A Multiscale Bayesian Framework for Designing Efficient and Sustainable Industrial Systems

Nandan U. Ukidwe and Bhavik R. Bakshi
The Ohio State University
Columbus, Ohio 43210

Gautham Parthasarathy
Solutia Inc.
Springfield, Massachusetts 01151

Abstract

Designing sustainable and efficient businesses requires techniques that can provide a holistic system-wide perspective. Such system may refer to a manufacturing plant or the life cycle of the business activity or enterprise. A plant-wide assessment (PWA) has to deal with different processing units on-site and their interactions with the objectives that include identification of process integration opportunities and avenues for capacity expansion. Similarly, a life cycle-wide assessment (LCA) has to deal with the business in question, and its demand and supply chains. The primary motivation behind doing a life-cycle assessment is to incorporate environmental considerations in product and process design and retrofit. System-wide assessments such as PWA and LCA often have to deal with multiple scales of operation and information. For instance, PWA has to consider a coarser plant-scale comprising various processes on-site and a finer equipment scale comprising specific unit operations within those processes. Similarly, LCA has to consider a coarser economy-scale comprising various industry sectors and a finer plant-scale comprising specific businesses within those industry sectors. Integration across scales is essential for efficient and sensible decision making. A decision maker who sets out with the aforementioned multiscale and systems perspective has to deal with data and models at different scales that vary substantially in their quality, subjectivity and completeness. Moreover, data and models at each scale may not be mutually compatible. To take prudent decisions in system-wide assessment it is necessary to systematically fuse all available information into a hybrid dataset that is as complete and internally consistent as possible. Moreover, the dataset must also comply with various constraints. Such hybrid dataset, once compiled, can be used as a basis for system-wide decisions that may range from identification of process bottlenecks to life cycle environmental implications of products and processes. This presentation proposes and applies a new data fusion framework to address the needs of system-wide assessments. The new approach uses a tree representation to connect multiple scales. The framework ensures intra-scale and inter-scale model and data consistency, and easily renders itself to the established multiscale and Bayesian techniques in statistics and systems engineering.

Keywords

Process design, Sustainability, Economic models, Life cycle assessment, Bayesian modeling.

1. Introduction

As businesses continue to realize the tangible and intangible benefits of sustainable development, there is a pressing need for systematic methods for environmentally conscious process engineering. Meeting this need requires expansion of the process engineering boundary beyond the process and its supply chain to consider the entire life cycle of a product or process (Bakshi and Fiksel, 2003). Environmental life cycle assessment (LCA) has become a popular technique for considering the "cradle-to-grave" resource consumption and emissions of a product or process. It is a systematic way of collecting and analyzing information about various stages of the life cycle including, resource extraction, manufacturing, use, and final disposal or recycling. Besides assessment, LCA can also play a crucial role in environmentally conscious process design (ECPD) by quantifying the environmental objective (Burgess and Brennan, 2001). Consequently, better methods for obtaining and assessing the life cycle inventory (LCI) are essential for improved LCA and ECPD.

Since life cycles are large and complex networks of interconnected systems, it is virtually impossible to capture them accurately. The most common type of LCA focuses on the most important processes, but the use of such an, often arbitrary, boundary can introduce significant errors in the LCA results (Lave et al., 1995). LCA based on economic data considers the entire economic network, but the data are highly aggregated. Detailed engineering knowledge is another source of data. However, this extremely detailed equipment scale data is usually computationally prohibitive for LCA.

Available LCI are typically at different levels of aggregation and represent multiple spatial scales. *Equipment scale* data are at the *finest* scale, and the most accurate. *Life cycle scale* data are at an *intermediate* scale and represent averages of equipment scale data. *Economy scale* data are at the *coarsest* scale, and represent averages over industries in a sector. Recent Hybrid LCA methods attempt to combine the comprehensiveness of EIO-LCA with the greater detail and accuracy of Process LCA (Joshi, 2000). However, such integration of data at widely different scales and with vastly different levels of uncertainty needs to be done carefully, if meaningful results are to be expected. Currently, no systematic framework exists for addressing the challenges of using multiscale data for obtaining the LCI for ECPD.

This paper develops such a framework by treating LCA as a multiscale statistical data fusion problem. It relies on the latest advances in deterministic and stochastic multiscale methods to develop a systematic framework for obtaining the life cycle inventory of products and processes. This approach uses a tree representation to connect data and models at multiple scales. The proposed framework ensures satisfaction of intra-scale and inter-scale model and data consistency, and easily renders itself to both, deterministic and stochastic formulations. The rest of this paper provides a brief background of LCI, and describes the MSLCI methodology.

2. Existing Methods for Life Cycle Inventory

2.1. Economy Scale

Data at this scale represent the entire economy or interaction between several economic sectors. Economic Input-Output data are compiled for many countries and contain information about monetary exchanges between multiple sectors. National statistics also provide information about resource use and emissions for these sectors. Such data form the basis of Economic Input-Output LCA (EIO-LCA) (Lave et al., 1995). An important advantage of this approach is that it considers the entire national economy and avoids defining an arbitrary boundary around selected processes. However, data representing each sector are significantly aggregated to maintain computational tractability. Thus, EIO-LCA uses a relatively complete network, but at a coarse scale.

2.2. Life Cycle Scale

This is the most popular scale for LCA. It focuses on the most important processes in a life cycle, and relies on detailed inventory about the inputs and emissions of the selected processes. Extensive databases of life cycle inventory continue to be compiled (Curran, 1996). Such Life Cycle Inventories usually represent average industry numbers for a selected geographical region and product or manufacturing process. Thus, data at the life cycle scale are more detailed than data at the economy scale, but fail to capture details about an individual process or equipment. Despite its popularity and standardization, the biggest shortcoming of Process LCA is that its results depend on the selected boundary. Consequently, it is not difficult for different users to obtain different LCI results for the same product.

2.3. Equipment Scale

Process engineering knowledge and tools enable the development of detailed and relatively accurate models of individual equipment and flowsheets at this fine scale. Public domain studies are also available for most processes and reasonably accurate simulation is possible via software packages. The benefits of using such information for LCA has been identified and “gate to gate” inventory models for many chemical equipment and processes are being developed (Jiménez-González et al., 2001). Although performing LCA at this fine scale is practically infeasible, the available data could be utilized in LCA studies for environmentally conscious design and manufacturing. Unfortunately, most LCA studies tend to ignore this trove of information.

2.4. Hybrid LCI

Having realized the shortcomings of existing methods, many researchers are combining the best features of LCA at different scales. Most efforts are hybridizing the comprehensiveness of the economy scale with the greater detail of the life cycle scale (Joshi, 2000). Many variations of hybrid LCA methods have been developed. Examples of such methods include the use of Process LCA to compensate for the absence of some sectors such as, the use phase in EIO-LCA; or the use of EIO-LCA at the boundary of Process LCA; or disaggregation of existing economic sectors to include more detailed information. Most hybrid LCA methods usually do not utilize equipment scale information. Moreover, integrated hybrid analysis is completely deterministic in nature, and cannot incorporate stochastic information and subjective and expert knowledge. The multiscale framework proposed in the next section can overcome many of these shortcomings.

3. Multiscale Statistical Framework for LCI

Since LCI information is available at multiple scales, a rigorous framework may be developed by treating it as a multiscale statistical data fusion problem. The multiple scales may correspond to the finest equipment scale, coarser plant or life cycle scales and the coarsest economy scale. Such a framework is described in this section for deterministic and stochastic multiscale LCI (MSLCI).

3.1. Deterministic MSLCI

Deterministic multiscale LCI may be formulated as the following data fusion and inference task.

Given: (1) Data at multiple scales represented by the set of life cycle state variables, $\mathbf{Y}_m$, m=0,…,L with m = 0 representing the finest scale (equipment), and m=L representing the coarsest scale (economy).
(2) Models relating the data at each scale (intra-scale models), $\Phi_m(\mathbf{Y}_m) = 0$

Determine: Estimated values of life cycle state variables for system at selected scale, m, based on knowledge at all scales, $\mathbf{Y}_{m,0:L}$.

Ideally, the estimated state variables obtained by combining data at multiple scales should be consistent with physical laws and maximally utilize available data and models at all scales. Since the assessment of individual processes or life cycles at any scale involves analysis of networks, each intra-scale model is represented by network algebra equations of the same general form. Assuming static networks, the total output of the nodes of a network at scale m, $\mathbf{x}_m$, is related to the output leaving the network, $\mathbf{f}_m$, and the interactions between the nodes, $\mathbf{A}_m$ as,

$$\mathbf{x}_m = [\mathbf{I}_m - \mathbf{A}_m]^{-1}\mathbf{f}_m = \mathbf{T}_m\mathbf{f}_m \tag{1}$$

The emissions from this network may be determined as, $\mathbf{e}_m = \mathbf{R}_m\mathbf{T}_m\mathbf{f}_m$, where $\mathbf{R}_m$ represents the emissions per unit of network output. For a network of n_m nodes, $\mathbf{x}_m$ and $\mathbf{f}_m$ are $n_m \times 1$ vectors, while $\mathbf{A}_m$, $\mathbf{T}_m$, and $\mathbf{R}_m$ are $n_m \times n_m$ matrices. Equation (1) is commonly encountered in network algebra, including economic input-output (EIO) analysis, and would be available at scales, m = 0, …, L. The set of LCI state variables, $\mathbf{Y}_m$ used in the problem formulation and in Equation (1) represents all the variables in the previous equations, that is,

$$\mathbf{Y}_m = \{\mathbf{x}_m, \mathbf{f}_m, \mathbf{e}_m, \mathbf{A}_m, \mathbf{R}_m, \mathbf{T}_m\} \tag{2}$$

The available data and MSLCI problem may be represented as grids at different resolutions, which are connected by edges, as shown in Figure 1. The approach for fusing the available data in a deterministic setting consists of two steps that go up and down the tree.

Fine-to-Coarse (FtC): Identify coarser scale nodes that are relevant to the selected node by propagating information about inputs and outputs from finer to coarser scales.
Coarse-to-Fine (CtF): Refine relevant coarse scale sectors identified in FtC to finer scale by using available finer scale data.

The FtC step is the information-gathering step, while CtF is the computation step. For example, consider the need for life cycle inventory relevant to equipment represented by unit, $S_{3(0)}$, shown lightly shaded in Figure 1. The FtC step would identify the coarser light shaded node, $S_{2(1)}$ as the parent, and obtain coarser scale information for all the inputs and outputs of the equipment, $S_{3(0)}$. This life cycle information would get propagated to the coarser scale of the economy to identify the relevant economic sector, shown as $S_{1(2)}$ in Figure 1.

$$Y_{1(2),1(2)} = Y_{1(1),1(1)} + Y_{1(1),2(1)} + Y_{2(1),(1)} + Y_{2(1),(1)} \tag{3}$$
$$Y_{1(2),(2)} = Y_{1(1),(2)} + Y_{2(1),(2)} \tag{4}$$
$$Y_{2(2),(2)} = Y_{2(2)1(1)} + Y_{2(2),(1)} \tag{5}$$

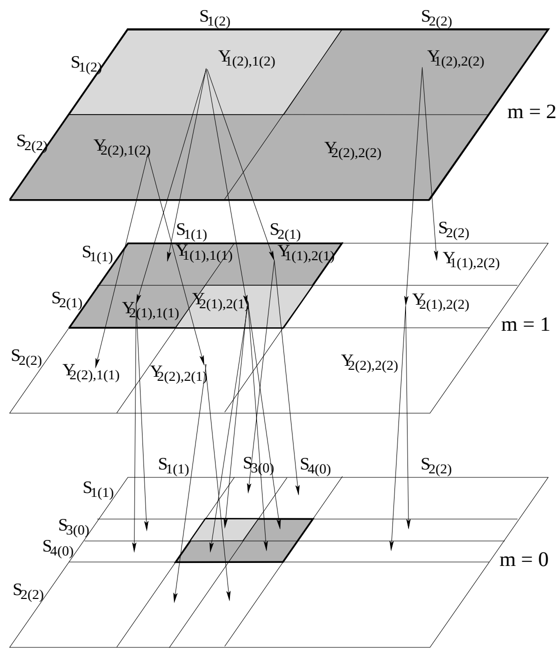

Figure 1. Tree representation of MSLCI.

In the CtF step, the coarsest scale sector identified in the FtC step, $S_{1(2)}$, is refined by using finer scale data and ensuring satisfaction of these conservation equations. These three equalities relate eight variables, indicating the need for additional information about the interaction of the refined sector with other sectors at the coarser scale. Iterative calculation between the coarse and fine scales may be required to reconcile data and models at all scales and to ensure consistency with physical principles such as conservation. These inter-scale models are represented by arrows connecting the scales in Figure 1. The grid at scale $m = 1$ is obtained by refining Sector $S_{1(2)}$ to $S_{1(1)}$ and $S_{2(1)}$ which requires satisfaction of the above balances along with the input-output models at each scale. Refining the shaded grids by combining data at multiple scales results in mixing of scales, as indicated by rectangles of different sizes at $m = 0$ and $m = 1$. The result of the MSLCI approach is represented by the grid at $m = 0$, which fuses the data at all scales.

3.2. Illustrative Example for deterministic MSLCI

This example illustrates the application of deterministic MSLCI approach. The case study considers following two scales

- *Coarser Plant Scale* ($m = 1$) comprising two processes, namely a polymer manufacturing process $PA_{(1)}$ and an on-site power plant, $PP_{(1)}$
- *Finer Equipment Scale* ($m = 0$) comprising two unit operations, namely a reactor, $R_{(0)}$ and a distillation column, $DC_{(0)}$, that together constitute $PA_{(1)}$.

Data at individual scales is available in the form of transaction matrices, $\mathbf{Y}_{(0)}$ and $\mathbf{Y}_{(1)}$, that could be in material or energy units. The flowsheet showing interactions between $PP_{(1)}$, $R_{(0)}$ and $DC_{(0)}$ is shown in Figure 2, whereas Equations (6) and (7) show transaction matrices for the equipment- and plant-scales respectively.

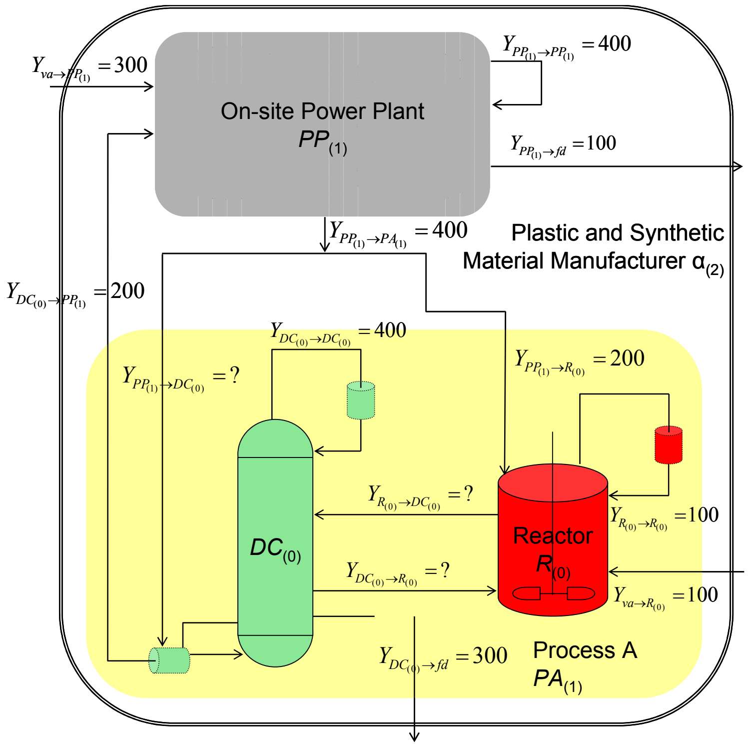

Figure 2. Flow diagram at scales m = 0 and m = 1

$$\mathbf{Y}_{(0)} = \begin{array}{c|cccc} & R_{(0)} & DC_{(0)} & f & Total \\ R_{(0)} & 100 & ? & 0 & ? \\ DC_{(0)} & ? & 400 & 500 & ? \\ va & 300 & ? & & \\ Total & ? & ? & & \end{array} \quad (6)$$

$$\mathbf{Y}_{(1)} = \begin{array}{c|cccc} & PP_{(1)} & PA_{(1)} & f & Total \\ PP_{(1)} & 400 & 400 & 100 & 900 \\ PA_{(1)} & 200 & 1000 & 300 & 1500 \\ va & 300 & 100 & & \\ Total & 900 & 1500 & & \end{array} \quad (7)$$

As seen from Equation (6) certain data points in $\mathbf{Y}_{(0)}$ are missing because of lack of measurement. These data points cannot be estimated by solving balance equations at m = 1 alone as there are more unknowns than equations. To estimate missing data points at m = 1, the inter-scale relations defined by Equations (3)-(5) need to be used. The corresponding algorithm resembles a W-cycle in multigrid analysis. The reconciled transaction matrix $\mathbf{Y}_{(0)}$ can then be fused with $\mathbf{Y}_{(1)}$ to obtain a hybrid transaction matrix $\mathbf{Y}_{hybrid}$ shown in Equation (8).

$$
\mathbf{Y}_{\text{hybrid}} = \begin{array}{c|ccccc} & R_{(0)} & DC_{(0)} & PP_{(1)} & f & Total \\ R_{(0)} & 100 & 400 & 0 & 0 & 500 \\ DC_{(0)} & 100 & 400 & 200 & 300 & 1000 \\ PP_{(1)} & 200 & 100 & 400 & 200 & 900 \\ va & 100 & 100 & 300 & & \\ Total & 500 & 1000 & 900 & & \end{array} \quad (8)
$$

$\mathbf{Y}_{\text{hybrid}}$ combines all available information from the two scales and complies with inter- and intra-scale balance constraints. $\mathbf{Y}_{\text{hybrid}}$ represents deterministic MSLCI over Equipment- and Plant-scales. Similar matrices can be compiled at coarser life cycle- and economy-scales as well. These matrices have many potential applications for environmentally conscious process design, and can be easily extended to include stochastic information. For instance, the hybrid transaction matrices can be used to propagate the effect of a broad policy decision (e.g. imposition of pollution tax or capacity expansion target) on fine-scale design variables and to determine the effect of fine-scale process modifications on broader sustainability objectives. Following paragraph discusses theoretical aspects of stochastic MSLCI.

3.3. Stochastic MSLCI

One of the shortcomings of conventional LCA is its inability to evaluate confidence bounds on LCA results. As a result, the practitioner of LCA can get very different results depending on how the LCA problem is set up. This, at the least, undermines the credibility of LCA and prevents its wide-spread use as a rigorous decision-support tool. An important and attractive feature of the deterministic MSLCI approach described above is that it can readily handle stochastic information, and utilize the powerful approach of Bayesian hierarchical modeling (Wikle, 2003). The stochastic problem formulation treats all variables to be random and all data to be contaminated by noise or errors. Each of the variables used in the deterministic formulation, $\mathbf{Y}_m$ are now considered to represent the noisy versions of the "true" or underlying but unknown values, $\tilde{\mathbf{Y}}_m$, which need to be estimated. Variables, $\tilde{\mathbf{Y}}_m$, are equivalent to the "state variables" in the jargon of stochastic systems, and $\mathbf{Y}_m$ represents the "measured variables". The deterministic input-output models at each scale and those relating different scales may be represented, in general, as the following multiscale stochastic models. These models capture all the available inter- and intra-scale information along with various kinds of uncertainties.

$$
\tilde{\mathbf{Y}}_m = h_m\left(\tilde{\mathbf{Y}}_{m+1}, \omega_m\right) \quad (9)
$$

$$
\mathbf{Y}_m = g_m\left(\tilde{\mathbf{Y}}_m, \nu_m\right) \quad (10)
$$

The function, $h_m(.)$ represents the inter-scale model, while $g_m(.)$ represents the intra-scale relationship between the measured and underlying variables and the model at each scale. Thus, Equation (9) is the stochastic counterpart of Equations (3), (4), and (5), and is expected to be linear, while Equation (10) is the stochastic counterpart of Equation (1), and may be nonlinear because of the product term. The variables, ω_m and ν_m represent the uncertainty in the data and models, respectively. Equations (9) and (10) are in the standard form commonly encountered in nonlinear estimation problems (Chen et al., 2004). Due to recent advances in statistics such as Markov Chain Monte Carlo methods and particle filtering, it is becoming possible to find the Bayesian solution to these equations in an efficient manner. The proposed approach for solving the Bayesian MSLCI problem will aim to obtain properties of the posterior distribution which may be written in a scale recursive form via Bayes rule as,

$$P\left(\tilde{\mathbf{Y}}_0 \mid \mathbf{Y}_{0:L}\right) = \frac{P\left(\mathbf{Y}_0 \mid \tilde{\mathbf{Y}}_0\right) P\left(\tilde{\mathbf{Y}}_0 \mid \mathbf{Y}_{1:L}\right)}{P\left(\mathbf{Y}_0 \mid \mathbf{Y}_{1:L}\right)} \tag{11}$$

The left-hand side of Equation (11) is the posterior probability of the unknown quantities at the finest scale, given information at all scales. Equation (11) shows that this posterior probability may be determined from the data at scale, m = 0, represented by the likelihood, $P\left(\mathbf{Y}_0 \mid \tilde{\mathbf{Y}}_0\right)$, and the information about the unknown quantity available from data and models at other scales, as represented by the prior, $P\left(\tilde{\mathbf{Y}}_0 \mid \mathbf{Y}_{1:L}\right)$. The posterior represents the distribution of the unknown quantities based on capturing all the data and models at all scales. Equation (11) represents recursion from the parent node to the children or the coarse to fine (CtF) change of the posterior. A similar equation may be written for the opposite direction or fine to coarse (FtC) change. These steps may be combined in different ways depending on the nature of the problem to result in a stochastic multiscale algorithm for life cycle inventory.

If Equations (9) and (10) are both linear with additive noise and Gaussian distributions, the posterior will also be Gaussian, and the above algorithm will become multiscale Kalman filtering (Chou et al., 1994). However, in general, due to the product terms in Equation (1), the state equation is not likely to be linear. In this case, the distributions will be determined via hierarchical methods based on MCMC methods. The multiscale approach described in this section will focus primarily on fusing life cycle inventory information at multiple scales.

3.4. Illustrative Example for stochastic MSLCI

This example builds upon the illustrative example presented in Section 3.2 by adding life cycle and economy scales. Thus $R_{(0)}$, $DC_{(0)}$ and $PP_{(1)}$ together comprise plastic and synthetic material manufacturer $\alpha_{(2)}$. $\beta_{(2)}$ represents basic organic chemical manufacturer in the supply chain of $\alpha_{(2)}$. $PR_{(3)}$ represents the petroleum refining sector at the coarsest economy-scale, m = 3. For this system, the stochastic transaction matrix is shown in Figure 3 along with CO_2-emission distributions for each of the system units. For this example all distributions are assumed to be Gaussian with their means corresponding to the values in the deterministic version. The stochastic components arise on account of various types of errors such as measurement error, aggregation error, sampling error and subset error. Figure 3 also shows error estimates in embodied CO_2 in the outputs from the reactor, $R_{(0)}$, and the distillation column, $DC_{(0)}$, due to arbitrary truncation of the system boundary. A narrow plant-scale system boundary leads to 89.9% and 88.5% truncation errors in embodied CO_2 in the reactor and distillation column outputs respectively. A broader life-cycle scale system boundary reduces these errors to 60.3% and 62.6% respectively, and the broadest economy-scale system boundary eliminates the truncation error altogether as it utilizes all available information. However, other types of errors such as measurement error, subset error, sampling error etc. still exist at all scales and can be estimated in a similar manner. Such stochastic analysis can assist designing of industrial systems under uncertain policy scenarios, changing socio-economic conditions and environmental regulations.

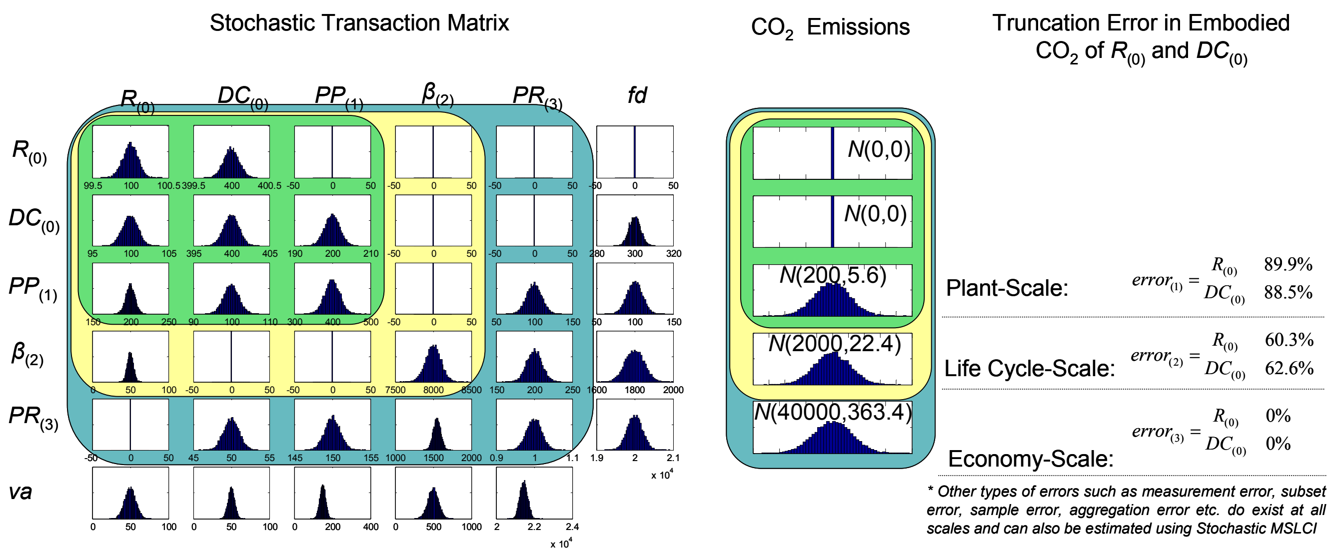

Figure 3. Illustrative example for stochastic MSLCI

4. Conclusions

This paper proposes a new framework for obtaining environmental life cycle inventory information by utilizing data and models at multiple spatial scales. Such information is essential for environmentally conscious process design and for improving the sustainability of industrial activities. The proposed framework fuses data at scales of individual equipment, process life cycle, and economy. Information about different kinds of uncertainties may be readily incorporated in this approach as well.

5. References

Bakshi, B. R. and J. Fiksel (2003). The quest for sustainability: Challenges for process systems engineering. *AIChE Journal*, **49**, 6, 1350-1358

Burgess, A. A. and D. J. Brennan (2001). Application of life cycle assessment to chemical processes, *Chem. Eng. Sci.*, **56**, 2589-2604

Chen, W.-S., B. R. Bakshi, P. K. Goel, and S. Ungarala (2004). Bayesian rectification of unconstrained nonlinear dynamic systems by sequential Monte Carlo sampling, *Ind. Eng. Chem. Res.*, submitted.

Chou, K.C., A.S. Wilsky and A. Benveniste (1994). Multiscale recursive estimation, data fusion and regularization. *IEEE Trans. Automat. Contr.*, **39**, 3, 464-478

Curran, M. A. (1996). *Environmental life-cycle assessment*, McGraw-Hill, New York

Fava J.A., R. Denision, B. Jones, M.A. Curran, B. Vigon, S. Selke and J. Barnum (1991). *A technical framework for Life Cycle Assessment*. SETAC, Washington D.C.

Jiménez-González C., M. R. Overcash and A. Curzons. (2001). Waste treatment modules – a partial life cycle inventory, *J. Chem. Tech. Biotech.*, **76**, 7, 707-716

Joshi, S. (2000). Product Environmental Life-Cycle Assessment Using Input-Output Techniques, *J. Ind. Ecol.*, **3**, 2-3, 95-120

Lave, L., E. Cobas-Flores, C. Hendrickson and F. McMichael (1995). Using Input-Output Analysis to estimate economy wide discharges. *Env. Sci. Tech.*, **29**, 9, 420

C. K. Wikle (2003). Hierarchical models in environmental science. *Int. Stat. Rev.*, **71**, 2, 181-199