Approaches for Addressing Life Cycle Assessment Data Gaps for Bio-based Products

Llorenç Milà i Canals, Adisa Azapagic, Gabor Doka, Donna Jefferies, Henry King, Christopher Mutel, Thomas Nemecek, Anne Roches, Sarah Sim, Heinz Stichnothe, Greg Thoma, and Adrian Williams

Keywords:

carbon footprint
data gaps
extrapolated data
industrial ecology
proxy data
surrogate data

Supporting information is available on the *JIE* Web site

Summary

There is an increasing need for life cycle data for bio-based products, which becomes particularly evident with the recent drive for greenhouse gas reporting and carbon footprinting studies. Meeting this need is challenging given that many bio-products have not yet been studied by life cycle assessment (LCA), and those that have are specific and limited to certain geographic regions.

In an attempt to bridge data gaps for bio-based products, LCA practitioners can use either proxy data sets (e.g., use existing environmental data for apples to represent pears) or extrapolated data (e.g., derive new data for pears by modifying data for apples considering pear-specific production characteristics). This article explores the challenges and consequences of using these two approaches. Several case studies are used to illustrate the trade-offs between uncertainty and the ease of application, with carbon footprinting as an example. As shown, the use of proxy data sets is the quickest and easiest solution for bridging data gaps but also has the highest uncertainty. In contrast, data extrapolation methods may require extensive expert knowledge and are thus harder to use but give more robust results in bridging data gaps. They can also provide a sound basis for understanding variability in bio-based product data. If resources (time, budget, and expertise) are limited, the use of averaged proxy data may be an acceptable compromise for initial or screening assessments. Overall, the article highlights the need for further research on the development and validation of different approaches to bridging data gaps for bio-based products.

Address correspondence to:
Dr Llorenç Milà i Canals
Safety and Environmental Assurance Centre
Unilever R&D Colworth
Colworth Science Park
MK44 1LQ Bedfordshire, UK.
Llorenc.Mila-i-Canals@Unilever.com
www.unilever.com/sustainability

DOI: 10.1111/j.1530-9290.2011.00369.x

Volume 15, Number 5

Introduction

There is an increasing need to describe and quantify the life cycle environmental impacts of bio-based[1] products. This is particularly true in the area of carbon footprinting (CF) and greenhouse gas (GHG) reporting of products and organizations (Weidema et al. 2008). Such life-cycle-based studies are very data intensive, and data availability in traditional life cycle assessment (LCA) databases is currently very limited for many bio-based products; for example, Ecoinvent V2 only provides data for potatoes from two countries (Switzerland and the United States), although potatoes are a staple food in many parts of the world. Even for products for which several studies exist, the geographical coverage of bio-based production is very limited. This is particularly an issue given the variability of agricultural production in different regions (see, e.g., Milà i Canals et al. 2006, 2007; Mouron et al. 2006; Sim 2006; Edwards-Jones et al. 2009). The coverage in terms of product groups is also concentrated on a few subsectors (e.g., meat, milk, a few cereals, some vegetable oils), with many basic food ingredients not covered, including most vegetables, fruit, herbs, and spices. In recognition of these data gaps, several recent studies provide a compilation of the carbon footprint data for different bio-based products, including the work of Defra (2009) and CCaLC (2010).

In the absence of data, LCA practitioners have to either fill the gap by creating a new data set, find a "surrogate" that bridges the data gap, or leave a data gap. The first option is often not possible due to time and resource constraints, and the last is not recommended due to the uncertainty related to the excluded data. Therefore, finding surrogate data may be a compromise between the two ends of the spectrum. By "surrogates," we refer to any data set (*source* data) that is sufficiently similar to the process, material, or product for which data do not exist (*target* data) and that is used to represent the target data.

This article focuses on the use of data surrogates in GHG assessments and CF of bio-based products. The principles discussed here should be applicable to other environmental impacts; however, spatial dependency is higher in impacts such as toxicity, acidification, eutrophication, and biodiversity, and readers should take care when interpreting the conclusions and recommendations from this article.

The variability in the production of bio-based products is larger than for abiotic ones. This is mainly due to the fact that bio-based products are derived from natural systems, which are subject to environmental conditions, as opposed to technical systems, where conditions are normally controlled (and often standardized). The main sources of variability in agricultural production include soil, climate, topography, crop or animal variety, farming system and intensity of farming, type of farm (specialized versus mixed farm, stockless versus farm with livestock), tradition, and education of the farmer (see, e.g., Milà i Canals et al. 2006; Mouron et al. 2006; Edwards-Jones et al. 2009; Nemecek et al. 2009a).

Developing ways to derive surrogate LCA data sets for bio-based products that account for their variability would help address the challenge of poor data availability. In addition, researchers need guidance to justify suitable surrogates when data on a particular product are not available. Thus, the motivation to progress in this area of work includes the following:

- the need for guidance on approaches to bridging the data gaps and the types of data that can be used as surrogates
- the need to understand how the inherent variability of bio-based products may be taken into account
- the need to understand the uncertainty related to the use of surrogate data.

This article focuses on the first point by exploring the main approaches that can be used in LCA to provide surrogates and discussing their consideration of variability and uncertainty. The next section offers definitions and a systematic classification of these approaches, which are later illustrated with examples using surrogate data for bio-based products; most examples are drawn from existing studies, and one was specifically developed for this article. The implications of such approaches are then discussed in terms of balancing the effort with quality of the results (uncertainty) in relation to potential applications; finally, the main conclusions and further research needs are highlighted.

Description of Existing Approaches

Two main approaches for bridging data gaps in LCA studies are distinguished in this article (table 1):

1. Use of proxy data sets: These describe alternative products for which data exist (*source* data) and are assumed to have similar environmental impacts to the products of interest (*target*). They are used to bridge data gaps without changing the original values

Table 1 Types of surrogate data used for bridging life cycle assessment (LCA) data gaps for bio-based products

Term	*Definition*	*Formal representation*	*Source inventory/LCI modeling*	*Illustrative example*
Scaled proxies	Use *known* LCIs of Products A through F (weighted according to the product mix) to represent the *unknown* LCIs of Products G through K, which may be functionally quite different	A through F represent G through K	Unchanged[a]	In a multi-ingredient product, such as pizza, accurate LCI data for only $85\%_{wt}$ of the ingredients are available. The data gaps are bridged by linearly scaling up the data for 85% of ingredients to 100%
Direct proxies	Use LCI of Product A to represent LCI of Product B; Products A and B are assumed to have similar characteristics and function	A represents B	Unchanged[a]	Using an LCI of apples to represent pears
Averaged proxies	Use weighted or nonweighted average or median LCI of Products A through F to represent LCI of Product G. Products A through F are considered similar to G	A through F represent G	Unchanged[a]	Using the average of LCIs of apples from four countries, weighted, for example, according to production volume, to represent, for example, French apples
Extrapolated data	Using LCI of Product A or LCIs of Products A through X to estimate LCI of Product B, making appropriate changes	A is used to generate data for B	Adapted to the target system	Taking the LCI of chicken production to derive an adapted LCI of turkey production, by changing production parameters

Note: LCI = life cycle inventory.
[a] That is, the source data are kept intact or only averaged to represent the target system.

beyond statistical calculations, such as averaging. The selection and use of proxy data sets is usually based on the knowledge and experience of the LCA practitioner, and the possibility to validate such choices is often limited. In table 1, three types of proxy data can be identified:

- *Scaled proxies* scale the existing data to estimate the missing data on the basis of, for example, quantities or composition of the products.
- *Direct proxies* represent direct substitutes of the target with source data with no changes.
- *Averaged proxies* refer to the data derived through averaging the data for a group of (source) products that can be assumed to be similar to the target product.

2. Data extrapolation: This approach generates new data by adapting data from *source* data sets outside the range of their original validity to reflect better the *target* situations, for example, between countries or regions, products, technologies, or time periods. The newly generated data are then used to bridge the data gap in the target situation. The level of detail and sophistication of the extrapolation algorithms can vary, but, in general, this approach requires thorough knowledge of the systems being extrapolated and involves some changes in life cycle inventory (LCI) models.

In table 1, the approaches are ranked according to the effort required from the practitioner, which is almost nil when one is scaling data (the study is focused on those products for which data exist, and the overall result is scaled up linearly to include the missing data); the effort increases in direct proxies (the practitioner has to find one suitable proxy for each data gap), and it is further increased in averaged proxies (two or more alternative data sets are needed for each gap). Finally, extrapolation requires significantly more effort, as the source system characteristics need to be explored and modified to represent better the target system for each data gap.

A qualitative relationship between the level of effort and the uncertainty expected with each of these approaches is illustrated in figure 1. This relationship is discussed in the following sections.

Proxy Approaches

The motivation for using proxy data is usually that LCA practitioners feel that leaving the data gap unfilled, thereby contributing a burden of zero to the overall LCA result, is more wrong than using proxy data. Using proxy data allows for the estimation of an approximate contribution of the target part of the whole system. As noted in table 1, three main approaches for deriving proxy data sets can be distinguished: scaled, direct, and averaged proxies; these are discussed below.

In scaled proxies, all parts of the system for which data are not available are ignored, and once the impact is calculated for the remaining parts, the result is linearly scaled up to 100%. The advantage of this procedure is that it enables quick estimates because it avoids the search for suitable (direct or average) proxies of the missing parts of the system; the obvious drawback is that the researcher does not know whether the source data (in the pizza example, e.g., for wheat and mozzarella) are a good proxy for the target data (e.g., mushrooms and artichokes).

In the case of direct proxies, a source data set is used to represent a similar (target) product. The uncertainty associated with the use of direct proxies is generally high, because even if variability between production systems is incorporated, one cannot know whether such variability would include the potential values for the target system. To reduce uncertainty, the researcher needs good knowledge of the target (and source) production systems to evaluate whether the source is a good proxy for the target. For example, the following key indicators related to agricultural (and possibly other biotic) systems should be as similar as possible between the source and target products: yield, nitrogen input, diesel use, amount of irrigation water, and the intensity of pesticide use. Other criteria may include duration of cultivation (permanent versus annual crops), taxonomy (genus, species, cultivar, etc.), harvested parts (seeds, leaves, roots, etc.), and type of farming system (organic or conventional).

In averaged proxies, the average (or median, weighted or unweighted) values of two or more

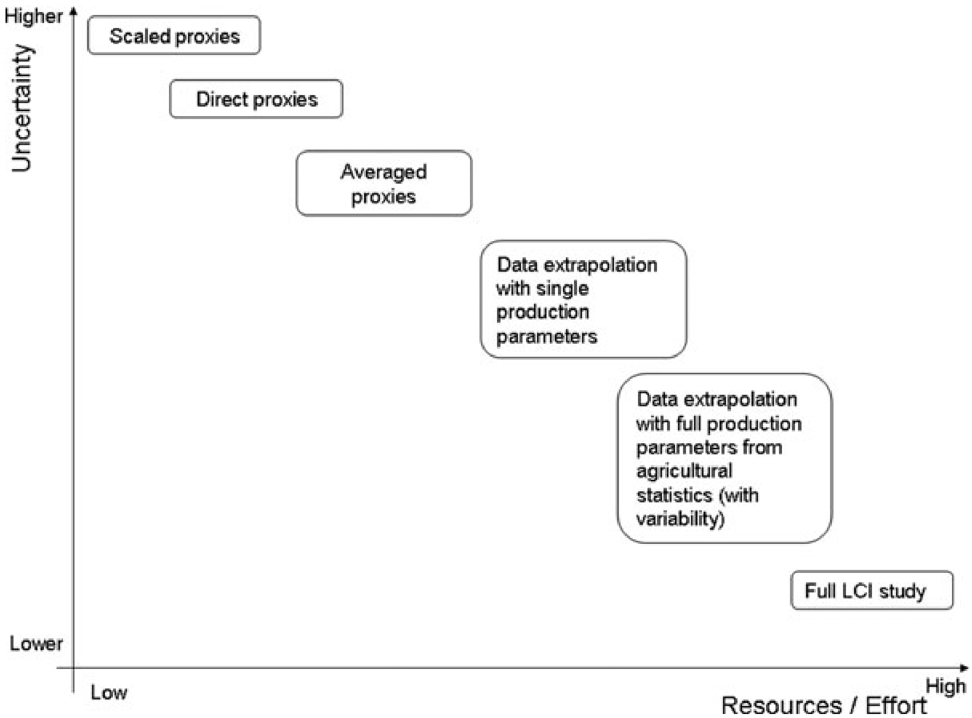

Figure 1 Trade-offs between effort and uncertainty in the approaches to managing environmental data gaps. LCI = life cycle inventory.

source data sets are used to represent a target product. If averaging is performed with a significant number of initial data sets that cover a broad range of products within the target group and if variability is incorporated into the final result, then the range of impacts from the resultant proxy data should be a reasonable representation of the expected impacts of the target process. The influence of the proxy data sets on the results can be examined as usual—that is, through sensitivity or contribution-to-variance analysis.

Extrapolation Approaches

As mentioned above, "extrapolation" refers to the adaptation of data from source data sets to the target situations by means of models. Such adaptation requires initial knowledge of the parameters that influence the values in the source data sets. Information on these parameters is usually more readily available than information on the environmental impacts; thus, when comparing the parameter values for the existing and the new situations, one should be able to extrapolate the environmental impacts from one to the other. In cases in which the source and target production systems are similar, the predicted environmental impacts are likely to be more indicative of the actual impacts.

Case Studies

This section presents several case studies that illustrate applications of both proxy and extrapolated data and the impact on the overall study results. Most examples are drawn from the literature and contain a mix of approaches, of which we have highlighted the most relevant for illustration purposes.

Use of Proxy Data: Scaled, Direct, and Averaged Proxies

Data gaps abound in assessments of complex products—for example, studies of people's diets, national food sector assessments, complex multi-ingredient food products, or product portfolio analyses. The next sections discuss some examples in which proxy approaches have been predominantly used.

Direct Proxies

An example that uses direct proxies is offered by Muñoz and colleagues (2010), who assess the cradle-to-grave impacts of the Spanish food sector. In that study, food is divided into 53 groups, for which direct proxy data sets are used (e.g., legumes are generalized with a data set for dry peas). This approach provides an estimate of the contribution of different food items to the environmental impacts of the food sector that is of the same order of magnitude as previous studies that followed different approaches to managing data gaps (e.g., Santacana et al. [2008], who used an input-output approach). Thus, even though in the absence of specific full LCA studies it is difficult to be certain which results are more correct, the fact that different approaches lead to similar results suggests that the conclusions at least point in the same direction.

A similar example is provided by a project offering a national-level consumption-orientated GHG inventory for food and drink in the United Kingdom (Audsley et al. 2010). The data sources used included detailed LCA data produced by the authors and literature data, which were used both as direct proxies (e.g., from greenhouse tomatoes to greenhouse peppers) and to average whole commodity groups from a limited number of data sets (e.g., field vegetables from the average of several other field vegetables). Extrapolation was also applied in some cases; for example, sugar beets were extrapolated from potatoes according to data from Tzilivakis and colleagues (2005) on fertilizer application rates, tillage methods, pesticide application rates, and yields. The study of Audsley and colleagues (2010) was very useful in quantifying the relative orders of magnitude of primary production, postfarm gate activities (distributing, processing, and consumption), and GHG emissions from land-use change, although uncertainties were large for small scale domestic production and for some overseas commodities.

In the two case studies mentioned above, the use of proxies allowed for the realization of complex assessments in a reasonable time frame and budget. It is not possible, however, to check the reliability of the outcomes other than by doing the whole study with specific data, which would require significant resources. In general, identifying clearly all data gaps and describing the approach used to fill them can help to improve transparency and confidence in the results.

Averaging Proxies From Statistical Analysis of Crop Life Cycle Impact Assessment Results

Mutel and colleagues (2009) generated data sets for a selected group of fruits and vegetables (field tomatoes, carrots, onions, pumpkins, pineapples, papayas, kiwi fruit, and bananas) to test whether different types of crops could be easily grouped into generic classes (e.g., tropical fruit versus temperate fruit). For each fruit or vegetable, the authors constructed a data set with uncertainty ranges representative of global production conditions using global databases or specific literature sources. In many cases, this involved adapting specific inventory data from countries with industrialized agricultural systems, with uncertainty ranges expected for general classes of inputs (e.g., fertilizer, farm machinery, and irrigation inputs). Uncertainty distributions were fitted from empirical cumulative distribution functions on the basis of data drawn from databases or a collection of published papers for each crop parameter. Figure 2 shows the ranges observed under Monte Carlo analysis for global warming potential (GWP), measured in kilograms of carbon dioxide equivalents (kg CO_2-eq.) per kilogram (kg) of product[2].

For the specific crops and indicators chosen in this case study, no groups could be statistically distinguished. For example, the sample set of tropical fruits (pineapple, papaya, kiwi, and banana) could not be considered different from the sample set of temperate fruits and vegetables (all others) at even a 50% confidence level (according to the Kolmogorov-Smirnoff two-sample test, $p = 0.98$). A generic data set was constructed with kernel density estimation from the samples generated for all crops. Although this generic data set is a poor fit for many of the individual crops, it could be used as a proxy if crop data were not available. The lack of observed grouping between different product categories is tempered by the limits of the sample data set. These results suggest that

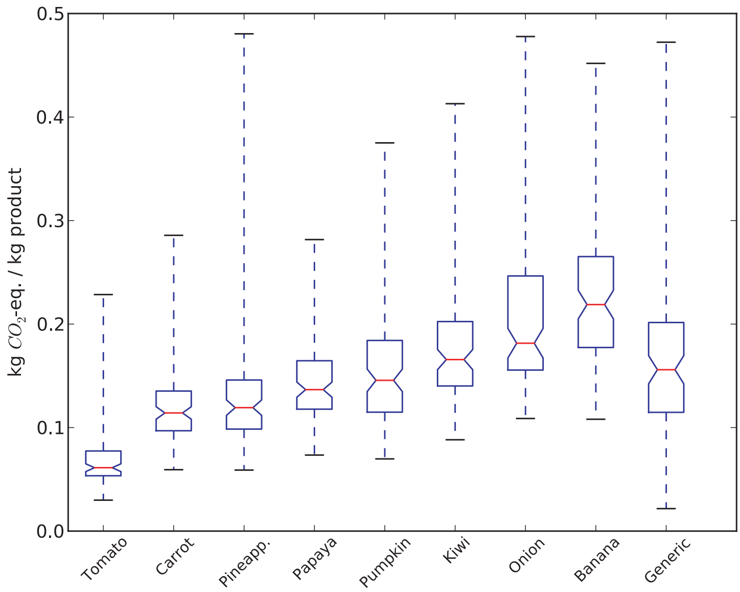

Figure 2 Box and whisker plots of global warming potential scores expressed in kilograms of carbon dioxide equivalents (kg CO_2-eq.) from Monte Carlo analysis (n = 100) of a set of fruits and vegetables. Boxes represent the 25% to 75% quantiles, the median is the line inside the boxes, and the dotted lines represent the data range. The generic distribution is constructed from a Gaussian kernel density estimator; input data were the complete sample set of all individual crop Monte Carlo runs (Mutel et al. 2009).

there is a research need for further species to be considered in investigations of the applicability of averaged data sets.

The analysis also indicates that life cycle impact assessment (LCIA) results vary as much within crops as they do between different crops. For example, the ratio of the upper and lower 95% confidence limits for each crop (2.1 to 4.1) was approximately the same as the ratio of the median values of the lowest and highest scoring crops (tomato and banana; ratio of 3.6). These results show that using only a mean value can be misleading when one is studying bio-based products and that variability should be incorporated whenever possible.

Averaging and Scaling

Milà i Canals and colleagues (2009b, 2011) assessed the global CF of the Knorr brand portfolio. The complexity of Knorr's product portfolio made a bottom-up, conventional product-based carbon footprint approach impractical. Ingredients and processes that were considered similar were aggregated in "building blocks" (e.g., "dairy products" instead of milk, cream, etc.; "drying" instead of air drying, spray drying, drum drying, etc.), and this facilitated assessments of different combinations of such building blocks for different products. Apart from the simplification in the analysis, such grouping was required due to the lack of specific data for most of the ingredients used in Knorr products. To assess the robustness of the results, they individually assessed the variability around the averaged proxies for most building blocks and propagated it through the calculations (Milà i Canals et al. 2011). Once the majority of the portfolio's volume had been assessed, the results were scaled to estimate the total impact of the brand. For example, the impact calculated for wet soups was scaled to the production volume of wet soups and wet sauces (the latter were not specifically assessed).

The variability assessment of the averaged proxies greatly supported the interpretation of the results by identifying a confidence range around the carbon footprints for both product format (e.g., dry soups, wet soups, bouillon cubes) assessment and target setting. Using only mean values, they estimated Knorr's carbon footprint as 3.5 (or "between 3 and 4," Milà i Canals et al. 2009b) million tonnes CO_2-eq. per year[3], whereas the variability analysis suggested that it lies between 3.4 and 4.8 with a 95% confidence interval (Milà i Canals et al. 2011). Such a range provides at least the order of magnitude for the brand's footprint and facilitates target setting for improvement. As long as progress against the target is measured with the same approach, it should be possible to assess with reasonable confidence whether the brand is moving in the right direction and at what rate.

The implications of using averaged proxies rather than specific data for most ingredients are clear in this case study. First, the use of proxies allowed the realization of the study, which could not have been carried out had all ingredients needed assessments with specific studies. By characterizing them with an averaged proxy and variability range, the probability density function (PDF) of the final impact could be more confidently stated to contain the real impact. When the studied product types were scaled up to cover the whole of the portfolio, the uncertainty increased, but in an informed manner: The brand consciously chose to study some products in more detail and accepted that minor product formats were more coarsely estimated. Therefore, the final result is valuable as an indication of where the impacts lie and to monitor progress against reduction targets. In comparison to direct and scaled proxies, the use of averaged proxies with variability ranges is considered to provide more robust and less uncertain results.

Extrapolation of Data Sets

When enough knowledge and resources are available, extrapolation allows researchers to better represent the target system by modifying the source data.

Extrapolation of Specific Life Cycle Inventory Parameters: From Chicken to Turkey

In trying to estimate the impacts for turkey production, one could assume that turkeys are sufficiently similar to other poultry for which LCA studies exist, such as chickens. Stichnothe and colleagues (2010) have used chicken data as an averaged proxy for turkey on the basis of a range of literature data for chicken meat (see table 2).

Alternatively, instead of using the chicken values for turkey as direct or averaged proxy, it may be possible to study the production practices for these two species to determine how they compare and what flows one needs to adapt in the inventory data to extrapolate from one to the other. Much of the data are available from standard management texts—for example, the work of Nix (2009) or ABC (2009). Values for the main production parameters for chicken and turkey have been gathered from the literature and are detailed in the Supporting Information available on the journal's Web site together with definitions for such parameters (see table S1); the differences between the work of Nix (2009) and ABC (2009) give an indication of the variability of typical production data.

Thus, with these readily available data and with an LCA model of broiler chicken production that includes feed requirements and conversion to meat (Williams et al. 2006), together with time-dependent variables (e.g., ammonia emission rate), we estimated turkey production on the basis of broilers, as described in detail in the Supporting Information on the Web. Some simplifying assumptions were needed regarding the parameters that could be used to extrapolate the various impacts; for example, we extrapolated impacts related to feed inputs and manure outputs using the feed conversion ratio (FCR), and we extrapolated impacts from direct energy use in proportion to the days of production per unit live weight.

As an example, the GWP related to the production of feed for chicken is 1,721 kg CO_2-eq./tonne of chicken meat (Williams et al. 2006), and the average FCR is 1.8 for chicken and 3.1 for turkey (table S1 in the Supporting Information on the Web); thus, by multiplying the GWP for feed production for chicken by the relationship between these parameters (3.1/1.8), we obtain the extrapolated value for GWP for feed production for turkey (3,052 kg CO_2-eq./tonne turkey meat). The full results are shown in table S2 in the Supporting Information on the Web. With the extrapolation, the GHG emissions for turkey are estimated at about 4.46 kg CO_2-eq./kg meat, or about 50% higher than for chicken, which

Table 2 Carbon footprint values for chicken and poultry (developed from Stichnothe et al. 2010)

Reference	*kg CO_2eq./kg functional unit*	*Functional unit*
Danish database on food (www.lcafood.dk)	1.86	Live chicken
Danish database on food (www.lcafood.dk)	3.11 *to* 3.28	Chicken meat
German database (www.probas.umweltbundesamt.de)	1.57 *to* 1.83	Live chicken
Jungbluth (2000)	2.32 *to* 2.90	Chicken meat
Williams and colleagues (2006)	2.57	Poultry meat
Wallén and colleagues (2004)	2.81	Poultry meat
Baumgartner and colleagues (2008)	2.8 *to* 3.3	Live chicken
Pelletier (2008)	1.4	Live chicken
Azapagic and colleagues (2011)	3.7	Chicken meat

Note: kg CO_2eq. = kilograms of carbon dioxide equivalents.

has an average GHG value of around 3 kg CO_2-eq./kg meat (if we consider the values per meat from table 2).

Stichnothe and colleagues (2008, 2010) report a similar value of 4.97 (turkey LCA based on the feed composition, the feed conversion rate, emissions from turkeys, waste, etc.), and Williams and colleagues (2006) suggest 5.13 kg CO_2-eq./kg turkey in a different LCI for turkey production. The values available for chicken (table 2) suggest that using direct or averaged proxy values for turkey on the basis of chicken data is inadequate (see also Stichnothe et al. 2010). Using chicken LCI data as a starting point, however, and extrapolating them with the relevant production parameters (available in agricultural literature) results in a reasonable estimate when compared with turkey LCI data, as illustrated here. In addition, the extrapolation process forces the analyst to critically explore the characteristics of the whole system and the relevance of the data being used. It is thus reasonable to assume that the outcome will be at least more reliable than using direct proxies without further adaptation.

Full Extrapolation of LCI Data for Crops

Nemecek and colleagues (2009a) and Roches and colleagues (2010) describe MEXALCA, a modular extrapolation of agricultural LCA. Within MEXALCA, the impacts of crop production are estimated for all producing countries by means of one single inventory and estimators for the main production parameters, developed for all the countries from statistical data from the Food and Agriculture Organization (FAO; see Roches et al. 2010). Such parameters include nitrogen (N), phosphorus (P), and potassium (K) fertilizers; pesticides; irrigation water; and water to be evaporated in drying as well as the mechanization and tillage intensity. The impacts are then weighted for each country by the production volume of the country, and the statistical distribution of the weighted impacts on the global scale is determined. Extrapolating values in this way offers a relatively quick estimate of the probability density function of production impacts in the world for a given crop. For instance, the extrapolated statistical distribution of GWP for potato production in the world is displayed in figure 3; the percentage in the *x*-axis is the cumulative share of global potato production that is produced with a GWP equal to or lower than the value shown in the *y*-axis. It shows that 50% of the potato production is estimated to have a GWP below 0.13 kg CO_2-eq./kg, the other 50%

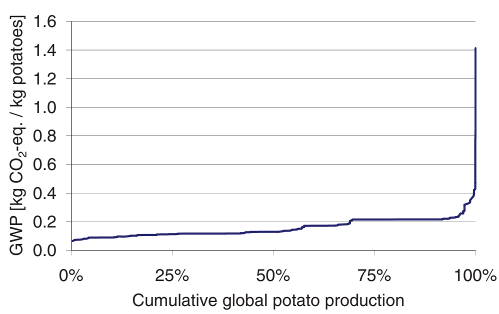

Figure 3 Global warming potential (GWP) as extrapolated by MEXALCA expressed as a function of the cumulative global potato production. Values are in kilograms of carbon dioxide equivalents per kilogram of potatoes (kg CO_2-eq. kg^{-1}).

above this value. The distribution is left-skewed, as 90% of producers have values in the 0.1 to 0.22 bracket and 10% of the producers are in the 0.22 to 1.4 bracket.

The possibilities for extrapolation are quite different depending on the impacts and products concerned (Roches et al. 2010). For example, for global resources, such as fossil energy or minerals (e.g., P and K), extrapolation works relatively well. This is also the case for environmental impacts, such as global warming and ozone formation. This is because such impacts are the same regardless of where resources are consumed or emissions are caused.

Regional and local environmental impacts are much more difficult to extrapolate, because they are influenced by a number of site-dependent parameters, such as soil characteristics, topography, and climate. This applies, for example, to the assessment of eutrophication, acidification, and biodiversity. Furthermore, toxicity impacts of agricultural systems are mostly dominated by pesticides; without very specific knowledge of the pesticide applied, extrapolation is hardly possible. In the same vein, the use of water for irrigation is highly dependent on the climate, and the severity of the impacts depends on the availability of the water in the watershed (Milà i Canals et al. 2009a). This requires quite a high geographical resolution and is thus not easily extrapolated.

MEXALCA provides a framework for deriving extrapolated values, helping the analyst to understand better the extrapolated system. Its main strength possibly lies in the explicit description of variability in production conditions of the studied area. For example, figure 3 suggests one order of magnitude variance in the GWP of world potato production; however, it also shows that about 90% of the world's potatoes are produced with a GWP that varies only two-fold.

Exploration of Variability in Food Commodities With Full Life Cycle Inventory

The main advantage of using the FAO statistics in such an exhaustive way as explained in the preceding section is that the results present a picture of where impacts are likely to lie for a crop from a certain region (continent) or even from nowhere in particular. This approach may be particularly useful when one is describing impacts of bio-based commodities traded in the open market, for which there is no traceability to production region and conditions (which can affect their environmental impacts significantly).

If variability (e.g., PDF) were not estimated through extrapolated data as done by MEXALCA, full LCIs would be required for a large number of producers in the region that needs representing, which is very resource intensive. For example, Thoma and colleagues (2010) provide GWPs for milk production in more than 500 U.S. farms (see figure 4 and figure S1 in the Supporting Information on the Web), which demonstrates the variability associated with these impacts across production practices. Analysis of these data shows that the variation in GWP is strongly correlated to farm management and not strongly correlated to region or herd size. Thus, because milk is traded as a commodity (at least within countries and regions), using its average impact together with probable ranges of variation that include the mix of production practices may be more appropriate than using a single value from a well-characterized practice (farm). Thoma and colleagues (2010) collected data by survey from more than 500 U.S. dairy farms that represented a wide range of production practices, from small pasture-based to large confined-animal facilities. Rations for each animal class were used to create an individual farm footprint. Direct and averaged proxies and extrapolation were all used to determine the footprint of the total farm ration. For example, for some farms incomplete rations were reported and regional average rations were substituted; in other cases data were unavailable for some specific feeds, and the authors created a proxy by taking an average of several similar feeds for which data did exist. Thoma and colleagues (2010) thus provide a level of depth for milk that is very uncommon to find in LCI databases; this could be seen as an aspiration in terms of data but is highly unlikely to be achieved for many bio-based products.

The high degree of variability that is shown in these data underscores the importance of attaching uncertainty information to LCI data. In this case, as shown in table 3, although there are outliers at both the high and low ends of the

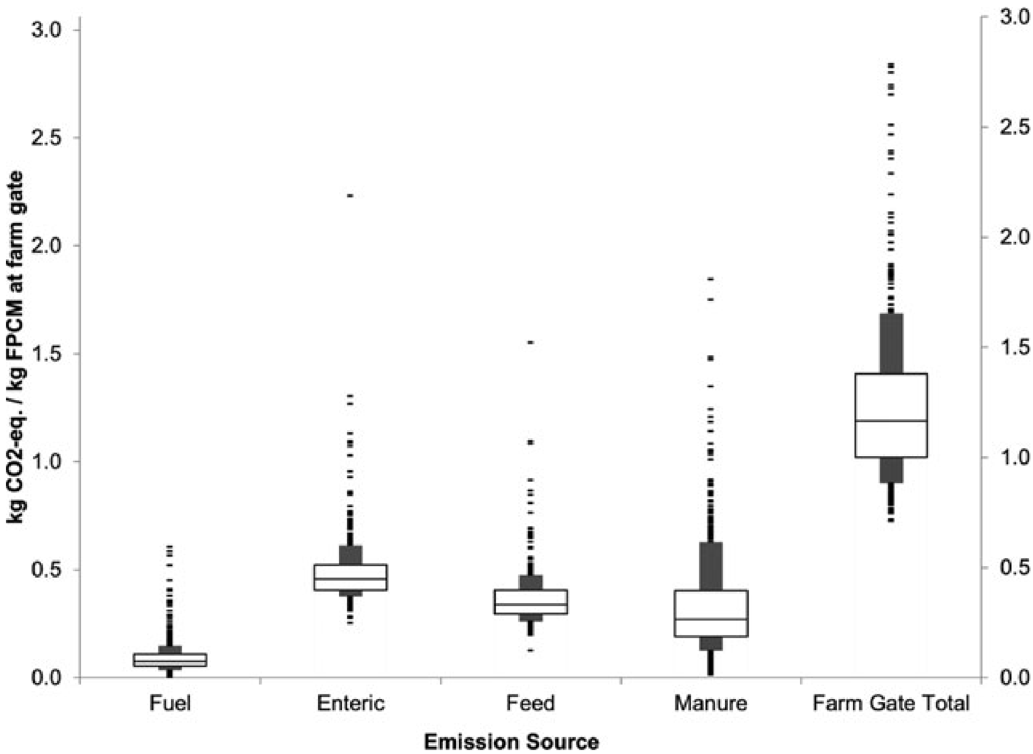

Figure 4 Global warming potentials (GWPs) of fat and protein corrected milk (FPCM) at the farm gate (4% fat; 3.3% protein) in U.S. farms (data from Thoma et al. 2010). The box represents the 25th and 75th percentiles, with the median represented by the horizontal line. The thick bars denote the 10th and 90th percentiles; outliers are shown as discrete points. Values are in kilograms of carbon dioxide equivalents per kilogram of FPCM (kg CO_2-eq./kg FPCM).

spectrum, the mean value for U.S. production could be reasonably substituted as a direct proxy for milk production in most regions of the world, and the reverse is also true. Here, the underlying production similarities appear to dominate milk production globally, in contrast to the previous example with fruits and vegetables (figure 2), for which a factor of 2 is approximately the best we can expect in terms of the accuracy of replacement data. This further underscores the critical role of the analysts' understanding of the source and target systems.

Discussion

The above examples illustrate some of the implications of different approaches to managing data gaps for bio-based products. They also indicate the balance between effort and quality of the results provided by the different approaches (figure 1). In general, expert judgment suggests that effort and quality increase as one moves from scaled proxies to direct to averaged ones (with consideration of variability) and then to extrapolation. Uncertainty in all the processes to find surrogate data still needs to be further resolved through comparison of surrogates with full studies, however.

In all approaches and particularly with proxies, sensitivity analyses should be carried out, but it is often difficult to identify sensible data ranges for such analyses. In any case, even direct proxies are probably closer to reality (i.e., impact higher than zero) than leaving a data gap, and thus practitioners will continue to use them as demand for bio-based LCA and CF studies increases more rapidly than the capacity to provide full studies of all products.

Table 3 Summary of previously published life cycle assessment (LCA) for fluid milk production and consumption

Study	*Emissions ($kg\ CO_2$-eq.)*	*Functional unit*	*Allocation % to milk*	*Characterization factors (CO_2, CH_4, N_2O)*	*Study description*
Basset-Mens and colleagues (2009)	0.72	kg milk at farm gate	85	1, 21, 310	New Zealand, estimated national average
Capper and colleagues (2009)	1.35	kg milk at farm gate			U.S. average, 1944 versus 2007
Cederberg and Flysjo (2004)	0.9 to 1.04	kg milk at farm gate	90	1, 21, 310	Sweden, 23 farms
Cederberg and Mattsson (2000)	0.90 to 1.1	kg ECM at farm gate	85	IPCC 1995	Sweden, organic versus conventional
Cederberg and colleagues (2009)	1.02	kg ECM at farm gate	85	1, 25, 298	Sweden, 1990 versus 2005
Cederberg and colleagues (2009)	1.08	kg ECM at retail	85	1, 25, 298	Sweden, 1990 versus 2005
DEFRA (2007)	1.18	kg milk at farm gate	100	-	United Kingdom
Eide (2002)	~0.54 to 0.65	kg milk at end of life	65	-	Norway, study of three dairies
Eide (2002)	~0.41 to 0.46	kg milk at farm gate	65	-	Norway, study of three dairies
Gerber and colleagues (2010)	2.4	kg FPCM at retail	~90	1, 25, 298	International average
Gerber and colleagues (2010)	1.3	kg FPCM at retail	~90	1, 25, 298	U.S. average
Gerber and colleagues (2010)	1	kg FPCM at farm gate	~90	1, 25, 298	U.S average
Haas and colleagues (2001)	1.0 to 1.3	kg milk at farm gate	-	-	Southern Germany, intensive, extensive, and organic
Guinard and colleagues (2009)	1.2	kg milk at end of life	Economic	1, 25, 298	Literature review of 60 studies, primarily European
Guinard and colleagues (2009)	1	kg milk at farm gate	Economic	1, 25, 298	Literature review of 60 studies, primarily European
Thoma and colleagues (2010)	1.23	kg FPCM at farm gate	Biophysical/ causal	1, 25, 298	U.S. national average, primary data collected from more than 500 farms

Note: kg CO_2-eq. = kilograms carbon dioxide equivalents; CO_2 = carbon dioxide; CH_4 = methane; N_2O = nitrous oxide; FPCM = fat and protein corrected milk; ECM = energy corrected milk, which is equivalent to FPCM.

When one is averaging data, it is possible to suggest ranges for variability and sensitivity analyses, although it is still uncertain whether such ranges cover the whole spectrum of the product or group represented. As shown in the Knorr example (Milà i Canals et al. 2011), propagating the expected variability may enhance the robustness of results and provide enough information to support decisions that require only a notion of orders of magnitude or trends (e.g., detection of hotspots, broad-brush strategic decisions, target setting).

A more common type of proxy data use (although practitioners may not often recognize it as such) is when data for one product grown in one country are used to represent the same product in another country (e.g., using GHG emissions for Dutch milk to represent English milk). Nemecek and colleagues (2009b) have shown for cereals, however, that the impacts of wheat and barley within a country can be more similar than the impacts of the same species in two neighboring countries. In this sense, it may be wiser to use slightly different products from the same country as proxies than to use the same product from a different country.

When more time is devoted to the analysis and adaptation of LCI data (data extrapolation), the results are likely to be more credible, but the effort is higher. This approach is thus good when a relatively small number of products are to be assessed and enough technical knowledge exists but perhaps there is limited access to data. The effort for the analysis increases if there are no main hotspots within the system or if the key influencing parameters are completely unknown. Extrapolation methods, such as MEXALCA, allow for a quick assessment of large amounts of statistical data describing agricultural production conditions (e.g., from FAO); such approaches give a broad picture of likely impacts and variability, which are enough for several types of decisions, but are associated with a relatively high level of uncertainty at the individual country level. Thus, extrapolation is a valuable approach for deriving a generic value for apple cultivation at the global (world) or at the regional (e.g., European) scale. When one first extrapolates the impacts to all producing countries and then calculates the median and some quantiles of these impacts (at the global or regional scales), the obtained value should give a more robust estimate for bio-based commodities than when one averages solely the regional values.

Apart from the considerations on uncertainty of the surrogates, this article has also illustrated the importance of considering variability in bio-based products. Approaches for considering such variability have been shown for averaged proxies and for extrapolation from agricultural statistics with MEXALCA. Indeed, there is no reason why a better consideration of variability should not be more common practice in bio-based LCA studies: Information on variability is available in agricultural statistics, yet this is not usually incorporated into the LCA results. Such statistical information may be used to provide, for example, 3–5-year averages, which may be more representative for bio-based production than very detailed information covering one single year. A more detailed analysis of this issue is offered by Röös and colleagues (2010) in a case study of table potatoes, where they report a variation of about +/−30% in a simple product (fresh potatoes) from a single region in Sweden, or the work of Thoma and colleagues (2010), who reported a range larger than a factor of 3 for milk produced in the upper Midwest in the United States.

The inclusion of variability reduces the discriminatory capacity of the results (due to overlaps of variability ranges) but reduces the uncertainty and increases the robustness of the conclusions when differences are shown. For example, the range of values used to derive the averaged proxy may overlap with values for other products (e.g., New Zealand apples); this could be seen as limiting the usefulness of results (we cannot say, "A is better than B") but above all would avoid spurious certainty based on mean values.

On the basis of discussions around the approaches illustrated in this article, table 4 suggests what types of applications may be supported when researchers use data derived following the above-discussed approaches. For instance, product labeling should not rely on 100% direct proxies, but a small percentage of proxy data (e.g., less than 10%) in a study supporting labeling may be acceptable. This is important because, as discussed throughout the article, there is an increasing focus on assessing the environmental impacts of

Table 4 Potential applications of approaches to finding surrogate data explored in this article

Application	*Scaled proxies*	*Direct proxies*	*Averaged proxies*	*Extrapolation from single parameters*	*Extrapolation with full parameters from agricultural statistics (with variability)*
Strategic planning	OK for initial screenings	OK for initial screenings	OK	OK	OK
Hotspot analysis and innovation	OK for initial screenings	OK for initial screenings	OK	OK	OK
Ingredient selection and product incremental changes	Not suitable	Not suitable	May be OK for changes between groups of ingredients (e.g., vegetables for meat)	OK	OK
Sourcing and supplier selection	Not suitable	Not suitable	Not suitable	May be OK if key influencing parameters are known for product type and supplier-specific parameters are used	May be OK if key influencing parameters are known for product type and supplier-specific parameters are used (rather than national statistics)
Labeling, Environmental Product Declarations, and external claims	Not suitable	Not suitable	Not suitable	Not suitable	Not suitable

bio-based products, and some of the approaches allowing quick access to data may be more or less appropriate depending on the intended application.

Scaled and direct proxies are the most uncertain of all the approaches reviewed here; however, they can still be informative in screening studies for strategic decisions and hotspot analysis to guide innovation. In addition to supporting decisions that only require orders of magnitude and trends, averaged proxies are useful for establishing environmental hotspots for future in-depth investigations to focus on these. They might help to reduce the list of thousands of ingredients or products to a few dozen, probably most relevant ones. It should be the goal in a next step to establish more accurate LCA data for those dozens (instead of probably unimportant ones). For less important ingredients, a direct proxy might be acceptable for many applications. Checking the developed proxy approaches against real, specific data also helps to establish the suitability of those approaches; however, this is not possible in many cases, as it is the lack of data that forces us to use proxies. It should not be the goal to keep forever the proxies that turn out to be relevant. Continuous improvement of data quality should be anticipated and planned for.

Averaged proxies may also be useful to select between different product types, depending on the overlap in variability ranges. For example, averaged proxies would probably support distinctions between animal and vegetable fats, although they would likely not suffice to distinguish one type of vegetable oil from another unless their impacts were clearly differentiated and the difference was clearly supported by enough studies. Nonetheless, 100% proxy data should not be used to select suppliers or sourcing regions, as the available data will usually not be sufficient to differentiate between options. With data extrapolation, conversely, it should be possible to construct new data sets that are representative enough of suppliers as long as the parameters driving environmental impacts (e.g., N fertilizer use) are known and values for such parameters are available for each supplier. In the extreme, this could be applied to select sourcing regions as long as statistical information is available on the relevant parameters (e.g., from FAO) and it can be incorporated in the calculations (e.g., with methods such as MEXALCA).

For other uses that require high precision (e.g., carbon labeling), the approaches presented here are not suitable, and more detailed assessments with study-specific data are recommended. Even for carbon labels, however, certain parts of the production system may be better represented by an extrapolated value with variability range than by a study-specific, well-characterized value. For example, it would be preferable to describe the commodity wheat with an extrapolated PDF representing the main growing regions in the world instead of a specific single-farm value from a very well-characterized study that is representative of less than 1% of the system suppliers. One additional note regarding carbon label application relates to the consideration of data variability: When such variability is incorporated (as shown, e.g., in figures 2, 3, and 4 and in the Knorr case study), one may question whether reporting results as a single number in carbon labels is appropriate.

Conclusions

As explored in this article, there are many approaches to bridging data gaps for bio-based products with surrogate data; these will continue to be used with the rise in LCA and, particularly, CF studies, and better understanding of their implications is needed. This article has defined some of the main approaches for bridging data gaps and has discussed their advantages and shortcomings. The obvious advantage common to all the surrogates is that they facilitate the impact estimation when a full assessment would be too costly (e.g., assessment of a whole product portfolio, diet, or sector). Conversely, surrogates are almost always wrong or, at the very least, poor approximations of reality, and the difficult question is how to determine whether the results are close to real data (and how much closer to, or perhaps further from, reality they are, as compared to leaving a data gap). An added difficulty is that it is often not possible to validate the results, for the same reason that makes surrogates necessary: lack of data. The hierarchy proposed in this article represents a first and important step in establishing guidelines for use of surrogate data and establishing a common vocabulary to allow an extension of the discussion.

The application of detailed environmental knowledge to life cycle management will probably generate more need for basic data, particularly if it keeps proving useful and robust enough to provide fit-for-purpose answers even when research moves into unexplored territories. The current situation suggests that there are more gaps than data, particularly for bio-based products. As a first step, before detailed knowledge is available, careful use of the approaches reviewed here could help support such application and guide research into major data gaps. The definitions and discussion provided in this article could help practitioners to describe better what data are being used in their studies and their potential implications.

Extensive data are available from agricultural statistics on key production parameters that drive environmental impacts; researchers should further exploit such data sources to derive expected values for a range of situations starting from detailed LCA studies (data extrapolation). This seems to be a more promising approach than using direct proxy data. When not enough resources are available (time, budget, or technical expertise) the use of averaged proxies with consideration of variability may be a good compromise for initial assessments and even for guiding broad product group selection (e.g., vegetable oils versus animal fats). Direct proxies with no or low variability may still be informative of the expected trends and orders of magnitude in environmental impacts, which are enough for many applications (e.g., strategic planning, identification of hotspots). Caveats should be placed, however, when common sense or rules of thumb are used to bridge data gaps between products or regions with direct proxies. For example, Swiss wheat may be a better proxy for Swiss barley than is French barley; chicken data are not a good proxy for turkey, even though both are poultry. Initial explorations are required by product experts who understand the key parameters driving the environmental impacts of a product life cycle and are able to make such rules of thumb.

The use of surrogate data as explored here emphasizes even more the unavoidably uncertain and incomplete nature of LCA results. More research is needed, in particular on the consideration of variability for bio-based ingredients. Ideally, proxy data should only be used with proper consideration of the potential effects of variability, and refinement of results should be planned for those products or components represented by proxies that influence the overall environmental impact. Once the variability and uncertainty around the data used are factored into the analysis, the use of single numbers for communication of environmental impacts (e.g., in carbon labels) seems inadequate, even when surrogates are not used and all data are specific to the study. Data expressing variability and uncertainty around a mean are preferable.

In conclusion, assessment of carbon footprints (among other environmental impacts) and business-to-business communication of results will remain positive drivers for environmental improvements (and LCA applications). It is therefore important that the analysis and reported results include proper consideration of data availability, variability, and uncertainty. This will help not only to obtain more reliable results but also to improve the trust in LCA and related approaches.

Acknowledgements

This article was initiated in a workshop organized by the Sustainable Consumption Institute (SCI) at the University of Manchester and sponsored by Unilever. The article has benefited from the authors' discussions with the other workshop participants, including Simon Aumônier, Claudine Basset-Mens, Gareth Edwards-Jones, Rolf Frischknecht, and Ulf Sonesson. The comments from three anonymous reviewers have helped to improve the clarity and focus of the article and are gratefully acknowledged.

Notes

1. For the purposes of this article, bio-based products are those derived from biotic systems, including agriculture, forestry, and fisheries.
2. One kilogram (kg, SI) $\approx$ 2.204 pounds (lb).
3. One tonne (t) = 10^3 kilograms (kg, SI) $\approx$ 1.102 short tons.

References

ABC (Agro Business Consultants). 2009. *The agricultural budgeting and costing book*. 63rd edition. Melton Mobray, UK: ABC.

Audsley, E., M. Brander, J. Chatterton, D. Murphy-Bokern, C. Webster, and A. Williams. 2010. *An assessment of greenhouse gas emissions from the UK food system and the scope for reduction by 2050: How low can we go?* London: World Wildlife Fund and Food Climate Research Network.

Azapagic, A., H. Stichnothe, and N. Espinoza-Orias. 2011. Sustainability issues in food provisioning systems. In *Sustainable development in practice: Case studies for engineers and scientists*, second edition, edited by A. Azapagic and S. Perdan. Chichester, UK: Wiley.

Basset-Mens, C., S. Ledgard, and M. Boyes. 2009. Eco-efficiency of intensification scenarios for milk production in New Zealand. *Ecological Economics* 68(6): 1615–1625.

Baumgartner, D. U., L. de Baan, T. Nemecek, and F. Pressenda. 2008. Life cycle assessment of feeding livestock with European grain legumes. In *Proceedings of the 6th international conference on LCA in the agri-food sector*. Zürich, Switzerland: Agroscope Reckenholz-Taenikon Research Station.

Capper, J. L., R. A. Cady, and D. E. Bauman. 2009. The environmental impact of dairy production: 1944 compared with 2007. *Journal of Animal Science* 87(6): 2160–2167.

CCaLC. 2010. CCaLC software and database. Manchester, UK: Sustainable Consumption Institute. www.ccalc.org.uk. Accessed 12 December 2010.

Cederberg, C. and A. Flysjö. 2004. *Life cycle inventory of 23 dairy farms in south-western Sweden*. Gothenburg, Sweden: SIK, the Swedish Institute of Food and Biotechnology.

Cederberg, C. and B. Mattson. 2000. Life cycle assessment of milk production: A comparison of conventional and organic farming. *Journal of Cleaner Production* 8(1): 49–60.

Cederberg, C., U. Sonnesson, M. Henriksson, V. Sund, and J. Davis. 2009. *Greenhouse gas emissions from Swedish production of meat, milk and eggs 1990 and 2005*. SIK Report 793. Gothenburg, Sweden: SIK, the Swedish Institute of Food and Biotechnology.

Defra (Department for the Environment, Food and Rural Affairs). 2009. *Scenario building to test and inform the development of a BSI method for assessing greenhouse gas emissions from food*. Project Reference No. FO0404. ADAS. London: Defra.

Edwards-Jones, G., K. Plassmann, and I. M. Harris. 2009. Carbon footprinting of lamb and beef production systems: Insights from an empirical analysis of farms in Wales, UK. *Journal of Agricultural Science* 147: 707–719.

Eide, M. H. 2002. Life cycle assessment (LCA) of industrial milk production. *International Journal of Life Cycle Assessment* 7(2): 115–126.

Gerber, P., T. Vellinga, C. Opoio, B. Henderson, and H. Steinfield. 2010. *Greenhouse gas emissions from the dairy sector: A life cycle assessment*. Rome: Food and Agriculture Organization of the United Nations.

Guinard, C., F. Verones, and Y. Loerincik. 2009. *Environmental/ecological impact of the dairy sector system*. Report 436/2009. Lausanne, Switzerland: International Dairy Federation.

Haas, G., F. Wetterich, and U. Kopke. 2001. Comparing intensive, extensified and organic grassland farming in southern Germany by process life cycle assessment. *Agriculture, Ecosystems and Environment* 83: 43–53.

Jungbluth, N. 2000. Umweltfolgen des Nahrungsmittelkonsums: Beurteilung von Produktmerkmalen auf Grundlage einer modularen Oekobilanz. [Environmental consequences of food consumption: A modular life cycle assessment to evaluate product characteristics.] PhD thesis No. 13499, Federal Institute of Technology Zurich, Zurich, Switzerland.

Milà i Canals, L., G. M. Burnip, and S. J. Cowell. 2006. Evaluation of the environmental impacts of apple production using life cycle assessment (LCA): Case study in New Zealand. *Agriculture, Ecosystems and Environment* 114: 226–238.

Milà i Canals, L., S. J. Cowell, S. Sim, and L. Basson. 2007. Comparing domestic versus imported apples: A focus on energy use. *Environmental Science and Pollution Research* 14(5): 338–344.

Milà i Canals, L., J. Chenoweth, A. K. Chapagain, S. Orr, A. Antón, and R. Clift. 2009a. Assessing freshwater use impacts in LCA Part I: Inventory modelling and characterisation factors for the main impact pathways. *International Journal of Life Cycle Assessment* 14(1): 28–42.

Milà i Canals, L., S. Sim, G. Neuer, P. Vermeiden, V. Markl-Moser, and C. Kerr. 2009b. The global carbon footprint of Knorr: Assessing one of the world's largest food brands. Paper presented at the fourth International Conference on Life Cycle Management: The Global Challenge of Managing Life Cycles, 6–9 September, Cape Town, South Africa.

Milà i Canals, L., S. Sim, T. Garcia-Suarez, G. Neuer, K. Herstein, C. Kerr, G. Rigarlsford, and H. King. 2011. Estimating the greenhouse gas footprint of Knorr. *International Journal of Life Cycle Assessment* 16(1): 50–58.

Mouron, P., T. Nemecek, R. W. Scholz, and O. Weber. 2006. Management influence on environmental impacts in an apple production system on Swiss fruit farms: Combining life cycle assessment with statistical risk assessment. *Agriculture, Ecosystems and Environment* 114: 311–322.

Muñoz, I., L. Milà i Canals, and A. Fernández-Alba. 2010. Food from farm to sewer: Life cycle assessment of the average Spanish diet including the human excretion stage. *International Journal of Life Cycle Assessment* 15(8): 794–805.

Mutel, C., F. Stössel, and S. Hellweg. 2009. *Sensitivity-based methods for estimating LCA results of selected bio-based ingredients*. ETH Report for Unilever. Zurich, Switzerland: Federal Institute of Technology Zurich.

Nemecek, T., A. Roches, F. Hayer, D. Baumgartner, G. Gaillard, S. Sim, and L. Milà i Canals. 2009a. Geographical extrapolation of crop life cycle inventories and impacts. In *LCAIX 2009: Toward the global life cycle economy*. Boston: American Center of Life Cycle Assessment.

Nemecek, T., A. Roches, and R. Freiermuth. 2009b. Multivariate analysis of similarities between environmental profiles of crops. Poster presented at the fourth International Conference on Life Cycle Management: The Global Challenge of Managing Life Cycles, 6–9 September, Cape Town, South Africa.

Nix, J. S. 2009. *Farm management pocketbook*. 40th (2010) edition. London: Andersons Centre, Imperial College London, Wye Campus.

Pelletier, N. 2008. Environmental performance in the US broiler poultry sector: Life cycle energy use and greenhouse gas, ozone depleting, acidifying and eutrophying emissions. *Agricultural Systems* 98(2): 67–73.

Roches, A., T. Nemecek, G. Gaillard, K. Plassmann, S. Sim, H. King, and L. Milà i Canals. 2010. MEXALCA: A modular method for the extrapolation of crop LCA. *International Journal of Life Cycle Assessment* 15(8): 842–854.

Röös, E., C. Sundberg, and P.-A. Hansson. 2010. Uncertainties in the carbon footprint of food products: A case study on table potatoes. *International Journal of Life Cycle Assessment* 15:478–488.

Santacana, M., J. Pon, D. Pon, I. Arto, and S. Casanovas. 2008. Greenhouse gas emissions from a consumption perspective in a global economy. Opportunities for the Mediterranean region. In *CP/RAC Annual technical publication*, volume 7:101–112. Barcelona, Spain: Regional Activity Centre for Cleaner Production.

Sim, S. 2006. Sustainable food supply chains. Eng.D. thesis portfolio, University of Surrey, Guildford, UK.

Stichnothe, H., A. Hospido, and A. Azapagic. 2008. Carbon footprint estimation of food production systems: The importance of considering methane and nitrous oxide. *Aspects of Applied Biology* 87:65–75.

Stichnothe, H., A. Hospido, and A. Azapagic. 2010. *Environmental impacts of whole meals: The case study of a Christmas meal*. Manchester, UK: CCaLC Project, University of Manchester.

Thoma, G., J. Popp, D. Nutter, D. Shonnard, R. Ulrich, M. Matlock, D. S. Kim, Z. Neiderman, N. Kemper, C. East, and F. Adom. 2010. Greenhouse gas emissions from production of fluid milk in the United States: Executive summary. Fayetteville, AK: University of Arkansas www.usdairy.com/Public%20Communication%20Tools/Carbon%20Footprint%20Study%20of%20Fluid%20Milk/CARBON%20FOOTPRINT%20STUDY%20EXEC%20SUMMARY.pdf. Accessed 18 July 2011.

Tzilivakis, J., D. J. Warner, M. May, K. A. Lewis, and K. Jaggard. 2005. An assessment of the energy inputs and greenhouse gas emissions in sugar beet (*Beta vulgaris*) production in the UK. *Agricultural Systems* 85: 101–119.

Wallén, A., N. Brandt, and R. Wennersten. 2004. Does the Swedish consumer's choice of food influence GHG emissions? *Environmental Science & Policy* 7: 525–535.

Weidema, B. P., M. Thrane, P. Christensen, J. Schmidt, and S. Løkke. 2008. Carbon footprint: A catalyst for life cycle assessment? *Journal of Industrial Ecology* 12(1): 3–6.

Williams, A. G., E. Audsley, and D. L. Sandars. 2006. *Determining the environmental burdens and resource use in the production of agricultural and horticultural commodities*. Report to Defra on Project IS0205. Cranfield, UK: Cranfield University.

About the Authors

Llorenç Milà i Canals is a sustainability scientist at Unilever in Sharnbrook, UK. **Adisa Azapagic** is a professor at the University of Manchester in Manchester, UK. **Gabor Doka** is a consultant at Doka Life Cycle Assessment in Zürich, Switzerland. **Donna Jefferies** is a sustainability scientist at Unilever in Sharnbrook, UK. **Henry King** is a sustainability science and technology leader at Unilever in Sharnbrook, UK. **Christopher Mutel** is a researcher at ETH—Swiss Federal Institute of Technology in Zürich, Switzerland. **Thomas Nemecek** is a deputy research group leader at ART—Agroscope Reckenholz-Tänikon Research Station in Zürich, Switzerland. **Anne Roches** was a researcher at ART in Zürich, Switzerland, at the time the article was written; she is currently a researcher at ETH in Zürich, Switzerland. **Sarah Sim** is a sustainability scientist at Unilever in

Sharnbrook, UK. **Heinz Stichnothe** was a researcher at the University of Manchester (UK) at the time this article was written; he currently works at the Institute of Agricultural Technologies and Biosystems Engineering in Braunschweig, Germany. **Greg Thoma** is a professor at the University of Arkansas in Fayetteville, Arkansas, USA, and is research director of the Sustainability Consortium at the University of Arkansas; **Adrian Williams** is a principal research fellow at Cranfield University in Cranfield, UK.

Supporting Information

Supporting information may be found in the online version of this article:

Supporting Information S1: The supporting information provides additional data and worked-out examples illustrating the contents of this article. First, concept definitions, further production parameter data (table S1), assumptions, and a table showing the step-by-step calculations (table S2) of the simplified extrapolation example (from chicken to turkey) are provided. Then, an alternative representation of figure 4 is given in figure S1.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.