A MULTISCALE BAYESIAN FRAMEWORK FOR ENVIRONMENTALLY CONSCIOUS PROCESS DESIGN

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

Abstract

Improving the sustainability of chemical processes requires techniques for including life cycle environmental considerations in product and process design. Data for life cycle assessment are usually available at different spatial resolutions such as, the economy, supply chain and equipment. These data vary in their degree of completeness in capturing the life cycle, and their level of uncertainty. For example, equipment scale data are relatively accurate, but narrow in scope, while economy scale data capture the economic network, but are aggregated. This paper presents a new framework for using all kinds of data and models for obtaining the life cycle inventory of a selected product or process. Such information is essential for assessing the life cycle impact of process alternatives and for incorporating environmental aspects in process design. The proposed multiscale framework is represented as a tree-based model and can be used in deterministic or stochastic formulations. The stochastic formulation provides the probability distribution for the life cycle state variables via a hierarchical Bayesian data fusion algorithm. Case studies illustrating the framework are in progress.

Keywords

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

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 product and process design (Bakshi and Fiksel, 2003). Environmental life cycle assessment (LCA) has become a popular technique for considering the "cradle-to-grave" resource consumption, emissions and their impacts 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 a large and complex network 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 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 models 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. Case studies illustrating the approach are in progress and will be presented at the conference.

Existing Methods for Life Cycle Inventory

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 exchange 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.

Life Cycle Scale

This is the most popular scale for LCA, and focuses on the most important processes in a life cycle. The resulting Process LCA relies on detailed inventory about the inputs and emissions of the selected processes, and extensive databases of life cycle inventory continue to be compiled (Curran, 1996). These LCI 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.

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 and for environmentally conscious design and manufacturing. Unfortunately, most LCA studies tend to ignore this trove of information.

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.

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 life cycle scale and the coarsest economy scale. Such a framework is described in this section for deterministic and stochastic multiscale LCI (MSLCI).

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,

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

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 three equations, that is,

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

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.

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.

$$Y_{1(2),1(2)} = Y_{1(1),1(1)} + Y_{1(1),2(1)} + Y_{2(1),(1)} + Y_{2(1),(1)} \qquad (3)$$

$$Y_{1(2),(2)} = Y_{1(1),(2)} + Y_{2(1),(2)} \qquad (4)$$

$$Y_{2(2),(2)} = Y_{2(2)1(1)} + Y_{2(2),(1)} \qquad (5)$$

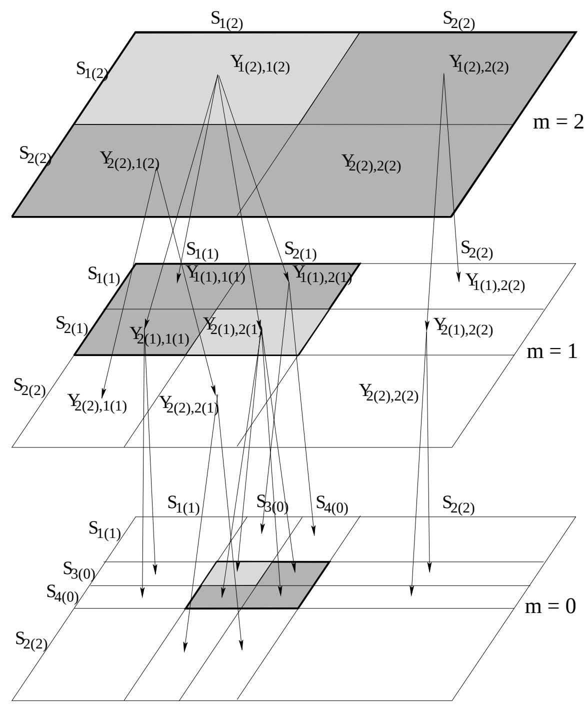

Figure 1. Tree **r***epresentation of MSLCI.*

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 the 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.

Illustrative Case Study

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

- *Coarser Economy Scale* (m = 2) comprising two sectors, namely Electric Power Generation (PG) and the Plastics and Synthetic Materials (SM)
- *Finer Life Cycle Scale* (m = 1) comprising two processes, namely a plastic manufacturer α and the residual Plastics and Synthetic Materials Sector (SM-α)

Data at individual scales is available in the form of monetary transaction matrices $\mathbf{Y}_{(2)}$ and $\mathbf{Y}_{(1)}$ as follows.

$$\mathbf{Y}_{(2)} = \begin{array}{c|cccc} & PG & SM & fd & Tot \\ PG & 400 & 400 & 100 & 900 \\ SM & 200 & 1000 & 300 & 1500 \\ va & 300 & 100 & & \\ Tot & 900 & 1500 & & \end{array} \quad \mathbf{Y}_{(1)} = \begin{array}{c|cccc} & \alpha & SM-\alpha & fd & Tot \\ \alpha & 100 & ? & 0 & ? \\ SM-\alpha & ? & 400 & 500 & ? \\ va & 300 & ? & & \\ Tot & ? & ? & & \end{array} \tag{6}$$

As seen from Equation (6) certain data points in $\mathbf{Y}_{(1)}$ 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, it is necessary to use inter-scale relations defined by Equations (3)-(5) and the corresponding algorithm resembles a W-cycle in multigrid analysis. The reconciled transaction matrix $\mathbf{Y}_{(1)}$ can then be fused with $\mathbf{Y}_{(2)}$ to obtain a hybrid transaction matrix $\mathbf{Y}_{\text{hybrid}}$ as shown below.

$$\mathbf{Y}_{\text{hybrid}} = \begin{array}{c|ccccc} & PG_{(2)} & \alpha_{(1)} & SM-\alpha_{(1)} & fd & Tot \\ PG_{(2)} & 400 & 200 & 200 & 100 & 900 \\ \alpha_{(1)} & 0 & 100 & 400 & 0 & 500 \\ SM-\alpha_{(1)} & 200 & 100 & 400 & 300 & 1000 \\ va & 300 & 100 & 0 & & \\ Tot & 900 & 500 & 1000 & & \end{array} \tag{7}$$

$\mathbf{Y}_{\text{hybrid}}$ 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 Economy and Life Cycle scales and is useful in environmentally conscious process design.

Stochastic MSLCI

An important and attractive feature of the deterministic multiscale LCI approach described above is that it can readily handle stochastic variables, and utilize the powerful approach of Bayesian hierarchical modeling (Lauritzen, 1996; 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 uncertainty.

$$\tilde{\mathbf{Y}}_m = h_m\left(\tilde{\mathbf{Y}}_{m+1}, \omega_m\right) \tag{8}$$

$$\mathbf{Y}_m = g_m\left(\tilde{\mathbf{Y}}_m, \nu_m\right) \tag{9}$$

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 (8) is the stochastic counterpart of Equations (3), (4), and (5), and is expected to be linear, while Equation (9) 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 (6) and (7) 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,

$$P\left(\tilde{\mathbf{Y}}_0 \mid \mathbf{Y}_{0:L}\right) = \frac{P\left(\mathbf{Y}_0 \mid \tilde{\mathbf{Y}}_0\right) P\left(\tilde{\mathbf{Y}}_0 \mid \mathbf{Y}_{1:L}\right)}{P\left(\mathbf{Y}_0 \mid \mathbf{Y}_{1:L}\right)} \tag{10}$$

The left-hand side of Equation (10) is the posterior probability of the unknown quantities at the finest scale, given information at all scales. Equation (10) shows that this posterior probability may be determined from information in 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 (10) 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 (8) and (9) 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.

Conclusions

This paper proposed 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 via a hierarchical Bayesian framework. Information about different kinds of uncertainties may be readily incorporated in this approach.

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