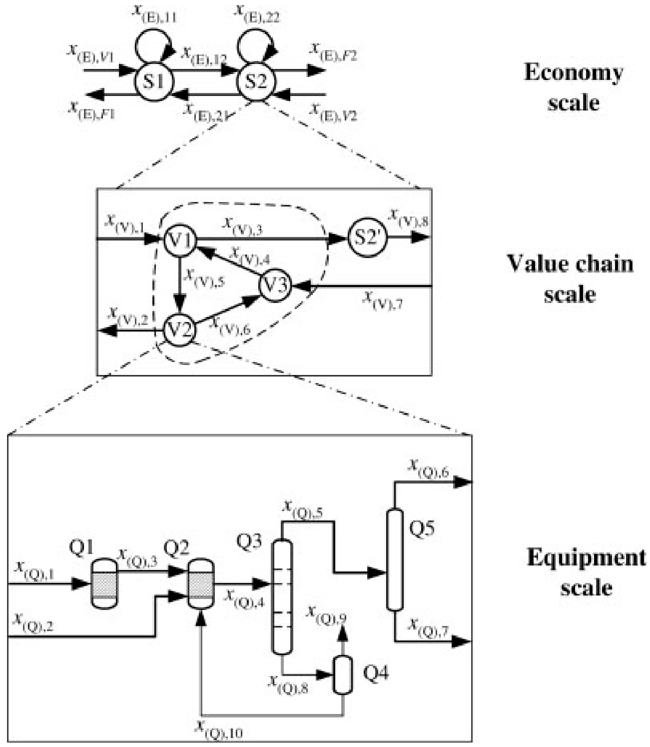


Figure 1. Conceptual diagram showing the characteristics of multiscale data.

Methods for Combining Multiscale Data

Problem formulation

Figure 1 is a conceptual diagram to show a typical multiscale system addressed in this article. This system is described in the context of LCA, but the proposed approach may be used for any multiscale rectification problem. The entire economy can be classified into two sectors, S1 and S2. The sectors have inputs, outputs, intersectoral, and intrasectoral flows, and it is assumed that these flows are measured in the units of physical quantities such as mass and energy. Usually, flows at the economy scale are measured in monetary units, but monetary flows can be transformed into physical flows when prices of products are available. Assume that partial information about the infrastructure of the economic sector S2, such as V1, V2, and V3, is available. The remaining part of S2 except V1, V2, and V3 is represented as S2′. V1, V2, and V3 have economic activity in S2, and can be processes such as chemical plants, iron-making plants, power plants, etc. It is assumed that every flow at the value chain scale is measured, and these flows can be measured in the units of money, mass, or energy. It is further assumed that V2 is a process plant, and its detail flow diagram is available as shown in Figure 1. V2 consists of the equipments, Q1 through Q5. Flow rates between equipments are considered to be measured at this equipment scale.

Multiscale measurements described in the previous paragraph can be interpreted in the context of LCA. Assume that LCA is implemented for the plant of V2. The detailed flowsheet information of V2 is available, and flow rates of its intermediate flows, products, and raw materials are measured at the equipment scale. The flow rates at this scale are usually measured by flow meters installed in production plants. The flow rates of raw materials such as $x_{(Q),1}$ and $x_{(Q),2}$ are used as the amount of resource consumption in V2 at the equipment scale for LCA. The resource consumption of V2 is also available at the value chain scale, which is aggregated as a single variable, $x_{(V),5}$. The flow rates of products such as $x_{(Q),6}$ and $x_{(Q),7}$ are measured at the equipment scale, as well as the value chain scale, which are $x_{(V),2}$ and $x_{(V),6}$. Measured values of $x_{(Q),i}$ ($i = 1,2,\ldots,10$) at the equipment scale are measured by flow meters, and can be domain-specific LCI data or LCI data at equipment scale. Available supply and demand chains of V2 contain V1 and V3, thus the system boundary of process LCA is represented as the dotted lines in Figure 1. Measured values of $x_{(V),i}$ ($i = 1,2,\ldots,7$) at the value chain scale may be collected from process inventories or public-domain databases. These represent LCI data at the value chain scale. Variables that are measured at the value chain scale may also be available at the economy scale. Thus, the measured variables of $x_{(V),1}$, $x_{(V),2}$, $x_{(V),7}$, and $x_{(V),8}$ at the value chain scale correspond to the measured variables of $x_{(E),12}$, $x_{(E),21}$, $x_{(E),V2}$, and $x_{(E),F2}$, respectively. The sum of $x_{(V),3}$, $x_{(V),4}$, $x_{(V),5}$, and $x_{(V),6}$ is equal to $x_{(E),22}$.

Such multiscale measurements are also found in the process industries. In production plants, flow rate and temperature are measured for operation and performance evaluation such as control, optimization, production planning, and yield accounting. Control and optimization are usually considered as low-level tasks, and utilize measured data from flow meters and thermocouples for entire streams. Planning and yield accounting are considered as high-level tasks that are concerned with the amounts of raw materials and products.[33,34] The high-level tasks can utilize measured data from flow meters and thermocouples at an equipment scale, but it is also possible to use the amounts of raw material purchased and products sold at a plant site scale. The two measurements at the equipment and plant site scales may not be equal to each other, and it is important to use common data for high and low-level tasks. For example, a production plant consumes raw materials, and these amounts are measured at the plant site scale, which can be the amounts of shipment and the level changes of storage tanks. The raw materials are supplied to reactors, where flow meters measure the flow rates of raw materials that are fed to the reactors. In Figure 1, the raw materials that are fed to the reactors are measured by flow meters, which gives $x_{(Q),1}$ and $x_{(Q),2}$. The amount of the raw material is also measured by the shipping amount or the level change of a storage tank, which gives $x_{(V),5}$. In practice, the sum of measured values ($x_{(Q),1} + x_{(Q),2}$) is likely to be different from the shipment amount or the level change of a storage tank ($x_{(V),5}$) due to measurement noise and errors. The raw materials react to form products that are sold to final users. The production amounts can be determined from the quantity shipped or the amounts in silos ($x_{(V),2}$ or $x_{(V),6}$), and can be measured by flow meters ($x_{(Q),6}$ and $x_{(Q),7}$). It is also possible that the measured values of production amounts at one scale are different from those at other scales. Consequently, it becomes necessary to determine meaningful values of the measurements in a statistically rigorous manner. This can be achieved via data rectification.

Existing LCA methods

Figure 2 shows a network diagram depicting process LCA, EIOLCA, and hybrid LCA. The process V3 is considered to be the target process for LCA study, and the upstream processes of process LCA are V1 and V2. The economy is assumed to consist of two sectors, S1 and S2. A raw material to V1 is supplied from S1. The subscripts (E) and (V) represent the scales of economy and value chain, respectively.

Formulation of Process LCA. Process LCA utilizes the scaled flow rates of several LCI data because it assumes that the amounts of resource consumption and waste emission have a linear relationship with the amount of a final product.[24] It usually utilizes flow information at the value chain scale. In Figure 2, the output or final demand for process LCA is the flow of $x_{(V),4}$, and each process must be scaled to satisfy the reference flow, $\phi_{(V)}$, which is a user-specified value of the final demand or output of $x_{(V),4}$. The calculation of process LCA is based on a *process matrix*, $\mathbf{P}$, that represents the relationship between the input and output flows. The process matrix of a life cycle composed of plants V1, V2, and V3 is represented in Eq. 1,

$$\mathbf{P} = [\mathbf{p}_1 \ \ \mathbf{p}_2 \ \ \mathbf{p}_3] = \left[\begin{array}{ccc} x_{(V),2} & -x'_{(V),2} & 0 \\ 0 & x_{(V),3} & -x'_{(V),3} \\ 0 & 0 & x_{(V),4} \\ \hline -x'_{(V),1} & 0 & 0 \\ 0 & 0 & 0 \\ \hline x_{(V),5} & 0 & 0 \\ 0 & x_{(V),6} & 0 \\ 0 & 0 & x_{(V),7} \end{array}\right] = \begin{bmatrix} \mathbf{T}_{(V)} \\ \mathbf{C}^u_{(V)} \\ \mathbf{B}_{(V)} \end{bmatrix}, \quad (1)$$

where $\mathbf{p}_i$ is a *process vector* of the *i*-th plant, negative sign represents input flow, and positive sign represents output flow. The element P_{ij} of the process matrix represents the *i*-th flow of the *j*-th process. Equation 1 also shows that the process matrix can be decomposed into a *technology matrix*, $\mathbf{T}$, an *upstream cutoff matrix*, $\mathbf{C^u}$, and an *intervention matrix*, $\mathbf{B}$. The upstream cutoff matrix represents that upstream processes interconnected with such streams are not evaluated in process LCA. The technology matrix shows technological input and output flows, and the intervention matrix represents the flows of raw materials and wastes. A *final demand vector* is defined using the reference flows shown in Eq. 2.

$$\mathbf{f}_{(V)} = [0 \ \ 0 \ \ \phi_{(V)}]^{\mathrm{T}} \quad (2)$$

An *inventory vector* for process LCA, $\mathbf{g}_{(V)}$, which represents the amounts of resource consumption and waste emission for the reference flow, is calculated by Eq. 3.

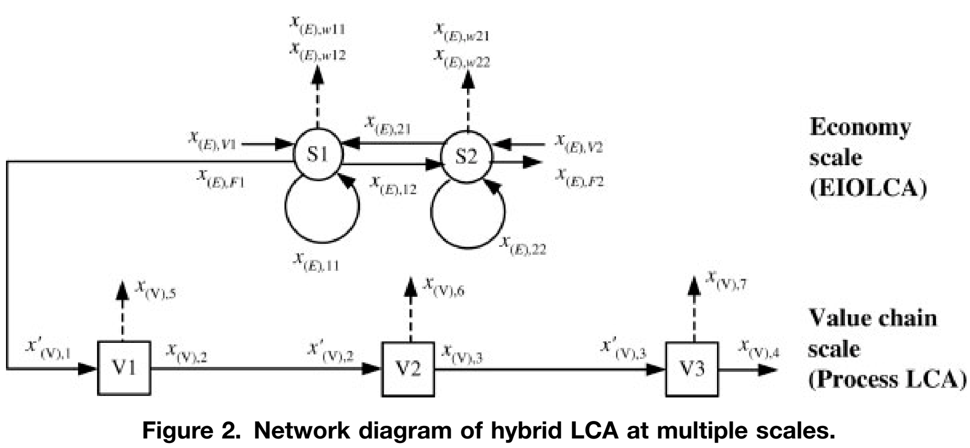


Figure 2. Network diagram of hybrid LCA at multiple scales.

$$\mathbf{g}_{(V)} = \mathbf{B}_{(V)}\mathbf{s}_{(V)} \tag{3}$$

where a *scaling vector* for process LCA using the information of the value chain scale, $\mathbf{s}_{(V)}$, is calculated by Eq. 4.

$$\mathbf{s}_{(V)} = \mathbf{T}_{(V)}^{-1}\mathbf{f}_{(V)} \tag{4}$$

Because the elements of the final demand vector except the reference flow must be equal to zero, the products of the vector of scaling factors and the technology matrix satisfy the conservation law of mass between processes Thus, $x_{(V),2}s_{(V)1} - x'_{(V),2}s_{(V)2} = 0$, $x_{(V),3}s_{(V)2} - x'_{(V),3}s_{(V)3} = 0$, and $x_{(V),4}s_{(V)3} - \phi_{(V)} = 0$. However, the method of process LCA does *not* guarantee satisfaction of the conservation law of mass around the system because $x'_{(V),1} - x_{(V),2} - x_{(V),5}$ need not be equal to zero. This computational method of process LCA can also be applied to the measurement data at the equipment scale using only the streams of resource and product when they are available.

Formulation of Economic Input–Output LCA. EIOLCA[25] is based on an economic input–output table that is composed of household demands, values added and transactions for economic sectors, and it encompasses the entire economy. Thus, it always uses the boundary of the entire economy. However, EIOLCA aggregates economic and industrial activities into broad economic sectors, and assumes that the amounts of waste emission and resource consumption for each economic sector are proportional to the household demands. Thus, a scaled vector of total outputs is utilized in EIOLCA to calculate an inventory vector. From Figure 2, the total outputs of each economic sector are defined by Eq. 5.

$$\begin{bmatrix} x_{\text{total (E)},1} \\ x_{\text{total (E)},2} \end{bmatrix} = \begin{bmatrix} x_{(E),11} + x_{(E),12} + x_{(E),F1} \\ x_{(E),21} + x_{(E),22} + x_{(E),F2} \end{bmatrix} \tag{5}$$

where, $x_{\text{total(E)},i}$ is the total output of the i-th economic sector, $x_{(E),ij}$ is the transaction from the i-th economic sector to the j-th economic sector, and $x_{(E),Fi}$ is the household demand of the i-th economic sector. Usually, the transactions and household demands are measured in a monetary unit such as dollars, but they can also be represented in physical units. However, the formulation of EIOLCA can be applied to an economic input–output table measured by either or both of the units. If a technical coefficient is defined as $D_{ij} = x_{(E),ij}/x_{\text{total(E)},j}$, Eq. 5 can be represented by Eq. 6.

$$\begin{bmatrix} x_{\text{total (E)},1} \\ x_{\text{total (E)},2} \end{bmatrix} = \begin{bmatrix} 1 - D_{11} & -D_{12} \\ -D_{21} & 1 - D_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_{(E),F1} \\ x_{(E),F2} \end{bmatrix}, \quad \text{or}$$
$$\mathbf{x}_{\text{total(E)}} = (\mathbf{I} - \mathbf{D})^{-1}\mathbf{x}_{(E),F} \tag{6}$$

where $\mathbf{x}_{\text{total(E)}}$ is a *total output vector*, $\mathbf{I}$ is an identity matrix, $\mathbf{D}$ is a *technical coefficient or direct requirements matrix*, and $\mathbf{x}_{\text{total(E)},F}$ is a *household demand vector*. For the example in Figure 2, the household demand vector is the resource consumption in V1, thus $\mathbf{y}_{(E)} = [s_{(V),1}x'_{(V),1} \;\; 0]^T$. The corresponding total output vector is represented by Eq. 7.

$$\mathbf{x}'_{\text{total(E)}} = (\mathbf{I} - \mathbf{D})^{-1}\mathbf{y}_{(E)} \tag{7}$$

The vector of $\mathbf{x}'_{\text{total(E)}}$ can be called the *scaled vector of total outputs* because the value of their elements, $x'_{\text{total(E)},i}$, has a linear relationship with $y_{(E)1}/x_{(E),F1}$ in this example. The inventory vector can be calculated by Eq. 8.

$$\mathbf{g}_{(E)} = \mathbf{B}_{(E)}\mathbf{x}'_{\text{total(E)}} \tag{8}$$

where $B_{(E),ij} = x_{(E),wij}/x_{\text{total},(E),i}$.

Formulation of Hybrid LCA. Methods for Hybrid LCA have been categorized into tiered hybrid LCA, economic input–output-based (IO-based) hybrid LCA, and integrated hybrid LCA.[26] Tiered hybrid LCA is the easiest to use, and considers EIOLCA and process LCA separately. For Figure 2, this implies that even though the processes of V1, V2, and V3 are a part of the economic sector S1, they are treated as if they were independent. The formulation of tiered hybrid LCA can be derived by combining the total output equations for each economic sector and mass balance equations between processes. Because the household demands of EIOLCA system must be supplied to the process LCA system, Eq. 9 must be satisfied.

$$\mathbf{C}^{u}_{(V)}\mathbf{s}_{(V)} - \mathbf{y}_{(E)} = \mathbf{0} \tag{9}$$

By substituting Eq. 9 into Eq. 7, Eqs. 4 and 7 can be rewritten as Eq. 10.

$$\begin{bmatrix} \mathbf{T}_{(V)} & \mathbf{0} \\ -\mathbf{C}^{u}_{(V)} & (\mathbf{I} - \mathbf{D}) \end{bmatrix} \begin{bmatrix} \mathbf{s}_{(V)} \\ \mathbf{x}'_{\text{total(E)}} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_{(V)} \\ \mathbf{0} \end{bmatrix} \tag{10}$$

The inventory vector of the tiered hybrid LCA, $\mathbf{g}_{t\text{Hybrid}}$, can be calculated by Eq. 11.

$$\mathbf{g}_{t\text{Hybrid}} = \begin{bmatrix} \mathbf{B}_{(V)} & \mathbf{B}_{(E)} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{(V)} \\ \mathbf{x}'_{\text{total(E)}} \end{bmatrix} \tag{11}$$

which is the sum of the corresponding equations for process LCA (Eq. 3) and EIOLCA (Eq. 8).

The formulations of the IO-based hybrid LCA and the integrated hybrid LCA are not shown in this article, but are easily derived from Eqs. 8 and 11, respectively. The integrated hybrid LCA approach does not separate process LCA and EIOLCA, but considers the systems of process LCA as a part of the EIOLCA system. Thus, a technical coefficient matrix is modified using the subtracted transactions and flows of values added and final demands, which make the model structure of the integrated hybrid LCA identical to that of the tiered hybrid LCA.[35] To utilize the subtracted information about transactions, values added, and final demands, the economic input–output tables must be prepared based on disaggregated economic sectors. The IO-based hybrid LCA utilizes disaggregated information about each economic sector. Thus, the matrix of technical coefficients and an intervention matrix are calculated using final demands, values added, and transactions for the disaggregated economic sectors, which makes the model structure of the IO-based hybrid LCA same as that of EIOLCA.[36] These existing methods for hybrid

LCA, like process LCA and EIOLCA, do not pay attention to thermodynamic consistency of the hybridized data. Consequently, these data can be improved by the data rectification approach described and illustrated in the rest of this article.

Rectification of LCI Data

Features of process data versus life cycle inventory data

LCI data have some unique features that do not permit direct application of traditional process data rectification methods. Some of the main differences are summarized in Table 1, and include the following: (1) In the process industries, it is common to measure and store physical quantities frequently and periodically. Thus, process data are usually in the form of time series, and a variance–covariance matrix of measurement errors may be estimated from such measurements. However, LCI data are much less common and only some databases report uncertainty; (2) It is quite common in process data rectification to ignore components with very small flow rates such as fugitive emissions. Such components are often very important for LCA due to their potential for environmental and human impact; (3) LCI data often contain aggregated streams such as dissolved solids and particulates, whose composition may not be specified. Aggregated energy streams are often reported in units of energy, and a lack of knowledge about their composition and net calorie value makes it difficult for rectification via combustion equations and the conservation law of mass. In an extreme case, emissions may be represented in terms of quantities such as biological oxygen demand (BOD), which does not represent the quantity or composition of the corresponding emission; (4) Accuracy of data rectification increases as the number of constraints that are composed of redundant variables increases. However, most LCI data do not contain information about a process flow diagram, making them "black-box" inventories. These inventories often contain only the amounts of resource consumption, waste emission, and product. Large number of missing flows often makes it impossible to rectify the data based on the available information. These challenges may be overcome by utilizing domain-specific process information, but collecting such information can be extremely tedious, since it is usually not included in LCI databases[10]; (5) Another characteristic of LCI data is that they are often available at multiple overlapping scales as described in the second section. Methods for hybrid LCA described in the second section combine such data but do not address any of the challenges due to inconsistency, uncertainty, and missing data; (6) The system boundary of LCA is usually much larger than that of process data rectification because several processes are included for LCA, while process data rectification focuses on a single process. This means that LCI rectification ought to consider a large number of reactions and other constraints. However, collecting entire chemical reaction equations for a life cycle network requires significant effort, and some chemical reactions may not be available. Thus, it is practically impossible to include all of the chemical reaction equations for the entire life cycle network, and a method may be required to guide which processes or economic sectors should utilize reaction equations as the constraints of data rectification.

Rectification of multiscale data for a single process

Data rectification of multiscale data for a single process is an extension of data rectification at a single scale. It is formulated to minimize errors in measured values at all scales, with constraints composed of intrascale and interscale balance equations. A general formulation of multiscale data reconciliation is shown in Eq. 12.

$$\begin{aligned} &\min_{\hat{\mathbf{x}}_{(m)}} \quad \sum_m \sum_i \frac{\left(x_{(m),i} - \hat{x}_{(m),i}\right)^2}{\sigma^2_{(m),i}} \\ &\text{s.t.} \\ &\mathbf{h}_{(m)}\left(\hat{\mathbf{x}}_{(m)}\right) = \mathbf{0}, \qquad (m = 1, 2, \ldots, M) \\ &\mathbf{g}_{(m,m+1)}\left(\hat{\mathbf{x}}_{(m)}, \hat{\mathbf{x}}_{(m+1)}\right) = \mathbf{0}, \qquad (m = 1, 2, \ldots, M-1) \end{aligned} \tag{12}$$

Here, $x_{(m),i}$ is the i-th measured variable at scale m, $\hat{x}_{(m),i}$ is the reconciled estimate of $x_{(m),i}$, $\sigma^2{}_{(m),i}$ is the i-th diagonal element of the variance–covariance matrix for measurement errors at scale m, and M represents the finest scale. The first constraint represents conservation laws at a single scale m, while the second constraint represents multiscale conservation laws between scales m and $m+1$. In Eq. 12, the off-diagonal elements of a variance–covariance matrix are assumed to be zero. Equation 12 can be extended to data rectification using several gross error compensation methods, and Eq. 13 shows a general model of data rectification using mixed integer programming.[19]

$$\begin{aligned} &\min_{\hat{\mathbf{x}}_{(m)}, \mathbf{B}_{(m)}, \delta_{(m)}} \quad \sum_m \sum_i \frac{\left(x_{(m),i} - \hat{x}_{(m),i} - \delta_{(m),i}\right)^2}{\sigma^2_{(m),i}} + W_{(m),i}\, B_{(m),i} \\ &\text{s.t} \\ &\mathbf{h}_{(m)}\left(\hat{\mathbf{x}}_{(m)}\right) = \mathbf{0}, \qquad (m = 1, 2, \ldots, M) \\ &\mathbf{g}_{(m,m+1)}\left(\hat{\mathbf{x}}_{(m)}, \hat{\mathbf{x}}_{(m+1)}\right) = \mathbf{0}, \qquad (m = 1, 2, \ldots, M-1) \\ &\forall m, i \qquad \left|\delta_{(m),i}\right| - U_{(m),i}\, B_{(m),i} \le 0 \\ &\qquad\qquad \left|\delta_{(m),i}\right| - \zeta_{(m),i}\, U_{(m),i}\, B_{(m),i} \ge 0 \end{aligned} \tag{13}$$

Table 1. Different Features of Rectification for Process Data and LCI Data

	LCI Data Rectification	Process Data Rectification
Time windows	Usually single time point	Time series of data are often available.
Chemical species	Often aggregated as solid waste, particulates, dissolved solids, etc.	Aggregated chemical species are negligible
Reaction models of (trace) wastes	Important	Not so important
Redundancy	Usually not redundant	Redundant
Infrastructure	Black-box	Detailed flowsheet is often available
Data sources	Multiple overlapping scales	Single scale
Processes included	Several processes	One process

Here, $\delta_{(m),i}$ is the gross error size, $W_{(m),i}$ is a penalty function, $B_{(m),i}$ is a binary variable to indicate the existence of a gross error, and $U_{(m),i}$ can be considered as the upper limit of the absolute size of a gross error in the i-th measured variable at scale m. The objective function is to minimize rectification errors while estimating rectified values and gross error sizes at all scales. The first two constraints are identical to those of the data reconciliation problem in Eq. 12. The third constraint places upper bounds on the absolute sizes of gross errors. The user-defined variable ζ must be selected such that the value of ζU is some fraction of the error standard deviation. Therefore, the fourth constraint is analogous to a statistical test based on the measurement error variance. The computational expense of solving the optimization problem in Eq. 13 may be reduced by identifying a candidate set of gross errors by statistical methods, thus reducing the number of binary variables.[20] If such candidate set of gross errors is found, binary variables are only used for the measured variables that belong to this set.

Rectification of multiscale data from networks

The approach for multiscale data rectification developed in the previous subsection needs to be extended from a single system to a network of systems. If each set of measurement data such as $\{x'_{(V),1}, x_{(V),2}, x_{(V),5}\}$ and $\{x'_{(V),2}, x_{(V),3}, x_{(V),6}\}$ in Figure 2 are collected independently, then the corresponding scaling factors need to be calculated as described in the second section. These factors should be multiplied to the measured variables so that the conservation laws between measurement sets are satisfied based on the scaled measurement sets (e.g., $s_{(V),1}, x_{(V),2} - s_{(V),2}\, x'_{(V),2} = 0$ in Figure 2). A general formulation of multiscale data rectification for networks is represented in Eq. 14.

$$\min_{\hat{\mathbf{x}}_{(m)},\mathbf{B}_{(m)},\delta_{(m)}} \sum_m C_{(m)} \sum_i \frac{\left(s_{(m),i}x_{(m),i} - \hat{x}_{(m),i} - \delta_{(m),i}\right)^2}{s_{(m),i}\sigma^2_{(m),i}} + W_{(m),i}\; B_{(m),i}$$

$$\begin{aligned}
&\text{s.t.}\\
&C_{(m)}\mathbf{h}_{(m)}\left(\hat{\mathbf{x}}_{(m)}\right) = \mathbf{0}, && (m = 1,2,\ldots,M)\\
&C_{(m,m+1)}\mathbf{g}_{(m,m+1)}\left(\hat{\mathbf{x}}_{(m)}, \hat{\mathbf{x}}_{(m,m+1)}\right) = \mathbf{0}, && (m = 1,2,\ldots,M-1)\\
&\forall m, i \quad |\delta_{(m),i}| - U_{(m),i}B_{(m),i} \le 0\\
&\qquad\quad |\delta_{(m),i}| - \zeta_{(m),i}U_{(m),i}B_{(m),i} \ge 0
\end{aligned} \tag{14}$$

where, $s_{(m),i}$ is a scaling factor related to the i-th variable at scale m, $C_{(m)}$ is a binary constant to represent the existence of scale m, and $C_{(m,m+1)} = C_{(m)} \wedge C_{(m+1)}$. When data at scale m are available, $C_{(m)}$ is equal to one; otherwise, it is equal to zero.

Multiscale rectification of LCI data

The data rectification model integrated with the computational methods of LCA, summarized in the second section, can be derived from Eq. 14. Considering the network diagram of hybrid LCA in Figure 2, and assuming that information about the equipment scale is not available, this implies that $C_{(E)} = 1$, $C_{(V)} = 1$, and $C_{(Q)} = 0$. Because LCA is concerned with the mass amounts of input and output flows, constraint equations are composed only of mass balance equations. The scaling factor of the value chain scale is already defined as Eq. 4. The scaling factors of the i-th economic sector is represented in Eq. 15. The numerator is the scaled vector of total outputs in Eq. 8, and denominator is the total output vector in Eq. 6.

$$s_{(E),i} = \frac{x'_{\text{total}(E),i}}{x_{\text{total}(E),i}} \tag{15}$$

Therefore, a generalized rectification formulation integrated with the computation methods of hybrid LCA can be represented in Eq. 16.

$$\min_{\hat{\mathbf{x}},\mathbf{B},\delta} \sum_{m=E,V}\left[C_{(m)} \sum_i \frac{\left(s_{(m),i}x_{(m),i} - \hat{x}_{(m),i} - \delta_{(m),i}\right)^2}{s_{(m),i}\sigma^2_{(m),i}}\right] + W_{(m),i}\; B_{(m),i}$$

$$\begin{aligned}
&\text{s.t.}\\
&C_{(E)}\mathbf{A}_{(E)}\hat{\mathbf{x}}_{(E)} + C_{(E)}\;\Theta^T_{(E)}\xi_{(E)} = \mathbf{0}\\
&C_{(V)}\mathbf{A}_{(V)}\hat{\mathbf{x}}_{(V)} + C_{(V)}\Theta^T_{(V)}\xi_{(V)} = \mathbf{0}\\
&C_{(E,V)}\mathbf{A}_{(E,V)}\begin{bmatrix}\hat{\mathbf{x}}_{(E)}\\ \hat{\mathbf{x}}_{(V)}\end{bmatrix} = \mathbf{0}\\
&\forall m, i \quad \left|\delta_{(m),i}\right| - U_{(m),i}B_{(m),i} \le 0\\
&\qquad\quad \left|\delta_{(m),i}\right| - \zeta_{(m),i}U_{(m),i}B_{(m),i} \ge 0
\end{aligned} \tag{16}$$

where, Θ is the matrix of stoichiometric coefficients, ξ is the vector of reaction extents, and $C_{(E,V)} = C_{(E)} \wedge C_{(V)}$. The first and second constraints are mass balance equations at the economy and value chain scales, respectively. These equations are usually not available in existing LCI data, and need to be obtained from engineering knowledge. The third constraint represents mass balance equations between scales. The scaling factors of the value chain scale must be calculated by Eq. 4, and the scaling factors for the economy scale must be calculated by Eq. 15. According to the values of C_V and C_E, the elements of the final and household demand vectors, and the technical coefficient matrix being aggregated or disaggregated, the general formulation of Eq. 16 can be converted into the data rectification model of process LCA, EIOLCA, or hybrid LCA, which is summarized in Table 2.

Rectification for Process LCA. The general formulation of Eq. 16 specializes to the data rectification model integrated with process LCA if $C_V = 1$, $C_E = 0$, and the reference flow, $\phi_{(V)}$, is used as the element of the final demand vector, $\mathbf{f}_{(V)}$. Only the scaling vector for the computational method of process LCA, $\mathbf{s}_{(V)}$, is utilized in the data rectification formulation. This model can be applied to data rectification for process LCA, but is also useful for hybrid LCA when there is no violation of the conservation laws in economic sectors.

Rectification for EIOLCA and IO-Based Hybrid LCA. The general formulation of Eq. 16 specializes to the data rectification model integrated with EIOLCA or IO-based hybrid LCA if $C_V = 1$, $C_E = 0$, and the reference flow, $\phi_{(V)}$, is used as the element of the household demand vector, $\mathbf{y}_{(E)}$. In this case, only the scaling vectors of economic sectors, $\mathbf{s}_{(E)}$, given by Eq. 15, are utilized in the data rectification formulation. As discussed in the second section, EIOLCA utilizes aggregated information of economic sectors, but IO-based hybrid LCA uses some disaggregated economic sectors. Thus, the

Table 2. Usage of a Generalized Data Rectification Model Integrated with LCA Computational Methods

	$C_{(V)}$	$C_{(E)}$	Vector of Household Demands, $\mathbf{y}_{(E)}$	Vector of Final Demands, $\mathbf{f}_{(V)}$	Scaling Vector of Economy $\mathbf{s}_{(E)}$	Scaling Vector of Value Chain Scale, $\mathbf{s}_{(V)}$	Matrix of Technical Coefficients, $\mathbf{D}$
Process LCA	1	0	None	$\phi_{(V)}$	None	$\mathbf{T}^{-1}{}_{(V)}\mathbf{f}_{(V)}$	None
EIOLCA	0	1	$\phi_{(V)}$	None	$\frac{x'_{\text{total(E)},i}}{x_{\text{total(E)},i}}$	None	Aggregated
Tiered hybrid LCA	1	1	$\mathbf{C}^{u}{}_{(V)}\mathbf{s}_{(V)}$	$\phi_{(V)}$	$\frac{x'_{\text{total(E)},i}}{x_{\text{total(E)},i}}$	$\mathbf{T}^{-1}{}_{(V)}\mathbf{f}_{(V)}$	Aggregated
IO-hybrid LCA	0	1	$\phi_{(V)}$	None	$\frac{x'_{\text{total(E)},i}}{x_{\text{total(E)},i}}$	None	Disaggregated
Integrated hybrid LCA	1	1	$\mathbf{C}^{u}{}_{(V)}\mathbf{s}_{(V)}$	$\phi_{(V)}$	$\frac{x'_{\text{total(E)},i}}{x_{\text{total(E)},i}}$	$\mathbf{T}^{-1}{}_{(V)}\mathbf{f}_{(V)}$	Disaggregated

general formulation of Eq. 16 becomes the data rectification model integrated with EIOLCA or with IO-based hybrid LCA if a technical coefficient matrix is calculated using aggregated or disaggregated information, respectively.

Rectification for Tiered and Integrated Hybrid LCA. The general formulation of Eq. 16 specializes to the data rectification model integrated with tiered or integrated hybrid LCAs if $C_V = 1$, $C_E = 0$, the reference flow, $\phi_{(V)}$, is used as the element of the final demand vector, $\mathbf{f}_{(V)}$, and the scaled upstream cutoffs are used as household demands (i.e., $\mathbf{C}^{u}_{(P)}\mathbf{s}_{(P)} - \mathbf{y}_{(E)} = \mathbf{0}$). Tiered hybrid LCA does not integrate the system of process LCA within economic sectors, while integrated hybrid LCA does. Thus, the general formulation of Eq. 16 becomes the data rectification model integrated with the tiered hybrid LCA method if a technical coefficient matrix is calculated using the aggregated information of economic sectors. If a technical coefficient matrix is calculated using disaggregated information of economic sectors, then the rectification model of Eq. 16 becomes the data rectification model integrated with the integrated hybrid LCA.

The rectification model integrated with LCA methods of Eq. 16 usually yields large problems, which may require much computation time. In this case, decomposition optimization methods may be more appropriate such as auxiliary problem principle and block coordination descent method.[37] These methods are mainly used for iterative solution of large-scale optimization problems that present a structure of interrelated systems. In addition to the benefit of dimension reduction, the decomposition optimization methods make it possible to partition a multiscale problem into single scale subproblems, with variables that connect different scales appearing in each single scale subproblem. In addition, methods for sequential processing of constraints or measurements[38] can be applied to avoid solving the full-scale rectification problem every time upon receiving new information. In particular, the sequential processing of constraints may be useful to LCI data rectification when additional constraints become available. Usually, all constraints such as the reaction equations of waste generation are not available at once, and the rectification problem should be solved with partial information. The accuracy of these results may be enhanced as more information becomes available, making such problems ideal for sequential processing of constraints because it is a recursive method. Detailed formulations and examples about decomposition optimization and constraints processing are found elsewhere.[39]

Table 3. Measured Data of First Case Study

Measured Variables	Equipment Scale Data, kg/h	Plant Scale Data, kg/h
$x_{(P),4}$		4.99
$x_{(Q),8}$	4.97	
$x_{(P),5}$		1005.40
$x_{(Q),4}$	1020.41	
$x_{(P),1}$		991.08
$x_{(Q),1}$	414.87	
$x_{(Q),5}$	602.32	
$x_{(P),2}$		8.91
$x_{(Q),6}$	5.01	
$x_{(Q),7}$	4.07	
$x_{(Q),2}$	412.20	
$x_{(Q),3}$	400.65	
$x_{(P),3}$		1.00

Case Studies

Rectification of multiscale data for a single process

This example illustrates the use of the proposed multiscale data rectification method for data from the process industries.

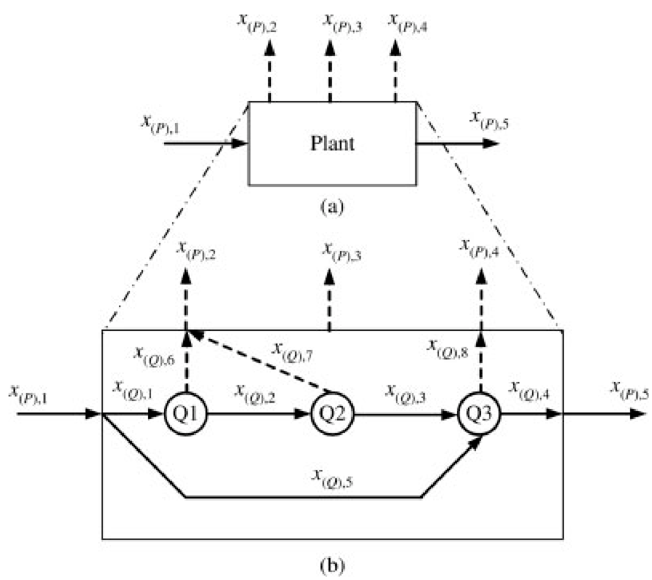


Figure 3. Network diagram of a single process at two scales.

(a) Only raw materials, product and wastes are represented at the plant scale; (b) detailed flowsheet and every intermediate stream are shown at the equipment scale.

Table 4. Rectification Results by Multiscale Model and Single Scale Model of First Case Study

		Rectified Results by Multiscale Formulation			Rectified Results by Single Scale Formulation		
Measured Variables	Simulated Values	Rectified Values	Binary Variables, B_i	Square Errors	Rectified Values	Binary Variables, B_i	Square Errors
$x_{(P),4}$	5	4.92	0	6.40×10^{-3}			
$x_{(Q),8}$	5	4.92	0	6.40×10^{-3}	4.92	0	6.40×10^{-3}
$x_{(P),5}$	1000	1000.84	0	7.06×10^{-1}			
$x_{(Q),4}$	1000	1000.84	1	7.06×10^{-1}	1009.47	0	89.68
$x_{(P),1}$	1015	1015.66	1	4.36×10^{-1}			
$x_{(Q),1}$	415	413.99	0	1.02	415.69	0	4.76×10^{-1}
$x_{(Q),5}$	600	601.67	0	2.79	608.78	0	77.09
$x_{(P),2}$	9	8.91	0	8.10×10^{-3}			
$x_{(Q),6}$	5	4.9	0	1.00×10^{-2}	5	0	0.0
$x_{(Q),7}$	4	4.01	0	1.00×10^{-4}	4.08	0	6.40×10^{-3}
$x_{(Q),2}$	410	409.09	0	8.28×10^{-1}	410.69	0	4.76×10^{-1}
$x_{(Q),3}$	405	404.09	0	8.28×10^{-1}	405.61	0	3.72×10^{-1}
$x_{(P),3}$	1	0.98	0	4.00×10^{-4}	1	0	0.0

Table 3 shows measurement data at the plant and equipment scales for Figure 3. Variables $x_{(P),1}$ and $x_{(P),5}$ represent the shipping amounts of a raw material and a product, respectively. Variables $x_{(P),2}$, $x_{(P),3}$, and $x_{(P),4}$ are wastes that are collected in storage tanks before being sent to a waste treatment firm. However, there is no flow meter for $x_{(P),3}$; hence its measurement is only available at the plant scale. $x_{(Q),i}$ is measured via flow meters on process equipment. Table 3 shows the measured values at the plant and equipment scales after their contamination by random and gross errors.

The data rectification approach of Eq. 13 is applied using the measured values at plant and equipment scales, and traditional data rectification at single scale is also applied using equipment scale information for comparison. The constraints for multiscale data rectification are mass balance equations at the plant and equipment scales, and across the scales. Mass balance equations at only the equipment scale are used for constraints of single scale rectification. It is assumed that a variance–covariance matrix of measurement errors is available, and that the off-diagonal values of this matrix are zero, with the diagonal values being 5% of the measured values. The value of ζ is selected as 50% of the corresponding standard deviation, and the value of U is selected as 50% of the corresponding measured value. Table 4 compares the simulated values with the rectified values obtained by the multiscale and single scale models. Two gross errors are found by the multiscale model, which are $x_{(P),1}$ and $x_{(Q),4}$, and their results of simultaneous data reconciliation and gross error compensation are shown in the third column of Table 4. No gross error is found by the single scale model even though $x_{(Q),4}$ contains large gross errors. Table 4 shows that the rectification errors of the multiscale model are not significant compared with the simulated values. However, the results by the single scale model cannot identify gross errors, and the gross error in $x_{(Q),4}$ is propagated to the variable $x_{(Q),5}$. Thus, the square errors of $x_{(Q),4}$ and $x_{(Q),5}$ by single scale rectification are much larger than those by multiscale rectification.

Rectification of the life cycle inventory data of a caustic soda process

This case study applies the developed method to the inventory data of a caustic soda process compiled by the National Renewable Energy Laboratory (http://www.nrel.gov/lci). This is LCI data at the value chain scale. It is assumed that detailed engineering information about the caustic soda process is available. Thus, data at the equipment scale are prepared by a material balance model, and random and gross

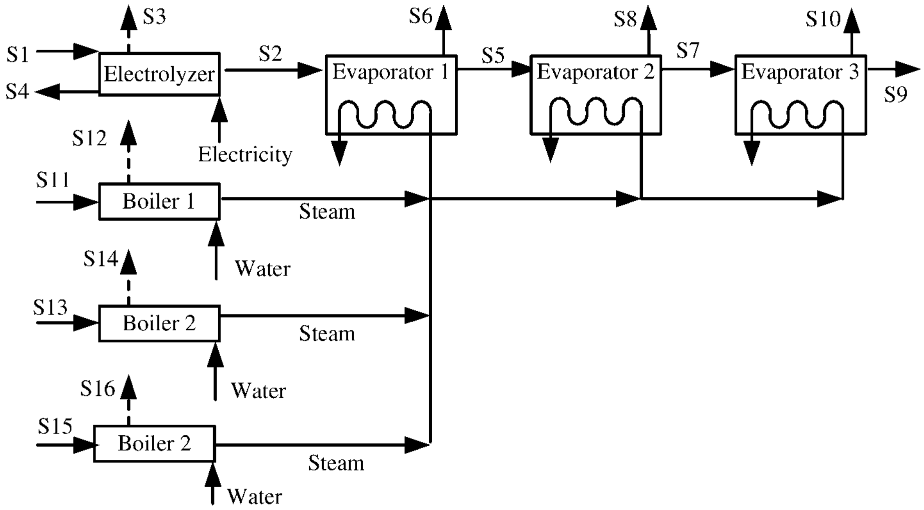
Table 5. LCI of the Caustic Soda Process from the National Renewable Energy Laboratory's Database

	Streams	Measured Values
Raw Material	Salt	1670.0 lb
Energy input	Natural gas	178.1 lb
	Coal	68.1 lb
	Residual oil	3.8 lb
	Electricity	813 kWh
Product and coproduct	Caustic soda	1000.0 lb
	Chorine	893.0 lb
	Hydrogen	16.9 lb
Air emission	Chorine	3.2×10^{-4} lb
	Mercury	5.4×10^{-4} lb
	Particulates	4.6×10^{-4} lb
	Sulfur oxides	1.0×10^{-3} lb
Water effluents	Dissolved solids	4.3 lb
	Lead	9.4×10^{-7} lb
	Mercury	1.5×10^{-6} lb
	Nickel	9.4×10^{-7} lb
	Sulfides	1.5×10^{-4} lb
	Zinc	9.4×10^{-7} lb
Solid waste	Solid waste	3.1 lb
Total input		1920.0 lb
Total output		1917.3 lb

Table 6. Atomic Balance of LCI for the Caustic Soda Process

Atoms	Inputs, lb	Outputs, lb	Differences, lb
Carbon, C	175.8	0.0	175.8
Chlorine, Cl	1013.1	893.0	120.1
Hydroger, H	48.9	42.1	6.8
Lead, Pb	0.0	9.4×10^{-7}	-9.4×10^{-7}
Mercury, Hg	0.0	5.4×10^{-4}	-5.4×10^{-4}
Nickel, Ni	0.0	9.4×10^{-7}	-9.4×10^{-7}
Nitrogen, N	0.5	0.0	0.5
Oxygen, O	13.7	400.0	−386.3
Sodium, Na	656.9	574.8	82.1
Sulfur, S	0.9	6.4×10^{-4}	0.9
Zinc, Zn	0.0	9.4×10^{-7}	-9.4×10^{-7}
Total	*1920.0*	*1917.3*	*2.7*

Table 7. LCI Data of the Caustic Soda Process at Equipment Scale



Streams	Components	Chemical Formula	Amount, lb
S1	Sodium chloride	NaCl	1022.8
	Water	H_2O	4187.3
S2	Sodium hydroxide	NaOH	1008.0
	Water	H_2O	3720.6
S3	Chloride	Cl_2	10.1
S4	Chloride	Cl_2	867.5
	Hydrogen	H_2	25.5
S5	Sodium hydroxide	NaOH	1030.0
	Water	H_2O	3078.6
S6	Steam	H_2O	633.8
S7	Sodium hydroxide	NaOH	1008.0
	Water	H_2O	2469.5
S8	Steam	H_2O	597.6
S9	Sodium hydroxide	NaOH	1000.0
	Water	H_2O	1906.2
S10	Steam	H_2O	582.5
S11	Coal		
	Sulfur	S	4.9
	Hydrogen	H	3.1
	Carbon	C	44.4
	Oxygen	O	7.8
	Nitrogen	N	0.5
	Water	H_2O	7.3
	Ash	–	22.4
	Oxygen	O_2	138.4
S12	Carbon dioxide	CO_2	158.7
	Steam	H_2O	34.8
	Sulfur dioxide	SO_2	9.8
	Nitrogen dioxide	NO_2	1.7
	Hydrogen sulfide	H_2S	4.4×10^{-3}
	Ash	–	21.9
S13	Natural gas	CH_4	100.0
	Oxygen	O_2	715.2
S14	Steam	H_2O	402.0
	Carbon dioxide	CO_2	401.0
S15	Residual oil		
	Sulfur	S	3.0×10^{-2}
	Hydrogen	H	0.4
	Carbon	C	2.8
	Oxygen	O_2	9.9
S16	Sulfur dioxide	SO_2	5.9×10^{-2}
	Steam	H_2O	3.8
	Carbon dioxide	CO_2	10.1

errors are introduced to emulate measured values. Table 5 contains the LCI data from the National Renewable Energy Laboratory. BOD is neglected because of the lack of a rigorous model for converting BOD into an equivalent mass of emission. Natural gas and residual oil are reported in units of volume in the original process inventory. The units of natural

Table 8. Reaction Equations for Data Rectification of a Caustic Soda Process

Equipment	Chemical Reactions
Eletrolyzer	$2NaCl + 2H_2O \rightarrow 2NaOH + Cl_2 + H_2$
Boiler 1	$S + O_2 \rightarrow SO_2$
	$2H + 0.5O_2 \rightarrow H_2O$
	$C + O_2 \rightarrow CO_2$
	$N + O_2 \rightarrow NO_2$
	$H_2O + SO_2 \rightarrow 1.5O_2 + H_2S$
Boiler 2	$CH_4 + 2O_2 \rightarrow 2H_2O + CO_2$
Boiler 3	$S + O_2 \rightarrow SO_2$
	$2H + 0.5O_2 \rightarrow H_2O$
	$C + O_2 \rightarrow CO_2$

gas and residual oil are converted into pounds by assuming the specific volume of 22.68 ft^3/lb for pure methane and the density of 7.83 lb/gal for residual oil, respectively.[40] Table 5 shows the amounts of total material input and output. The difference between the total input and the total output is about 2.7 lb, and it seems that the inconsistency of mass conservation is not large.

Table 6 shows the inconsistency between inputs and outputs by atomic balances. The standard compositions of coal and residual oil are utilized for atomic balances.[40] Table 6 shows that considerable difference between inputs and outputs exists in atomic balances even though the difference between total input and output is not significant. Lead, mercury, nickel, and zinc are assumed to come from process equipment in this caustic soda process.[41] Alternatively, these elements may be assumed to originate from coal or any other source, without any effect on other rectified values.[10] Carbon is consumed as an energy source, but there is no carbon output, which may be in the form of carbon oxides. Nitrogen is consumed because it is contained in coal, but there is no output of nitrogen, which may be in the form of nitrogen oxides. There are large discrepancies in chlorine, hydrogen,

Table 9. Comparison of Data Rectification Results with Raw LCI Data

					Rectified Results by Multiscale Model		Rectified Results by Single Scale Model	
	Components	Chemical Formula	Measured Values at Equipment Scale	Measured Values at Value Chain Scale	Rectified Values	Binary Variables ($B_{(Q)}$, $B_{(V)}$)	Rectified Values	Binary Variables ($B_{(Q)}$)
S1	Sodium chloride	NaCl	1022.8	1670.0	1461.2	(1, 1)	1461.2	1
	Water	H_2O	4187.3		4169.7	(0, ·)	4169.7	0
S2	Sodium hydroxide	NaOH	1008.0		1000.0	(0, ·)	1000.0	0
	Water	H_2O	3720.6		3719.3	(0, ·)	3719.3	0
S3	Chlorine	Cl_2	10.1	3.2×10^{-4}	0.6	(1, 1)	10.2	0
S4	Chlorine	Cl_2	867.5	893.0	883.2	(1, 1)	876.2	0
	Hydrogen	H_2	25.5	16.9	25.2	(0, 1)	25.2	0
S5	Sodium hydroxide	NaOH	1030.0		1000.0	(1, ·)	1000.0	1
	Water	H_2O	3078.6		3082.6	(0, ·)	3082.6	0
S6	Steam	H_2O	633.8		636.7	(0, ·)	636.7	0
S7	Sodium hydroxide	NaOH	1008.0		1000.0	(0, ·)	1000.0	0
	Water	H_2O	2469.5		2483.1	(0, ·)	2438.1	0
S8	Steam	H_2O	597.6		599.5	(0, ·)	599.5	0
S9	Sodium hydroxide	NaOH	1000.0	1000.0	1000.0	(0, 0)	1000.0	0
	Water	H_2O	1906.2		1901.9	(0, ·)	1901.9	0
S10	Steam	H_2O	582.5		581.2	(0, ·)	581.2	0
S11	Coal							
	Sulfur	S	4.9	0.9	4.6	(1, 1)	4.9	0
	Hydrogen	H	3.1	2.8	3.1	(0, 0)	3.1	0
	Carbon	C	44.4	39.3	43.6	(0, 1)	43.5	0
	Oxygen	O	7.8	7.6	7.7	(0, 0)	7.8	0
	Nitrogen	N	0.5	0.5	0.5	(0, 0)	0.5	0
	Water	H_2O	7.3	6.8	7.1	(0, 0)	7.3	0
	Ash	–	22.4	10.2	12.5	(1, 1)	22.2	0
	Oxygen	O_2	138.4		138.4	(0, ·)	138.6	0
S12	Carbon dioxide	CO_2	158.7		159.6	(0, ·)	159.4	0
	Steam	H_2O	34.8		34.5	(0, ·)	34.8	0
	Sulfur dioxide	SO_2	9.8	1.0×10^{-3}	9.1	(1, 1)	9.8	0
	Nitrogen dioxide	NO_2	1.7		1.7	(0, ·)	1.7	0
	Hydrogen sulfide	H_2S	4.4×10^{-3}	1.5×10^{-4}	2.9×10^{-4}	(0, 0)	4.4×10^{-3}	0
	Ash	–	21.9	7.4	12.5	(1, 1)	22.2	0
S13	Natural gas	CH_4	100.0	178.1	179.0	(1, 0)	179.2	1
	Oxygen	O_2	715.2		714.2	(0, ·)	714.8	0
S14	Steam	H_2O	402.0		402.1	(0, ·)	402.4	0
	Carbon dioxide	CO_2	401.0		491.1	(1, ·)	491.5	1
S15	Residual oil							
	Sulfur	S	3.0×10^{-2}	3.8×10^{-2}	3.1×10^{-2}	(0, 0)	2.9×10^{-2}	0
	Hydrogen	H	0.4	0.6	0.4	(0, 0)	0.4	0
	Carbon	C	2.8	3.2	2.7	(0, 0)	2.7	0
	Oxygen	O_2	9.9		10.6	(0, ·)	10.4	0
S16	Sulfur dioxide	SO_2	5.9×10^{-2}		6.2×10^{-2}	(0, ·)	5.9×10^{-2}	0
	Steam	H_2O	3.8		3.7	(0, ·)	3.6	0
	Carbon dioxide	CO_2	10.1		10.0	(0, ·)	9.8	0

oxygen, and sodium. Especially, oxygen must be consumed for the combustion of coal, residual oil, and natural gas, but there is no oxygen input in the LCI data of the caustic soda process. As discussed in the Introduction, these inconsistencies can be particularly important if LCI data are used for methods such as mass, energy, or exergy analysis.

Table 7 shows the LCI data of the caustic soda process at the equipment scale that includes a flowsheet and flow rates, while the LCI data at the value chain scale contains only the amounts of raw material, waste emission, coproduct, and product as shown in Table 5. Brine is supplied to an electrolyzer for producing sodium hydroxide, hydrogen, and chlorine. The produced sodium hydroxide is an aqueous solution, thus it is fed to evaporators to concentrate it. Steam used in the evaporators is supplied from three boilers that consume coal, natural gas, and residual oil. Lead, mercury, nickel, and zinc are removed from the formulation of data rectification because these chemical species are assumed to come from the process equipment of the caustic soda process as discussed in the previous paragraph.

The model of data rectification for hybrid LCAs requires an incidence matrix, reaction equations, and component flows at the economy scale. However, reactions and component flows are difficult to find at the economy level. Thus, the data rectification model integrated with process LCA discussed in the third section is applied to the LCI data at value chain and equipment scales. After that, EIOLCA is applied to the cutoff streams for the tiered hybrid LCA. It is assumed that covariance values are zero and variance values are equal to 5% of measured values in this study. The value of ζ is selected as 50% of standard deviation, and the value of U is selected as 50% of the measured value. Component balance equations for all equipment are used as constraints, and reaction equations are shown in Table 8.

The reaction equations in this study are based on our previous work on single-scale LCI data reconciliation of a caustic soda process.[10] In their study, four alternatives of flowsheet information are utilized for data reconciliation of the caustic soda process with fixed fuel compositions, and the number of identified gross errors is examined for the alternatives. However, no gross error compensation method has been applied to remove gross errors from measured values. If gross errors are not compensated, their effect may be propagated into other variables. Consequently, the accuracy of data reconciliation decreases. This study implements simultaneous data reconciliation and gross error compensation to estimate gross error sizes. Furthermore, the single scale data rectification of Hau et al. is extended to a multiscale formulation integrated with the computational method of LCA.[24] Thus, results of LCA with multiscale rectification can include the entire economy using a hybrid LCA method.

Table 9 compares the rectified results by the multiscale and single scale models. Measured values at the equipment scale are used in a single scale model. LCI data at the value chain scale have many missing flows such as water consumption in a caustic soda process, oxygen consumption in boilers, and the emission of carbon dioxide in boilers, which means that data rectification can not be applied to the LCI data at the value chain scale. LCI data at the value chain and equipment scales may contain gross errors, and it might be helpful for estimating accurate rectified values if LCI data at both scales are utilized in data rectification. The sixth and eighth columns in Table 9 show the rectified values by the multiscale and single scale models, respectively. The seventh and ninth columns in Table 9 show the results of binary variables to determine the existence of gross errors. The consumption amounts of sodium chloride (S1) at the value chain and equipment scales contain gross errors, which are compensated by rectification with the multiscale and single scale models. The gross error in the emission amount of chlorine (S3) can not be identified and removed by the data rectification of the single scale model. Thus, the amount of chlorine in S4 cannot be rectified accurately by the single scale model. However, the gross errors in the measured values of the emission amount of chlorine (S3) at the equipment and value chain scales can be identified and removed by the data rectification of the multiscale model. The large difference between the rectified values by the single scale model and multiscale model makes the results of LCA by the single scale model significantly different from those by the multiscale model. In addition, the amounts of sulfur and ash contained in coal (S11) at the equipment and value chain scales also include gross errors, which are removed by the data rectification of the multiscale model, while the single scale data rectification cannot identify a gross error in the equipment scale LCI data. The compensation of gross errors in the sulfur and ash contents of coal makes it possible to estimate the emission amounts of sulfur dioxide and ash in S12 accurately. However, the rectified values of sulfur dioxide and ash by the single scale model are similar to the measured values at the equipment scale because gross errors cannot be identified. Consequently, data rectification with a multiscale model has the benefit of being able to identify and compensate gross errors more accurately than rectification with a single scale model.

Tiered hybrid LCA is applied to the rectified values by a multiscale model and a single scale model, and the LCI data at the equipment and value chain scales. Figure 4 shows the resource consumption of hybrid LCA results, and Figure 5 represents the emission of hybrid LCA results. In both figures, HLCA I represents hybrid LCA using LCI data at the equipment and economy scales, while HLCA II represents

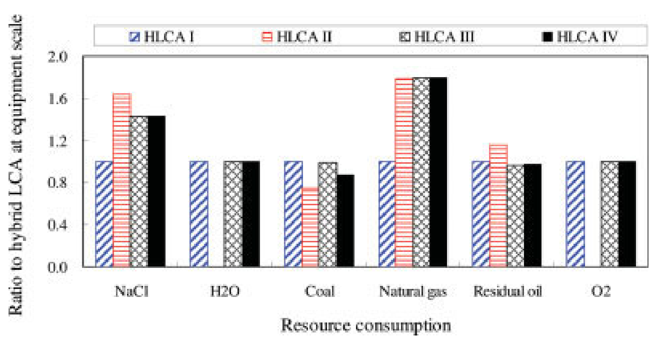


Figure 4. Hybrid LCA results for resources.

Without rectification of resource consumption by LCI data at equipment (HLCA I) and value chain (HLCA II) scales. Rectified data with single (HLCA III) and multiscale (HLCA IV) models. All results are scaled with respect to the results of HLCA I. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

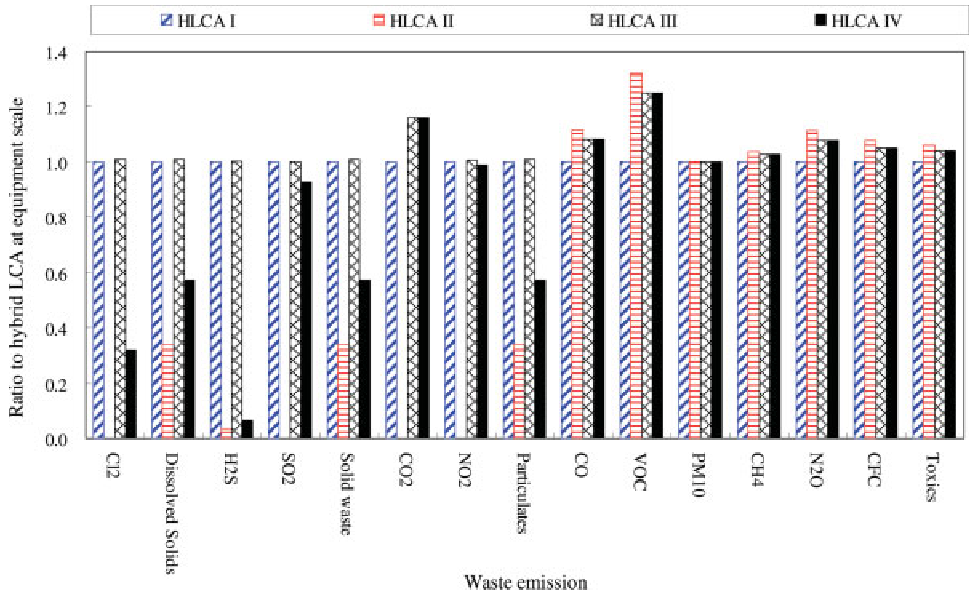


Figure 5. Hybrid LCA results for emissions.

Without rectification of emissions by LCI data at equipment (HLCA I) and value chain (HLCA II) scales. Rectified data with single (HLCA III) and multiscale (HLCA IV) models. All results are scaled with respect to the results of HLCA I. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

hybrid LCA using LCI data at the value chain and economy scales. These two sets of results are without rectification. Results of HLCA III are obtained by hybrid LCA using the rectified LCI data by a single scale model at the equipment scale and economic LCI data, and HLCA IV are obtained by hybrid LCA using the rectified LCI data by a multiscale model and economic LCI data. For easier comparison, all hybrid LCA results are divided by the results of HLCA I. Thus, the results of HLCA I are equal to 1. In Figure 4, the consumption amounts of water and oxygen by HLCA II are equal to zero even though they are utilized to generate power, while the results by HLCA I, III, and IV have no missing value. In addition, the LCI data for HLCA I and II do not satisfy the conservation law of mass due to which these results may contain more errors. Data rectification removes measurement errors, thus the LCA results by the rectified values such as HLCA III and IV can be more accurate than those by the raw LCI data such as HLCA I and II. The consumption amounts of resources by HLCA III are similar to those by HLCA IV except coal consumption. The gross errors in the ash and sulfur contents of coal are removed by multiscale data rectification, while no gross error is identified by single scale data rectification. Thus, the consumption amount of coal by HLCA III is similar to that by HLCA I, but the coal consumption by HLCA IV utilizes the rectified values based on the LCI data at the equipment and value chain scales.

The results of the tiered hybrid LCA for waste emission are shown in Figure 5. The LCA results of lead, mercury, nickel, and zinc are omitted because these variables are not redundant, and rectified values are equal to measured values. The emission amounts of carbon monoxide, VOC, PM10, methane, nitrous oxide, CFC, and toxics by rectified values and raw LCI data are similar to each other. However, the emission amounts of carbon dioxide and nitrogen dioxide by HLCA II are nearly zero because carbon dioxide and nitrogen dioxide are omitted from LCI data at the value chain scale. This is physically impossible because boilers consume fossil fuels for energy. However, the results by HLCA I, III, and IV can evaluate the emission amounts of carbon dioxide and nitrogen dioxide. The emission results of chlorine and sulfur dioxide by HLCA III are similar to those by HLCA I because gross errors can not be removed. However, the emission results of chlorine and sulfur dioxide by HLCA IV can remove the gross errors from the raw LCI data, which gives significantly different results. Usually, the LCA results by the rectified values using the single scale model are similar to those by LCI data at the equipment scale, and the LCA results by LCI data at the value chain scale have a few missing flows. However, the LCA results by the multiscale data rectification can estimate the missing flows, and gives weighted average values of LCI data at the equipment and value chain scales satisfying the conservation laws.

Conclusion

Satisfaction of mass and energy balances is a fundamental requirement according to the first law of thermodynamics. As is well known in process engineering, measured data may violate this law due to contamination by noise and errors. Similarly, many LCI data also do not satisfy the conservation law of mass, and the LCA results based on such inconsistent data may not be reliable. Data rectification techniques that are widely used for improving the quality of process data may also be applied to inconsistent LCI data, but the unique characteristics of LCI data prohibits the direct use of traditional rectification methods. These characteristics include the lack of information about the underlying reactions, flowsheet, and error structure, as well as the availability of data representing multiple overlapping spatial scales. This article presents the first effort for multiscale rectification of LCI data. This approach is integrated with the computational methods of process LCA, economic input–output LCA, and hybrid LCA. The case study of the life cycle for a caustic soda process shows that gross errors can exist in LCI data, and the developed method can remove the errors so that LCA results become more accurate.

This work represents only an early step in providing much needed rigor to LCA, and many challenges remain. Data rectification works best when information about the flow network, reaction equations, and component flow rates and compositions are available. Such information about reactions and component flows at the economy scale is currently unavailable, making it difficult to apply multiscale LCI data rectification to hybrid LCA and EIOLCA. A practical approach utilized in this article is to apply LCI data rectification to data where uncertainty information and constraints are available, such as at the equipment, process, and value chain scales, followed by applying EIOLCA to the cutoff streams of the network at finer scales. In addition, the representation of LCI data must be standardized for giving the correct results of data rectification. LCI data should be shown in the form of overall and atomic material balances, which makes it easy to examine the satisfaction of conservation laws. Information about energy consumption must include the net-calorie values to permit conversion between mass and energy units. In addition, the chemical composition of fuel must be shown to permit the use of combustion equations as constraints in the data rectification model.

LCI data are usually based on information from a variety of sources, including proprietary data. These data are often averaged to avoid revealing proprietary information. It should be possible to report the variance of the underlying values along with the mean, and many commercial databases are starting to provide such information. This should allow application of the proposed approach with fewer assumptions. LCI data are the foundation of LCA, which is an essential approach for evaluating the broader implications of technology. Such analysis is increasingly crucial for greener and economically feasible decision making. LCA is currently plagued by problems such as the use of arbitrary boundaries and uncertain data. The proposed approach is expected to enhance the quality of LCI data, and should become an integrated part of all LCA studies.

Acknowledgments

This work was financially supported by the National Science Foundation (ECS-0424692 and ECS-0524924).

Literature Cited

1. Burgess AA, Brennan DJ. Application of life cycle assessment to chemical processes. *Chem Eng Sci*. 2001;56:2589–2604.
2. Azapagic A. Life cycle assessment and its application to process selection, design and optimization. *Chem Eng J*. 1999;73:1–21.
3. Skone TJ, Curran MA. LCAccess—global directory of LCI resources. *J Cleaner Prod*. 2005;13:1345–1350. Available at http://www.epa.gov/ORD/NRMRL/lcaccess/.
4. NREL. U.S. Life-cycle inventory database. Available at http://www.nrel.gov/lci/. 2007.
5. TRI. Toxics release inventory program. Available at http://www.epa.gov/tri/. 2007.
6. SimaPro, PRé Consultants. Available at http://www.pre.nl/simapro/.
7. GaBi. PE Europe GmbH. Available at http://www.gabi-software.com/. 2007.
8. Miller SA, Theis TL. Comparison of life-cycle inventory databases—a case study using soybean production. *J Ind Ecol*. 2006;1/2:133–147.
9. Peereboom EC, Kleijn R, Lemkowitz S, Lundie S. The influence of inventory data sets on life cycle assessment results: a case study on PVC. *J Ind Ecol*. 1998;2:109–130.
10. Hau JL, Yi HS, Bakshi BR. Enhancing life cycle inventories via reconciliation with the laws of thermodynamics. *Tech. Rep. Ohio State Univ*., 2006.
11. Ayres RU, Life cycle analysis: a critique resources. *Conservat Recycl*. 1995;14:199–223.
12. Narashimhan S, Jordache C. Data Reconciliation and Gross Error Detection. Houston, TX:Gulf Publishing Company, 2000.
13. Romagnoli JA, Sánchez MC. Data Processing and Reconciliation for Chemical Process Operations. San Diego, CA:Academic Press, 2000.
14. Rosenberg J, Mah RSH. Iordache C. Evaluation of schemes for detecting and identifying gross errors in process data. *Ind Eng Chem Res*. 1987;26:555–564.
15. Serth R, Heenan W. Gross error detection and data reconciliation in steam metering systems. *AIChE J*. 1986;32:733–742.
16. Narashimhan S, Mah RSH. Generalized likelihood ratio method for gross error identification. *AIChE J*. 1987;33:1514–1512.
17. Rollins DK, Davis JF. Unbiased estimation of gross errors in process measurements. *AIChE J*. 1992;38:563–572.
18. Sanchez R, Romagnoli J, Jiang Q, Bagajewicz M. Simultaneous estimation of biases and leaks in process plants. *Comput Chem Eng*. 1999;23:841–858.
19. Soderstrom TA, Himmelblau DM, Edgar TFA. Mixed integer optimization approach for simultaneous reconciliation and identification of measurement biases. *Contr Eng Prac*. 2001;9:869–876.
20. Yi HS, Han C. Industrial application of MILP-based simultaneous compensation to a large-scale byproduct gases network in an iron- and steel-making plant. *Ind Eng Chem Res*. 2004;43:119–126.
21. Crowe CM, Campos YAG, Hrymak A. Reconciliation of process flow rates by matrix projection. I. Linear case. *AIChE J*. 1983;29:881–888.
22. Sanchez M, Romagnoli J. Use of orthogonal transformations in data classification-reconciliation. *Comput Chem Eng*. 1996;20:483–493.
23. Crowe C. Data reconciliation—progress and challenges. *J Process Contr*. 1996;6:89–98.
24. Heijungs R, Suh S. The Computational Structure of Life Cycle Assessment. Dordrecht:Kluwer, 2002.
25. Hendrickson C, Horvath A, Joshi S, Lave L. Economic input-output models for environmental life-cycle assessment. *Environ Sci Technol*. 1998;32:184A–191A.
26. Suh S, Lenzen M, Treloar GJ, Hondo H, Horvath A, Huppes G, Jolliet O, Klann U, Krewitt W, Moriguchi Y, Munksgaard J, Norris G. System boundary selection in life-cycle inventories using hybrid approaches. *Environ Sci Technol*. 2004;38:657–664.
27. Rickard JG, Macchietto S, Shah N. Integrated decision support in flexible multipurpose plants. *Comput Chem Eng*. 1999;23:S547–S550.
28. Das BP, Rickard JG, Shah N, Macchietto S. An investigation on integration of aggregate production planning, mater production scheduling and short-term production scheduling of batch process operations through a common data model. *Comput Chem Eng*. 2000;24:1625–1631.
29. Bakshi BR, Bansal P, Nounou MN. Multiscale rectification of random errors without fundamental process models. *Comput Chem Eng*. 1997;21:S1167–S1172.
30. Bakshi BR, Nounou MN, Goel PK, Shen X. Multiscale bayesian rectification of data from linear steady-state and dynamic systems without accurate models. *Ind Eng Chem Res*. 2001;40:261–274.
31. Ungarala S, Bakshi BR. A multiscale bayesian and error-in-variables approaches for linear dynamic data rectification. *Comput Chem Eng*. 2000;24:445–451.
32. Chou KC, Willsky AS, Benveniste A. Multiscale recursive estimation, data fusion, and regularization. *IEEE Trans Automat Contr*. 1994;39:464–478.

33. Power Y, Bahri PA. Integration techniques in intelligent operational management: a review. *Knowl Base Syst*. 2005;18:89–97.
34. Shobrys DE, White DC. Planning, scheduling and control systems: why can they not work together. *Comput Chem Eng*. 2000;24:163–173.
35. Suh S. Functions, commodities and environmental impacts in an ecological economic model. *Ecol Econ*. 2004;48:451–467.
36. Joshi S. Product environmental life-cycle assessment using input-output techniques. *J Ind Ecol*. 2000;3:95–120.
37. Beltran C, Heredia FJ. Unit commitment by augmented lagrangian relaxation: testing two decomposition approaches. *J Optim Theor Appl*. 2002;112:295–314.
38. Romagnoli J. On data reconciliation: constraints processing and treatment of bias. *Chem Eng Sci*. 1983;38:1107–1117.
39. Yi HS, Bakshi BR. Data reconciliation by decomposition methods and constraints processing. Technical Report,The Ohio State University, 2005.
40. Szargut J, Morris DR, Steward FR. Exergy Analysis of Thermal, Chemical and Metallurgical Processes. New York:Hemisphere, 1988.
41. ENERGETICS.Energy and environmental profile of the U. S. chemical industry,Chapter 6: The Chlor-Alkali Industry,U.S. Department of Energy, 2000. Available at http://www.eere.energy.gov/industry/chemicals/tools.html.

Manuscript received July 21, 2006, and revision received Dec. 16, 2006.