Bridging data gaps in environmental assessments: Modeling impacts of fine and basic chemical production†

Gregor Wernet,*[a] Stavros Papadokonstantakis,[a] Stefanie Hellweg[b] and Konrad Hungerbühler[a]

Received 19th March 2009, Accepted 7th August 2009
First published as an Advance Article on the web 7th September 2009
DOI: 10.1039/b905558d

The chemical industry is increasing its efforts to reduce the environmental burdens of chemical production. One focus is to implement energy-efficient processes and green technologies early in the process design to maximize environmental efficiency and to reduce costs. However, as data on many chemical products are scarce, many sustainability studies are hampered by the lack of information on production processes, and chemicals are often neglected or only crudely estimated. Models that estimate production data and environmental burdens can be vital tools to aid sustainability efforts. In addition, they are useful for the environmental assessment of chemicals without access to production data, *i.e.* in supply-chain management or for the assessment of products using chemicals as materials. Using mass and energy flow data on the petrochemical production of 338 chemicals, we developed models that can estimate key production parameters directly from the molecular structure. The data sources were mostly production data provided by industrial partners, extended by data from the ecoinvent database. The predicted parameters were the Cumulative Energy Demand (CED), the Global Warming Potential (GWP), the Eco-indicator 99 score, a Life Cycle Assessment (LCA) method, and the electricity and heat use over the production cycle. Model outputs include a measure of the prediction uncertainty. The median relative errors of the models were between 10% and 30%, within acceptable ranges for estimations. The modelled parameters offer a thorough insight into the environmental performance of a production process and the model estimates can be of great service in process design, supply-chain management and environmental assessments of chemical products in the early planning and design stages where production data are not available.

Introduction

The optimal time to minimize the environmental burdens of chemical production is in the early planning and design stages. In the later stages, when the processes are actually implemented, improvements are much more costly and time-consuming to implement. Later changes to synthesis approaches are generally prohibitively expensive, and end-of-pipe improvements are expensive and less efficient than planning for sustainable production. Therefore, the sooner in the planning phase information about the sustainability of a planned process can be acquired, the more helpful it is. Life Cycle Assessment (LCA)[1,2] is a method to determine the environmental impacts of processes and products and has been used extensively to assess chemical production.[3-6] LCA results can be very helpful in process design. Supply chain management and assessment is another important application. However, inventory data of chemicals, *i.e.* emissions and resource uses throughout the production life cycle, are scarce, mostly due to missing information or confidentiality issues. Thus, they often have to be modeled or estimated.[7,8]

In order to fill data gaps, inventory models for processes have been put forward,[4,9] as well as tools to model process parameters, such as energy use.[10] Inventory databases can be help in cases of missing process data and can be used to supply upstream data on auxiliaries and energy production processes. For instance, publicly available databases such as the ecoinvent database[11,12] or national LCI initiatives (*e.g.*[13]) are used. Moreover, several chemical companies are using LCA to assess their products[14,15] and have created internal databases to support their decision-making. Other companies use eco-efficiency analysis, an LCA-based tool to carry out comparative assessment of products and processes.[16] However, process-based inventory modeling often suffers from large uncertainties when little or no information is available. This may lead to severe estimation errors. Models that are not process-based circumvent these problems. Such models are the molecular-structure-based models (MSMs).[17] These models based on neural networks (NNs)[18-20] can estimate key inventory parameters and environmental impacts over the whole production cycle using only molecular features as input data. The results are cradle-to-gate inventory data or Life Cycle Impact Assessment (LCIA) results. Their functionality and

[a]*ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093, Zurich, Switzerland. E-mail: gregor.wernet@chem.ethz.ch; Fax: +41 44 632 1189; Tel: +41 44 632 3471*
[b]*ETH Zurich, Schafmattstrasse 6, 8093, Zurich, Switzerland*

† Electronic supplementary information (ESI) available: List and definition of all descriptors, training parameters, description of alternative approaches, full models for the CED, GWP and Eco-indicator 99 and guidelines for using them. See DOI: 10.1039/b905558d

potential has been demonstrated[17] using a very basic dataset that limited the resulting models to only basic chemicals.

Neural networks' basic premise is that the interactions among a large number of simple computing elements, called nodes or neurons, arranged in input, hidden and output layers, can effect complex information processing. As universal approximators, neural networks are suitable for handling multivariate systems of any complexity level, while as "black box" models they do not require knowledge of the underlying principles of the system under consideration. Instead, they are extracting the necessary information directly from training data, being tolerant to faults and noise included in them.[18] In training procedure, during which the network learns the relationships between input and output variables, the network parameters are determined and optimized in a way to minimize the errors between the true outputs and the network predictions.[19] While the "black box" structure of neural networks can be a burden when analyzing results, it allows the use of confidential production data. Although they are not new in concept, the interest in them has increased significantly in the last two decades due to evolution of digital computing and limitations of traditional modeling techniques.

Our goal in this work was to create MSMs that can accurately predict inventory parameters and LCIA results for a large variety of chemicals, ranging from basic to advanced and fine chemicals. Such models would be of substantial use to process designers as fast and simple tools to assess the sustainability of different productions, to use screening LCAs as decision-making support and to assess their supply chains. They would also be a significant improvement for industrial ecologists and LCA practitioners.

Methods

To create estimation models that can accurately predict production parameters, the underlying data needed to cover a large amount of chemicals from different classes of chemical structures. We gathered a dataset of 392 complete cradle-to-gate inventory data of organic petrochemical production from several sources. 296 datasets were based directly on industrial production data. These were gathered by industrial partners with direct access to production data and measurements. Most data were cradle-to-gate, but a few were missing data on auxiliaries and upstream processes. In these cases the ecoinvent database was used to complete the production data. These data are highly confidential. However, the nature of neural networks allowed their use to generate the models that render any introspective into the production data impossible. The final 96 inventories were taken from ecoinvent itself. As the models are designed to predict energy consumption, the training data had to include accurate information on energy use. Therefore, only datasets not or minimally relying on estimations of energy use in their production life cycle were included in the training data.

All datasets were based on petrochemical synthesis. Products that included biosynthesis steps in their production were not included in the data. Data quality was judged to be very high, with the industry data based on direct production documentation and measurements and the ecoinvent data based on detailed production data and thorough research of public sources of information. The chemicals in the dataset were diverse, ranging from basic chemicals to more complex and fine chemicals. As the composition of the electricity mix can affect the environmental impacts, all data were based on an average European (UCTE) electricity mix.

While all data were based on high quality sources, some inconsistencies should be expected when combining existing data sources. Different production processes can be one explanation for the differences. Furthermore, some production parameters were still estimated by the data collectors, and differing standards and assumptions led to different results. Cases of multiple datasets for the same chemicals allowed us to quantify these variations and determine parameters with fundamental differences. For training purposes, these datasets were averaged to have only one input per chemical, reducing the number of chemical inventories from 392 to 338. Table 1 offers insight into the dataset composition. The major output selected for the models was the Cumulative Energy Demand (CED).[21] The CED represents the total amount of primary energy potential used during the production life cycle. It is therefore a purely resource-oriented method. Nevertheless it has been shown to correlate well with many other LCIA methods,[22] meaning that the CED may be used as an indicator for other environmental impact methods, the calculation of which might otherwise require additional emission data. In addition, a large fraction of the environmental impacts of chemical production in developed countries can be linked to energy and resource use;[6] for these reasons the CED can serve as a proxy of general environmental impact. Resource-oriented methods also have the advantage that production data are usually of better quality than emission data, which are less well documented. Furthermore, models for the Global Warming Potential[23] were created to address an issue important for many process designers nowadays, as the focus in sustainability continues to shift towards Global Warming. As mass flows can be estimated more easily than energy flows in chemical synthesis, a special effort was undertaken to estimate heat and electricity use separately in the production cycle to aid practitioners with mass flow but no energy data. Many datasets did not offer this information however, so the available dataset for these models was only a

Table 1 Composition of the training dataset of 338 chemical inventories

	Molecular Weight [g/mol]	Number of functional groups†	Cumulative Energy Demand [MJ-eq/kg]	Global Warming Potential 100a [kg CO_2-eq/kg]
Minimum	30	0	30	0.5
Maximum	1200	30	1400	9
Mean	160	3.5	170	7
Median	130	3	100	4

fraction of the total dataset. These models estimate direct heat and electricity use. The Eco-indicator 99[24] method offers an overall estimation of all environmental damages weighed against each other. The weighting allows a direct comparison of different environmental impacts.

However, it is of course a subjective step, so results are biased by the opinions represented by the weighting scheme. Again, only a fraction of the dataset allowed calculation of Eco-indicator 99 results and could be used to create models. In theory, there are many ways to describe molecules. In the present study, the selection was limited to descriptors directly calculable from the molecular structure and occurring with some regularity within the available dataset.

A Multi-Layer Perceptron (MLPs) structure was used to develop the models. Other neural network structures were also tested but did not improve prediction accuracy.† The MLPs were developed with a pre-selection of descriptors based on chemical expertise. 10 descriptors were selected from a larger group to best describe production-relevant molecular features to the models:

- molecular weight
- number of functional groups (total)
- number of oxygen atoms in keto and aldehyde groups
- number of oxygen atoms not in keto and aldehyde groups
- number of nitrogen atoms
- number of halogen atoms
- number of aromatic or aliphatic rings
- number of ternary or quaternary carbon atoms
- number of heteroatoms in rings
- number of unique substitutes on aromatic rings

A large data-to-weights ratio, *i.e.* greater than 8, is recommended by a rule of thumb[18] to ensure the resulting models are capable to generalize well. Smaller ratios may lead to overfitting, an effect where the models are fitted too strongly to the specifics of the training set while their predictive capabilities for other chemicals are inferior. If the ratio drops too low below the threshold, the benefit of having more descriptors to describe the models may be outweighed by the drawbacks of overfitting and prediction accuracy decreases. The trade-off between communicating as much information as possible to the models while minimizing input size was reduced by reducing the input dimensions through Principal Component Analysis (PCA). PCA is a technique to reduce the number of input variables by analyzing the descriptors for correlations and then creating independent factors, which are a linear combination of the original descriptors. These factors still contain most of the information of the original descriptors while being fewer in number.

A PCA was carried out and the 10 descriptors could be reduced to 8 factors, which still accounted for over 95% of the variance of the 10 descriptors. These were then used as model input. The models used 7 hidden neurons.† This resulted in 71 weights in the network and a still acceptable data-to-weights ratio of 4.8. The models were therefore likely to be susceptible to some overfitting, however a test series confirmed that predictive capabilities decreased when reducing the input size. The benefits of more information outweighed the problems of overfitting in this case. For the Eco-indicator 99 models, fewer factors were needed to perform well.† All models were trained with a Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, a quasi-Newton method. For training, the input was randomly separated into a training set and a test set 30 times. Each test set contained 15% of the data, or 51 chemicals in this case. After training, model quality was assessed by calculating the coefficient of determination (q^2) for the model predictions over the test data. The final predictive models were a combination of the best-performing models with each of the 30 test sets. Applying the final models therefore yields 30 results, which are averaged and can be subjected to uncertainty analysis.

For the models using smaller datasets, the procedure was identical.

Results and discussion

Dataset analysis

The multiple inventories that existed for some chemicals allowed for a statistical analysis of the output parameter variation for the outputs that were available from more than one source. Of these, the CED had the lowest variation between sources. On average, the relative standard deviation of the CED is 24% and the median relative standard deviation is 17%. This means that for most of the observed cases, the standard deviation was low compared to the absolute CED value, implying consistency in method among the different datasets. Similarly, the Global Warming Potential shows both a mean and a median relative standard deviation of 40%. These parameters were deemed acceptable for predictive models, although the inherent uncertainty of the training data limits the effectiveness of the GWP models. Some production parameters on aqueous emissions were available but showed strong variations between the different data sources. For example, for the Total Organic Carbon (TOC) emissions the mean relative standard deviation was 113% and the median 139%. In other words, the standard deviation was usually larger than the mean value itself. In addition, the TOC and some other parameters showed a systematic difference between the different data sources, most likely due to different assumptions and models during data gathering in the industry and by ecoinvent. These uncertainties led us to conclude that models based on the available output data were not feasible.

Prediction results

To assess the predictive qualities of the models, the coefficient of determination (q^2) between the test data and the model predictions were calculated. For each output parameter, 30 networks were created and validated using the 30 test sets. Each model was evaluated on the average performance over these networks. Table 2 shows the q^2 results for the different models and outputs.

Table 2 Coefficients of determination for the structure-based models. All values are means over the 30 test sets, standard deviations in parentheses. The size of the available data sets is given as n for each output

CED (n = 338)	0.58 (0.13)
GWP (n = 338)	0.41 (0.23)
Eco-indicator 99 (H/A) (n = 142)	0.69 (0.12)
Electricity Use (n = 250)	0.59 (0.13)
Heat Use (n = 250)	0.47 (0.20)

The coefficients of determination are generally in the range of 0.4 to 0.7. The coefficient of determination is lower for the GWP—this is most likely due to the heterogeneousness of the input set (see dataset analysis). Note that the models based on only fewer data sources (Eco-indicator 99, electricity and heat) suffer less from variations within the data sets. In addition, q^2 scores become more significant as the number of data points increases, so a direct comparison is difficult.

The 30 networks performing best for the 30 test sets were selected to comprise the final models. Therefore, each model for a specific output actually consists of 30 individual networks trained and tested with different combinations of inventory data. Our suggested application is to use all 30 models to acquire a set of predictions which can be averaged to achieve a final prediction. This has the added benefit of providing a standard deviation which can be used as a measure of prediction uncertainty. When all networks agree in their prediction, the likelihood of a wrong prediction is expected to be lower. As an example of the model predictive capabilities the CED predictions are plotted against the inventory data in Fig. 1.

Fig. 1 Relative inventory data and model predictions of the CED for the 338 inventories. The line represents the ideal line where data and predictions match. The mean relative error is 29.1%.

To assess the predictive qualities of the final models, they were applied directly to predict the full, available dataset (338 chemicals, or less depending on the target parameter). With these results, several measures of model quality could be calculated: the absolute and the relative errors as measures of the models' preciseness and the error as a multiple of the standard deviation for the prediction.

The last measure utilizes the fact that each prediction has an individual uncertainty value attached to it—the standard deviation of the 30 individual predictions. By assessing the true error as a multiple of this standard deviation, the reliability of the ranges given by the model could be evaluated. To analyze how these measures can describe the capabilities of the models, the relative errors are plotted in Fig. 2, together with the model predictions and the inventory data for the chemicals. The figure shows that many of the largest relative errors occur due to chemicals with low CEDs being overestimated by the models.

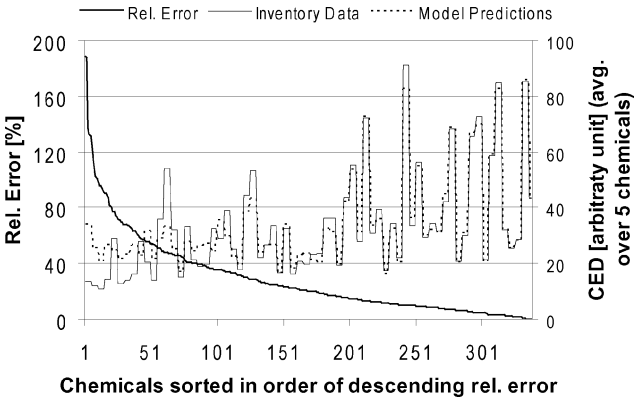

Fig. 2 Final model results for the CED prediction for relative error. These are the results of the final models selected for prediction, applied to the 338 datasets. 95% of the chemicals in the dataset have a relative error of less than 87.5%. For comparison, the secondary axis on the right shows the model predictions compared with the inventory data for the same chemicals.

The full results of all models can be seen in Table 3. The CED models show adequate prediction capabilities and perform well within the requirements of screening LCAs with a mean relative error of 29.1% and a mean absolute error of 36 MJ-eq/kg, given that the models were applied to a data range spanning from 30 to roughly 1400 MJ-eq/kg. Considering the uncertainty of roughly 20% in the original training data, this result is satisfying. In addition, the error in standard deviations is a measure of how well the models' own estimation of their uncertainty describes the true model uncertainty. A range of 1.97 standard deviations plus and minus the model result included the true value in 95% of the cases. The approach presented therefore allows both a prediction of a value and an assessment of the predictive uncertainty. The GWP models perform less well, as could be expected after the lower coefficients of determination and the higher variation in the original data. The Eco-indicator 99 models perform very well, with both low relative errors and a good measure of models uncertainties.

The Electricity and Heat models also perform well. The mean relative errors for both inventory parameters are heavily affected

Table 3 Predictive measures for the final models

	CED	GWP	Eco-indicator 99	Electricity Use	Heat Use
Mean absolute error	36 MJ-eq/kg	2.1 kg CO_2-eq/kg	0.071 Pt	3.0 MJ/kg	18.4 MJ/kg
Median absolute error	24 MJ-eq/kg	1.4 kg CO_2-eq/kg	0.047 Pt	1.9 MJ/kg	12.8 MJ/kg
Mean relative error [%]	29.1	58.2	20.7	94.6	76.4
Median relative error [%]	20.2	29.5	11.5	27.0	23.9
Mean error in standard deviations	0.74	0.66	0.74	0.60	0.71
Median error in standard deviations	0.55	0.50	0.49	0.42	0.42

Table 4 Results for basic and advanced chemicals according to different criteria. All predictions refer to the CED

	Divided by the Molecular Weight	Divided by the number of functional groups	Divided by the CED
Lower Half (Range)	30–126 g/mol	0–3 FG	31-99 MJ-eq/kg
Mean relative error [%]	33.0	29.9	37.8
Median relative error [%]	25.1	23.2	28.6
Mean error in standard deviations	0.83	0.90	0.71
Median error in standard deviations	0.65	0.64	0.58
Higher Half (Range)	126–1200 g/mol	3–30 FG	100–1400 MJ-eq/kg
Mean relative error [%]	25.2	28.3	20.4
Median relative error [%]	14.8	16.9	14.7
Mean error in standard deviations	0.65	0.59	0.78
Median error in standard deviations	0.43	0.46	0.51

by a few data points for chemicals requiring either almost no heat or almost no electricity in their production. The causes are exothermic reactions, where heat and electricity use can be avoided through use of high energy feedstocks, resulting in a very low or even negative heat or electricity use. These are well predicted in absolute terms, but the closeness of the results to zero results in high mean relative errors, much higher than the median relative errors. Furthermore, processes can be combined to utilize excess heat, which is not reflected in the heat use models. Some uncertainties are therefore unavoidable. Nevertheless, the errors for all predictions are within acceptable ranges for screening and design uses, especially given the inherent basic uncertainty of the training data (*e.g.* 24% for the mean relative CED error, see Dataset Analysis).

As seen in Fig. 2, the models can predict most substances in the dataset quite well, while a limited number of chemicals show a larger error in the prediction of the CED.† The model predictions were therefore analyzed to identify the classes of chemicals which were well and not well modeled. Using the CED models as an example, Table 4 shows the errors if the dataset is split into two equally sized halves using the molecular weight and the number of functional groups as measures of complexity. In addition, as the CED itself can be regarded as a measure of the complexity of the production, we assessed the model results for basic and advanced chemicals defined as having a CED below or above 100 MJ-eq/kg respectively (this also separated the dataset in 2 equal halves). There is no common agreement on how to define basic *vs.* fine chemicals but these indicators are all qualified in some degree to be a measure of the complexity of the chemical.

Naturally, the more complex chemicals have higher absolute errors, this is due to the fact that the average CED is higher for complex chemicals regardless of the type of definition of complexity. The relative errors, however, decrease when the models are applied to more complex chemicals as seen in Table 4, a finding supported by the results shown in Fig. 2. The model errors are actually lower for more complex chemicals, while the prediction of very basic chemicals is more difficult. This may be due to the increased use of highly specialized processes for basic, mass-produced chemicals. The restrictive costs of developing specialized processes limit these to mass-produced basic chemicals such as solvents. Fine chemicals on the other hand are often produced in multi-purpose batch-plants, sharing equipment and utilizing standard reactions, which facilitates modeling their production. Furthermore, the models also describe the prediction error more correctly for complex chemicals, and the errors in relation to the standard deviations are usually lower for the complex productions. The models perform best for the classes of chemicals for which production data is scarcest and where models are needed most.

One limitation of the models is that the estimation error may increase if the tested molecule is significantly different than the chemicals in the dataset of 338 chemicals used for this work. Chiral substances for example fall in this category. It is likely that the production requirements for chiral substances are underestimated to some degree using these models. As the models can be easily updated when data become available, this is one area of future work. However, as of yet LCIs of chiral productions are very scarce.

Application example

To demonstrate the use of the models, the CED and GWP models have been applied to 5 chemical productions of organic chemicals. After the descriptors were entered as the input, the calculations ran automatically. The model results, each given as the mean of the 30 individual models and the standard deviation of the 30 results to serve as an uncertainty approximation of the models, are shown in Fig. 3.

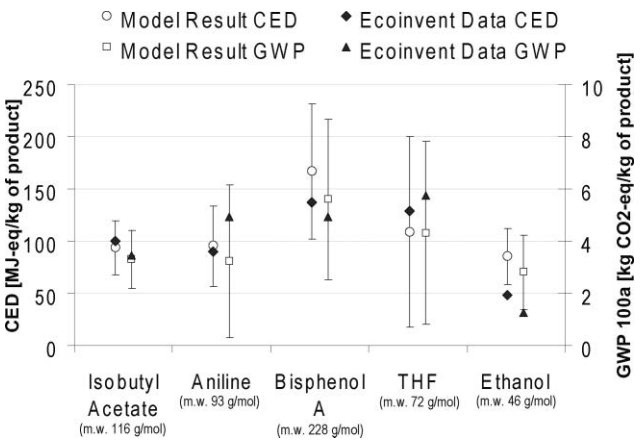

Fig. 3 Sample results for the application of the CED and GWP models. Molecular weights are shown under each chemical. The ecoinvent values are given for comparison. The individual standard deviations cover most of the true values (THF = tetrahydrofuran).

As can be observed, most of the examples are covered by the uncertainty ranges of the model predictions. The results for

ethanol are higher than the ecoinvent values for both the CED and the GWP, and both fall outside of the uncertainty ranges. This is an example of the models' problems with very basic and mass-produced chemicals discussed earlier. However, since basic chemicals with very high production volumes are more likely to be covered by databases such as ecoinvent than more specialized chemicals, the models should not be applied to these chemicals in any case.

For the other chemicals, the predictions are adequately close to the ecoinvent data to be of use for the intended purposes. More importantly, the standard deviations of the model results can serve as a meaningful indicator of the prediction quality. As can be observed, most errors are significantly smaller than the uncertainty ranges provided by the models. In the end, the user may decide how important the inclusion of the true value is for his or her purposes and may scale the ranges accordingly.

Conclusions

We have presented an advanced and thorough method to generate molecular structure-based models for inventory parameters and Life Cycle Impact Assessment results and have analyzed the resulting models. We tested several network structures and optimization strategies for neural networks to generate models that show high prediction capabilities. The models show satisfactory predictive capabilities for the purposes of screening LCA, supply chain management and process design. The models are "black box" type models, which allow the free publication of the models for the CED, GWP and Eco-indicator 99 for public use. The CED models combine the scope of the full dataset and an accurate data basis. As the CED correlates well with several other LCIA methods,[22] this model is especially useful. The three models are available in the supporting information† as well as through direct download (http://www.sust-chem.ethz.ch/tools/finechem).

The models offer both estimations of key parameters and an uncertainty value for the result, allowing an assessment of the prediction accuracy. Our analyses show that the models are especially suited to predicting properties of the more complex productions of advanced and fine chemicals, the areas for which data are needed most. The models can be updated and extended in scope and capabilities as new data become available for training. These models are a significant advancement over the previous options available for sustainability managers, LCA practitioners and process designers and can assist them by providing estimations for the many chemicals on which no accurate production data is available due to lack of process information or confidentiality issues.

Acknowledgements

We thank the Swiss Federal Office for the Environment (project no. 810.3189.004) and the Swiss Federal Office for Energy (project no. 101711) as well as our industrial partners for their support in this project.

Notes and references

1 ISO 14040, European Committee for Standardisation, 2006.
2 A. Azapagic, *Chem. Eng. J.*, 1999, **73**, 1–21.
3 C. Capello, U. Fischer and K. Hungerbühler, *Green Chem.*, 2007, **9**, 927–934.
4 C. Capello, S. Hellweg, B. Badertscher and K. Hungerbühler, *Environ. Sci. Technol.*, 2005, **39**, 5885–5892.
5 D. Kralisch, A. Stark, S. Korsten, G. Kreisel and B. Ondruschka, *Green Chem.*, 2005, **7**, 301–309.
6 G. Wernet, C. Mutel, S. Hellweg and K. Hungerbühler, unpublished results.
7 R. Bretz and P. Frankhauser, *Int. J. Life Cycle Assess.*, 1996, **1**, 139–146.
8 R. Hischier, S. Hellweg, C. Capello and A. Primas, *Int. J. Life Cycle Assess.*, 2005, **10**, 59–67.
9 G. Geisler, T. B. Hofstetter and K. Hungerbühler, *Int. J. Life Cycle Assess.*, 2004, **9**, 101–113.
10 A. Szijjarto, S. Papadokonstadakis, U. Fischer and K. Hungerbühler, *Ind. Eng. Chem. Res.*, 2008, **47**, 7323–7334.
11 R. Frischknecht, N. Jungbluth, H.-J. Althaus, G. Doka, R. Dones, S. Hellweg, T. Heck, R. Hischier, T. Nemecek, G. Rebitzer and M. Spielmann, *Int. J. Life Cycle Assess.*, 2005, **10**, 3–9.
12 The ecoinvent centre, Ecoinvent 2.01 database, http://www.ecoinvent.ch.
13 National Renewable Energy Laboratory, U.S. Life-Cycle, Inventory Database, http://www.nrel.gov/lci/.
14 R. Bretz and P. Frankhauser, *Chimia*, 1997, **51**, 213–217.
15 A. D. Curzons, C. Jiménez-Gonzalez, A. L. Duncan, D. J. C. Constable and C. V. L, *Int. J. Life Cycle Assess.*, 2007, **12**, 272–280.
16 P. Saling, A. Kicherer, B. Dittrich-Krämer, R. Wittlinger, W. Zombik, I. Schmidt, W. Schrott and S. Schmidt, *Int. J. Life Cycle Assess.*, 2002, **7**, 203–218.
17 G. Wernet, S. Hellweg, U. Fischer, S. Papadokonstantakis and K. Hungerbuhler, *Environ. Sci. Technol.*, 2008, **42**, 6717–6722.
18 K. Hornik, M. Stichcombe and H. White, *Neural Networks*, 1989, **2**, 359–366.
19 D. T. Pham, in *Neural Networks for Chemical Engineering*, ed. A. B. Bulsari, Elsevier, Amsterdam, 1995.
20 K. Swingler, *Applying Neural Networks*, Academic Press, Inc., 1996.
21 Verein Deutscher Ingenieure, VDI-richtlinie 4600: Cumulative Energy Demand, Terms, Definitions, Methods of Calculation, Verein Deutscher Ingenieure, Düsseldorf, 1997.
22 M. A. J. Huijbregts, L. J. A. Rombouts, S. Hellweg, R. Frischknecht, J. Hendriks, D. Van de Meent, A. M. J. Ragas, L. Reijnders and J. Struijs, *Environ. Sci. Technol.*, 2006, **40**, 641–648.
23 The Intergovernmental Panel on Climate Change, Fourth Report. Climate Change 2007, http://www.ipcc.ch/ipccreports/ar4-wg1.htm.
24 M. Goedkoop and R. Spriensma, *The Eco-Indicator 99: A Damage Orientated Method for Life-Cycle Impact Assessment*, Pre Consultants, 2000.