

Life cycle inventory data quality issues for bioplastics feedstocks

Alix Grabowski · Susan E. M. Selke · Rafael Auras ·
Martin K. Patel · Ramani Narayan

Received: 3 June 2014 / Accepted: 12 January 2015 / Published online: 27 January 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Purpose Bioplastics are a growing field, but with their expansion come unique environmental issues associated with the cultivation and processing of feedstocks. Availability of appropriate, high-quality data is a problem in life cycle assessment (LCA) of biopolymers and other bio-based materials that limits the accuracy and usefulness of study results. It is therefore critical that these data gaps be closed. To determine what data is needed to close these gaps, this study reviews currently available life cycle inventory data for biopolymer feedstocks and assesses the data quality for the selected feedstocks of corn, sugarcane, and soy.

Methods Life cycle inventory databases and relevant publications were searched for appropriate data, and the results collected into a summary table. The quality review was conducted using a pedigree matrix type scoring system which was adapted from the ILCD handbook, and an overall quality score for each dataset was calculated based on the matrix scores.

Results A total of 287 datasets were collected during the review for a total of 22 different feedstocks. The majority of these datasets are from Europe and the USA, with most of Asia, the Middle East, and Africa having very limited data available.

Conclusions From the quality analysis, it was determined that more datasets that capture regional variations in crop cultivation are needed, as well as more data on land use change.

Keywords Bio-based polymers · Corn · Data quality · Feedstocks · LCA · Soy · Sugar cane

Responsible editor: Thomas Koellner

Electronic supplementary material The online version of this article (doi:10.1007/s11367-015-0853-3) contains supplementary material, which is available to authorized users.

A. Grabowski · S. E. M. Selke (✉) · R. Auras
School of Packaging, Michigan State University, East
Lansing, MI 48824-1223, USA
e-mail: sselke@msu.edu

M. K. Patel
Institute for Environmental Sciences and Forel Institute, University
of Geneva, Energy Group, Route de Drize 7, 1227 Carouge/
Geneva, Switzerland

R. Narayan
Chemical Engineering and Materials Science, Michigan State
University, East Lansing, MI 48824-1223, USA

1 Introduction

Life cycle assessment (LCA) is an objective method of evaluating the sustainability of a system and assessing its impacts on the environment (ISO 14040 2006). One of the strengths of LCA is its ability to quantify and categorize the individual impacts of a system and facilitate comparisons. It is an effective tool for this purpose because of the level of detail and complexity it is able to convey; but like any method, it has its limitations. The quality of information available from an LCA study is largely limited by the quality and completeness of the data on which it is built. Without complete, high-quality data, one cannot be confident that the results of an LCA study accurately represent the situation. Therefore, the quality of the data used in a study affects the quality of the recommendations that come out of the study (ISO 14044 2006).

Bioplastics and biomaterials are a growing field, often heralded as a sustainable alternative to traditional products; yet their environmental impacts remain uncertain and variable. This is perhaps even truer for a related product—biofuels. In a 2012 comparative analysis, Álvarez-Chávez

et al. (2012) state “A bio-based plastic is not necessarily a sustainable plastic; this depends on a variety of issues, including the source material, production process, and how the material is managed at the end of its useful life.” In other words, to correctly assess a biopolymer’s environmental footprint one must have accurate and representative data about many aspects of the polymer’s life cycle, including the production of that polymer. A particular challenge of LCA studies of biopolymers is understanding the impact of growing the polymer feedstock since, as discussed by Nemecek et al. (2012), “Environmental impact data for crops in the literature and the LCA databases are scarce.” Since bioplastics and biofuels are often derived from the same feedstocks, one can infer that they likely suffer from the same data quality issues as bioplastics. However, while there is a significant overlap in the cultivation and processing phases for bioplastics and biofuels, this critical data quality analysis is focused on bioplastics. Therefore, there may be other significant quality issues in biofuels data that are not captured here.

This means LCA data for biopolymers must be obtained; but what type of data should be collected? Before more information can be sought out, it is necessary to identify what types of data are most needed. Identifying critical data gaps for LCA of biopolymers will provide a foundation for future research. By indicating what areas are weak in data availability, researchers can identify the types of data that should be focused on when collecting primary data. Understanding the quality and completeness of currently available data also has the potential to clarify the strengths and limitations of existing studies that use this data, as it may shed new light on how representative the study is of the system it is trying to model.

The first step in the quest to solve data issues in LCA is addressed by van der Voet et al. (2010) in a review of the state of LCA for the related field of biofuels: “The way forward to remedy data problems is clear: identify data needs, collect more and better data and make them accessible.” A review of the state of currently available LCA data is the first step to improving the quantity and quality of available data, because without being aware of what is already published it is impossible to know where future efforts in data acquisition should be focused. The goal of this project was to identify the critical data gaps present in LCA data for biopolymer feedstocks in order to illuminate the next steps in the path to more sustainable bioplastics.

2 Methods

In order to identify critical data gaps for bio-based plastics, a review of currently available data was conducted. Information was collected from the LCA databases Ecoinvent, GaBi-PE (Ecoinvent data modified by GaBi), USLCI, LCA Food DK, and LCA Commons. The

software programs SimaPro and GaBi were used to access all of these databases except LCA Commons, which has not been integrated into the software programs. The websites of LCA Food DK and USLCI were also searched for information. Additionally, publications were searched for relevant data that had not yet been integrated into the databases. Since it is not feasible to search all publications, a critical selection of publications was chosen based on the likelihood that they would contain relevant information. Table 1 contains the list of publications and timeframes searched.

All data pertaining to feedstocks that are viable for use in the production of bio-based plastics were considered in the search; in total, 287 datasets were collected during the review for a total of 22 different feedstocks. These data relate to many aspects of bioplastic production, including raw agricultural feedstocks (e.g., corn, soybeans) through processing steps (e.g., sugar, oil), platform chemical production (e.g., ethanol, biodiesel), and polymer production.

Because of the large number of data sources amassed, it was necessary to narrow the focus of the project to conduct a deeper analysis of the data. Corn, sugarcane, and soy were the feedstocks selected to be analyzed in more detail. These feedstocks were chosen because they are the most common feedstocks used to produce both bioplastics and biofuels (e.g., Evans 2010; Smith 2013). The first step was to conduct a completeness check on the available data for these crops. Information from ISO 14044 (2006) and the ILCD handbook (EC 2010) was used to inform this process. The completeness check involved searching each data file in detail and noting what inputs were considered and, when possible, how they were accounted for. For the database files, this meant looking through the data directly, at the flow level. For the data contained in journal articles, it involved searching for information within the article and looking up data source information from the references. An effort was made to include all significant input categories for these feedstocks in the check. The input categories included are as follows: carbon

Table 1 List of publications searched

Publication name	Time frame
Journal of Cleaner Production	2013–2005
International Journal of LCA	2012–2000
Bioresource Technology	2012–2005
Biomass and Bioenergy	2012–2005
Environmental Science and Technology	Selected
Packaging Technology and Science	Selected
Journal of Industrial Ecology	2012–2008
Sustainability	2012–2005
Science	2012–2008
PNAS	2012–2005

sequestration, seed production, soil preparation, transport of materials to the farm, fuel used on the field, power for farm activities, machines, machine shelter, sowing, tilling, fertilizer, pesticides, herbicides, lime, crop residue management, irrigation, harvesting, grain drying, direct field emissions, crop storage, land occupation, and land use change.

Not all inputs were relevant to all datasets. For example, a study that only considered GHG emissions would not include water use. It was noted in the completeness check results when an input was specifically excluded from a study because it was not within the study scope.

Next, a pedigree matrix scoring system was used to evaluate the datasets for corn, soy, and sugarcane. The data were evaluated on their technological, geographical, and temporal representativeness on a scale from 1–5. When information was available, data uncertainty was also considered in the evaluation. From this evaluation and the completeness evaluation, a data quality rating (DQR) was calculated. Agricultural data was considered separately from processing data. Some datasets therefore have two DQRs, one for agricultural data quality and one for processing data quality. The evaluation criteria and the formula used to calculate the overall score were adapted from the ILCD handbook (EC 2010) and van der Berg et al. (1999). Table 2 explains the different categories considered in the evaluation of the data, and Tables 3 and 4 explain the ranking of the quality ratings.

It was necessary to consider the quality ranking recommendations of both the ILCD (EC 2010) and van der Berg et al.

(1999) to complete a consistent analysis. The ILCD rankings are described in general terms and do not have category-specific requirements. Therefore, the criteria put forth by van der Berg et al. were used to supplement the ILCD recommendations. Specific category ranking requirements from van den Berg et al. were applied. For example, data that was less than 3 years older than the study date was given a score of 1 in temporal correlation. It should be noted that the category “reliability” from van den Berg et al. and the category “precision/uncertainty” from the ILCD handbook are equivalent. The ILCD handbook also includes another evaluation category, methodological appropriateness and consistency, which was not used in this evaluation. This category was excluded because it is dependent on the goal and scope of the intended application of the dataset.

Data was examined in the greatest detail possible when scoring each category since most datasets have multiple sources of information that can be of varying quality. These differences are accounted for by scoring each source of data and then resolving their respective scores into a single category score. Background process data were weighted less heavily when resolving scores, as were data that are not sensitive to a specific category. For example, using a US electricity mix for sugar production in Brazil is a significant difference because Brazilian sugarcane processing is generally powered by the burning of bagasse, a sugarcane co-product. This substitution therefore represents a significant technological difference. However, if US diesel tractor emissions were substituted for Brazilian diesel tractor emissions, the difference is not as significant since the technology is nearly identical.

Additionally, half scores were given to categories when deemed appropriate. For example, a dataset published in 2010 for which half of the relevant data was from 2009 (less than 3 years difference) and half was from 2005 (less than 6 years difference) would be given an overall score of 1.5 in the category of temporal correlation. The individual justifications for scores are available in the report “LCA studies for bio-based plastics: Closing critical data gaps and agreeing on representative generic datasets” (Patel et al. 2014).

Equation 1 was used to calculate the DQR for each dataset (EC 2010). Note that the lowest criteria score is weighted in the formula by fivefold. This is done because the weakest quality indicator significantly weakens the overall quality of the dataset being evaluated. The Precision (P) quality indicator was only included in the calculation of the DQR for Ecoinvent data. This was necessary due to the lack of information about precision for most datasets. When evaluating the precision of the Ecoinvent data, the Ecoinvent uncertainty scoring criteria were used to interpret the uncertainty scores for each flow (Ecoinvent 2007). Then, a

Table 2 Definition of evaluation categories adapted from the ILCD handbook (EC 2010)

Indicator/component	Definition/comment
Technological representativeness (TeR)	“Degree to which the data set reflects the true population of interest regarding technology, including for included background data sets, if any.”
Geographical representativeness (GR)	“Degree to which the data set reflects the true population of interest regarding geography, including for included background data sets, if any.”
Time related representativeness (TiR)	“Degree to which the data set reflects the true population of interest regarding time/age of data, including for included background data sets, if any.”
Completeness (C)	“Share of (elementary) flows that are quantitatively included in the inventory. Note that for product and waste flows this needs to be judged on a system’s level.”
Precision/uncertainty (P)	Measure of the variability of the data values for each data expressed (e.g., low variance = high precision). Note that for product and waste flows, this must be judged on a system level.

Table 3 Quality rating definitions adapted from the ILCD handbook (EC 2010)

Quality level	Quality rating	Definition
Very good	1	“Meets the criterion to a very high degree, having no relevant need for improvement. This is to be judged in view of the criterion’s contribution to the data set’s potential overall environmental impact and in comparison to a hypothetical ideal data quality.”
Good	2	“Meets the criterion to a high degree, having little yet significant need for improvement. This is to be judged in view of the criterion’s contribution to the data set’s potential overall environmental impact and in comparison to a hypothetical ideal data quality.”
Fair	3	“Meets the criterion to still sufficient degree, having the need for improvement. This is to be judged in view of the criterion’s contribution to the data set’s potential overall environmental impact and in comparison to a hypothetical ideal data quality.”
Poor	4	“Does not meet the criterion to a sufficient degree, having the need for relevant improvement. This is to be judged in view of the criterion’s contribution to the data set’s potential overall environmental impact and in comparison to a hypothetical ideal data quality.”
Very poor	5	“Does not at all meet the criterion, having the need for very substantial improvement. This is to be judged in view of the criterion’s contribution to the data set’s potential overall environmental Impact and in comparison to a hypothetical ideal data quality.”
Not applicable	0	Criteria could not be applied

single uncertainty score was chosen for the file based on this interpretation.

A dataset with DQR less than or equal to 1.6 is considered high quality, while a dataset with a DQR between 1.6 and 3 is considered to be of basic quality. Any dataset with a DQR between 3 and 4 is considered to be an estimate (EC 2010).

DQR formula and definitions modified from the ILCD handbook

$$DQR = \frac{TeR + GR + C + P + 4X_w}{i + 4} \tag{1}$$

Table 4 Pedigree matrix data quality rating level definitions adapted from van den Berg et al. (1999)

Indicator score	1	2	3	4	5
Reliability	Verified data based on measurements	Verified data partly based on assumptions or non-verified data based on measurements	Non-verified data partly based on assumptions	Qualified estimate (e.g., by industrial expert)	Non-qualified estimate
Completeness	Representative data from a sufficient sample of sites over an adequate period to even out normal fluctuations	Representative data from a smaller number of sites but for adequate periods	Representative data from adequate number of sites but from shorter periods	Representative data but from a smaller number of sites and shorter periods or incomplete data from an adequate number of sites and periods	Representativeness unknown or incomplete data from a smaller number of sites and/or from shorter periods
Temporal correlation	Less than 3 years difference to year of study	Less than 6 years difference	Less than 10 years difference	Less than 15 years difference	Age of data unknown or more than 15 years difference
Geographical correlation	Data from area under study	Average data from larger area in which the area under study is included	Data from area with similar production conditions	Data from area with slightly similar production conditions	Data from unknown area or area with very different production conditions
Technological correlation	Data from enterprises, processes, and materials under study	Data for processes and materials under study but from different enterprises	Data from processes and materials under study but from different technology	Data on related processes or materials but same technology	Data on related processes or materials but different technology

DQR	Data quality rating of the LCI dataset
TeR	GR, TiR, C, P: see Table 2
X_w	Weakest quality level obtained (i.e., highest numeric value) among the data quality indicators
i	Number of applicable data quality indicators

3 Results

3.1 Quantitative analysis

Figure 1 describes the collected datasets. Of the 287 datasets found, 110 are for raw agricultural feedstocks, 114 are for processing of agricultural feedstocks, and 56 are for production of platform chemicals (mainly ethanol and biodiesel). Only seven datasets were found for the complete production of biopolymers.

The feedstock with the largest amount of datasets is wood, with a total of 60. However, this is largely because the same base data is available in multiple iterations. For example, data for board trimmings, sawdust, and woodchips in the US Pacific Northwest are all available and build from the same base data. Therefore, the high total is somewhat misleading. Agricultural residue is the feedstock with the second highest amount of datasets collected at 42. This category includes things like corn stover, sugarcane bagasse, and wheat straw. This number is almost exclusively made up of Ecoinvent files. Ecoinvent has separated the data associated with the co-products of crop production into separate files, which means that for each crop Ecoinvent has data available, it also has data for the associated agricultural residue. Here, all of these residues are collected into a single category, which is why they are so numerous.

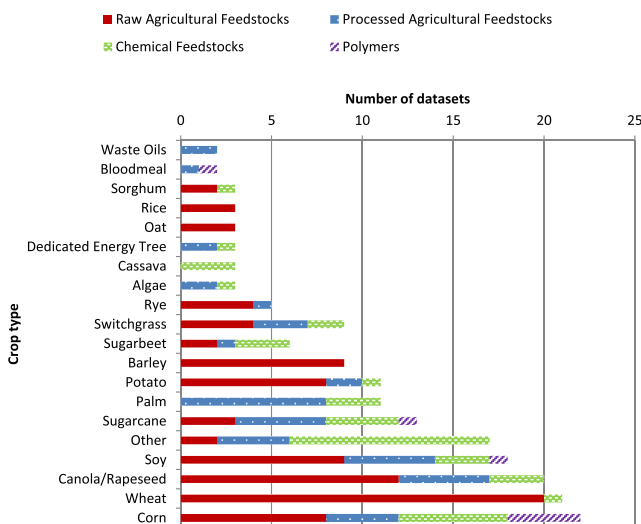


Fig. 1 Summary of data collected by crop and category. Note: wood is excluded from the figure for reasons of convenience

The crop with the next largest amount of available data is corn (22), which is closely followed by wheat (21), rapeseed (20), and soy (18). The category “other” is mostly composed of data for ethanol from mixed feedstocks, which is available from Ecoinvent, and generic data for unspecified biomass. Sugarcane (13) has slightly less data available in relation to the other commodity crops commonly used to produce biopolymers. It should also be noted that over half (4 of 7) of the polymer datasets are based on corn.

The table and figure show that eight datasets are available for raw corn. However, the 131 datasets available for corn from the US database LCA Commons are represented by a single entry in the summary table. This was done for reasons of practicality, since datasets in LCA Commons are state specific for a single harvest year. The short data collection time period is a quality issue, but 18 of the 131 datasets are aggregated over multiple years. These aggregated datasets are still state specific; LCA Commons does not have a file meant to represent averages for the entire USA. LCA Commons’ datasets are also included for soy (137), oats (12), rice (6), and wheat (155). They are represented in the table as described above.

Geographically, the data is skewed to Europe and North America with about 60 % of the datasets (175 of 287) from one of the two regions (105 for Europe and 70 for North America). Additionally, about half of the European data (58 of 106) is for Switzerland. This is likely related to the fact that the Ecoinvent project is based in Switzerland. Likewise, nearly all of the North America data is for the USA. Only one of the 70 datasets for North America is explicitly for another country, a study about biodiesel production in Costa Rica. There are a number of USLCI datasets whose region is described in general as “North America” but these files are more reflective of the USA than of North American averages, being mainly built from US data and modeling technology typical of the US. Figure 2 describes the geographic distribution of the datasets.

As illustrated by Fig. 2, the most obvious geographical data gap is that zero datasets were found that represent anywhere on the continent of Africa. Additionally, the category “Asia and the Middle East” covers a very large amount of area in

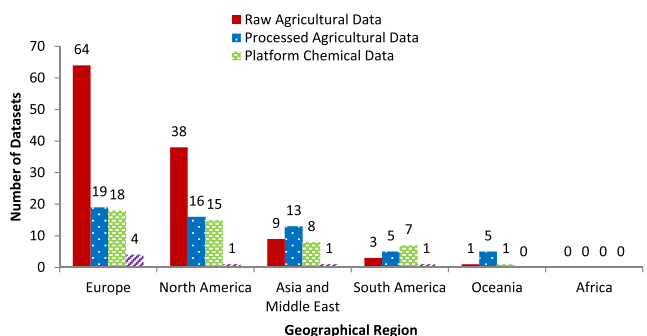


Fig. 2 Geographical distribution of datasets

theory, but in reality the datasets are focused on Southeast Asia and China. Only one dataset, potato production in Iran, breaks this pattern. The data designated Oceania is composed of one dataset for New Zealand with the remainder representing Australia. The South American data consists of entries from Brazil and Argentina and is mostly for sugarcane and soybeans. In general, the geographical concentration of the datasets is in line with the demand for LCA data in each region. It makes sense that Europe and the USA, which both have strong policies in place that promote biofuels and LCI database projects, have more data available. Furthermore, the pattern of data concentration is consistent with areas that produce large amounts of commodities that are commonly used to produce biofuels and biopolymers.

A notable exclusion is that neither Canada nor Russia is represented in the datasets. The two largest countries (geographically) in the world, Canada and Russia are both also major producers of grains that are commonly used to produce biodiesel and ethanol. Canada is consistently the largest producer of canola (Canola Council of Canada 2011), and Russia was the third largest producer of wheat in 2013 (USDA 2014). The largest producer of wheat in 2013, India, is also underrepresented in data availability (USDA 2014). There are four datasets for India, but none of them are for wheat. Three are for agricultural residues suitable for cellulosic ethanol production (jute and kenaf stalks), and the fourth is for jatropha (an oilseed crop that can be grown on degraded land) (Arvidsson et al. 2011).

3.2 Qualitative analysis

A data quality evaluation was undertaken for cultivation of corn, soy, and sugarcane. Figure 3 illustrates the composition of the corn DQRs by breaking them down into their constituent categories of completeness, uncertainty, and technological, geographical, and temporal representativeness. Figures 4 and 5 illustrate the equivalent for sugarcane and soy, respectively. Tables 5, 6, and 7 match the identifiers in the figures with the corresponding dataset names for each feedstock.

Overall, the DQRs for the corn datasets ranged from 1.4 to 2.6, with five of the datasets ranked in the high-quality range and the remaining eight datasets in the basic quality range. Selective comparisons of two corn datasets (i.e., corn, production average, US 2012 (NREL 2012) (N) DQR=2.6 and corn, at farm (Ecoinvent 2013) (G) DQR=1.7) were conducted using ReCiPe Midpoint (H) V1.09/World ReCiPe H impact assessment method. Figure 6 shows that a large difference is encountered between the two selected datasets which results in a large difference in many ReCiPe midpoint impact indicators. It is important to notice the lack of numerical values for freshwater eutrophication, ionizing radiation, agricultural and urban land occupation, natural land transformation, and metal depletion for the NREL 2012 dataset.

The sugarcane datasets ranged from 1.8 to 3.7, with seven of the datasets in the basic-quality range and the remaining two classified as estimates. Sugarcane has no datasets with a DQR in the high-quality range. These less

Fig. 3 Corn data quality ratings and component scores by dataset

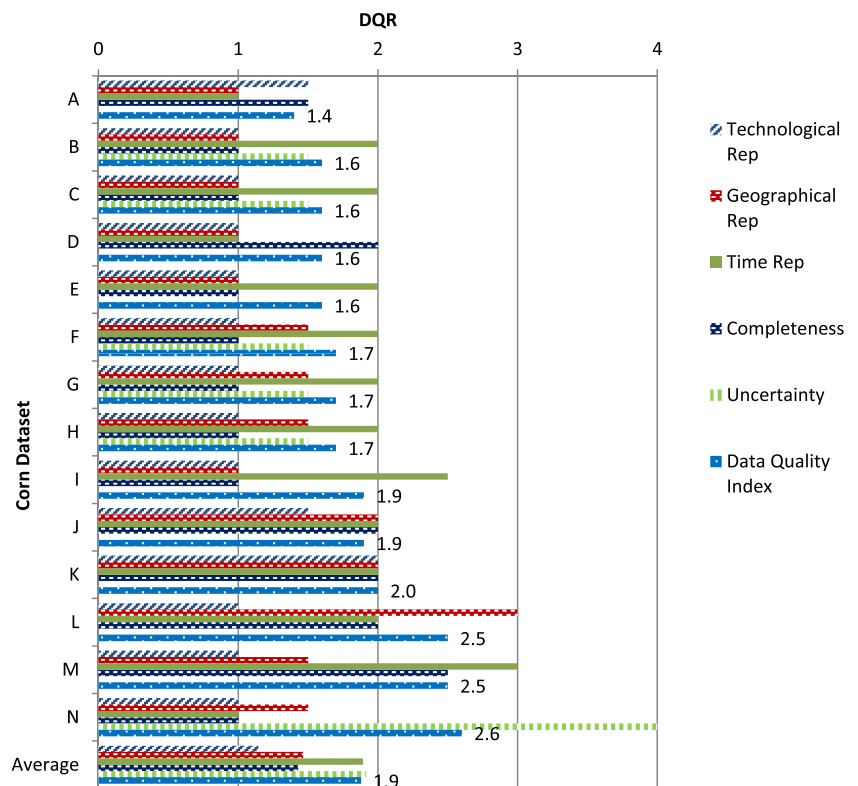
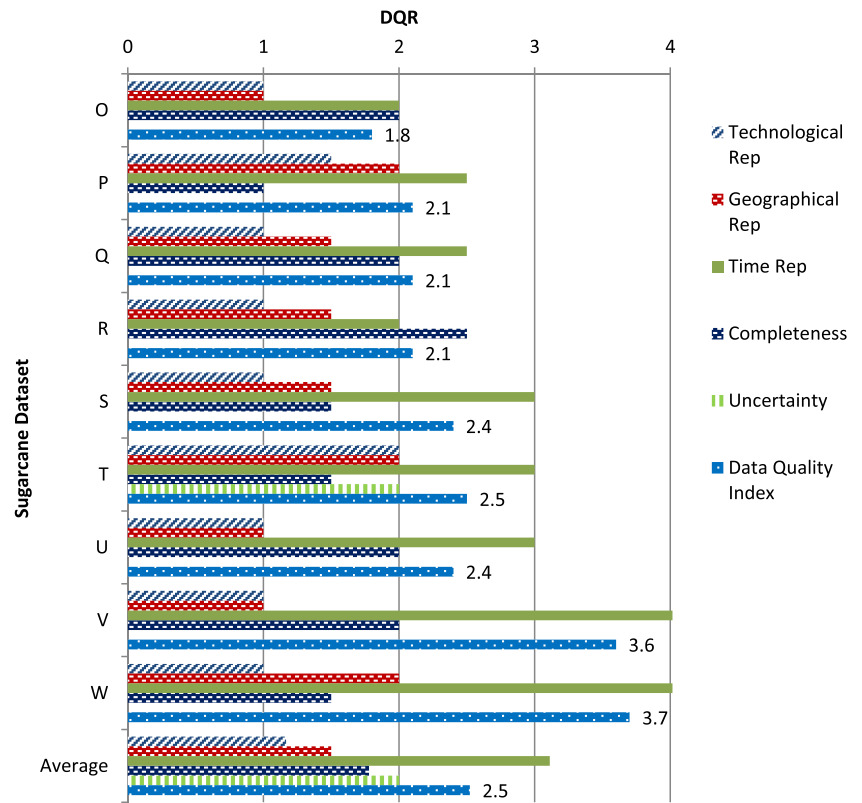


Fig. 4 Sugarcane data quality ratings and component scores by dataset



desirable DQRs for sugarcane are mainly a result of poorer scores in the category of temporal

representativeness. The soy datasets have the largest range of DQRs, from 1.6 to 4.5. Two of the datasets are

Fig. 5 Soy data quality ratings and component scores by dataset

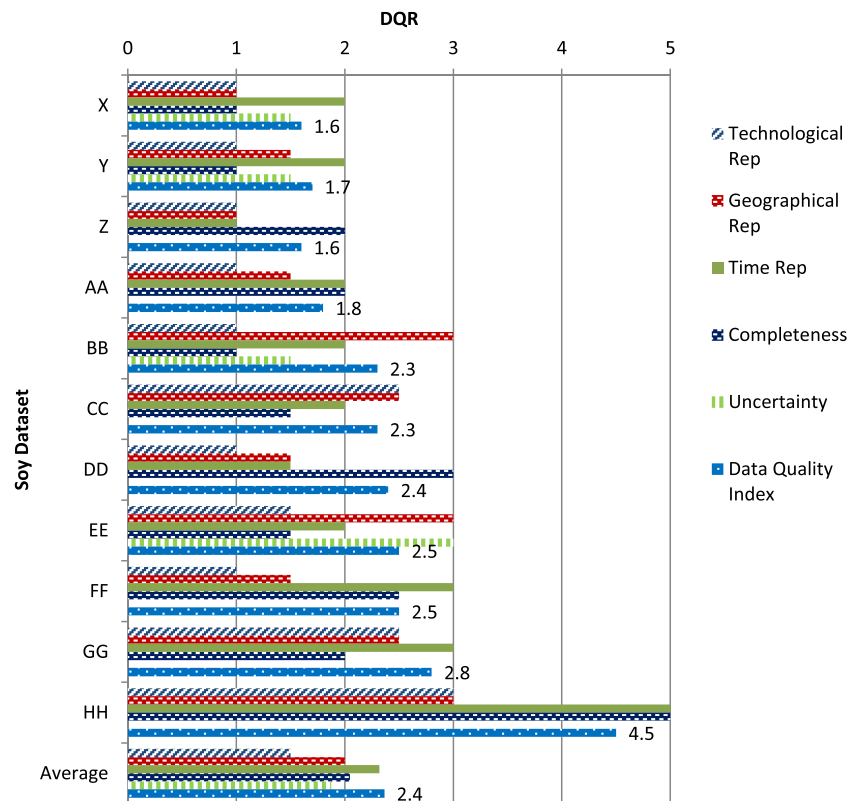


Table 5 Corn dataset names and identifiers

Corn dataset name	Identifier
Annual report: life cycle assessment to improve the sustainability and competitive position of biobased chemicals: a local approach (Kim and Dale 2009a)	A
Grain maize organic, at farm (Ecoinvent 2013)	B
Silage maize organic, at farm (Ecoinvent 2013)	C
Corn grain, at harvest in (year), at farm 85–91 % moisture (state) (LCA Digital Commons 2013)	D
Regional variations in GHG emissions of bio-based products in the United States—corn based ethanol and soybean oil (Kim and Dale 2009b)	E
Corn, at farm (Ecoinvent 2013)	F
Grain maize IP, at farm (Ecoinvent 2013)	G
Silage maize IP, at farm (Ecoinvent 2013)	H
Measuring ecological impact of water consumption by bioethanol using life cycle impact assessment (Chiu et al. 2012)	I
Life cycle assessment of various cropping systems utilized for producing biofuels: bioethanol and biodiesel (Kim and Dale 2005)	J
Improvements in life cycle energy efficiency and greenhouse gas emissions of corn-ethanol (Liska et al. 2009)	K
LCA of cropping systems with different external input levels for energetic purposes (Goglio et al. 2012)	L
Corn, whole plant, at field (NREL 2012)	M
Corn, production average, US, 2012 (NREL 2012)	N

classified as high quality, with the majority (eight) in the basic-quality range.

Table 6 Sugarcane dataset names and identifiers

Sugarcane dataset name	Identifier
Life cycle assessment of fuel ethanol from sugarcane in Brazil (Roberto Ometto et al. 2009)	O
A comparative life cycle assessment of PE based on sugar cane and crude oil (Liptow and Tillman 2012)	P
An environmental life cycle assessment comparing Australian sugarcane with US corn and UK sugar beet as producers of sugar for fermentation (Renouf et al. 2008)	Q
Carbon footprint of sugar produced from sugarcane in eastern Thailand (Yuttitham et al. 2011)	R
Bioproduction from Australian sugarcane: an environmental investigation of product diversification in an agro-industry (Renouf et al. 2013)	S
Sugarcane, at farm/BR (Ecoinvent 2013)	T
Life cycle assessment of Australian sugar cane production with a focus on sugarcane growing (Renouf et al. 2010)	V
A decision support tool for modifications in crop cultivation method based on LCA: a case study on GHG emissions reduction in Taiwanese sugarcane cultivation (Fukushima and Chen 2009)	V
Life cycle assessment of sugarcane ethanol and palm oil biodiesel joint production (Souza et al. 2012)	W

Table 7 Soy dataset names and identifiers

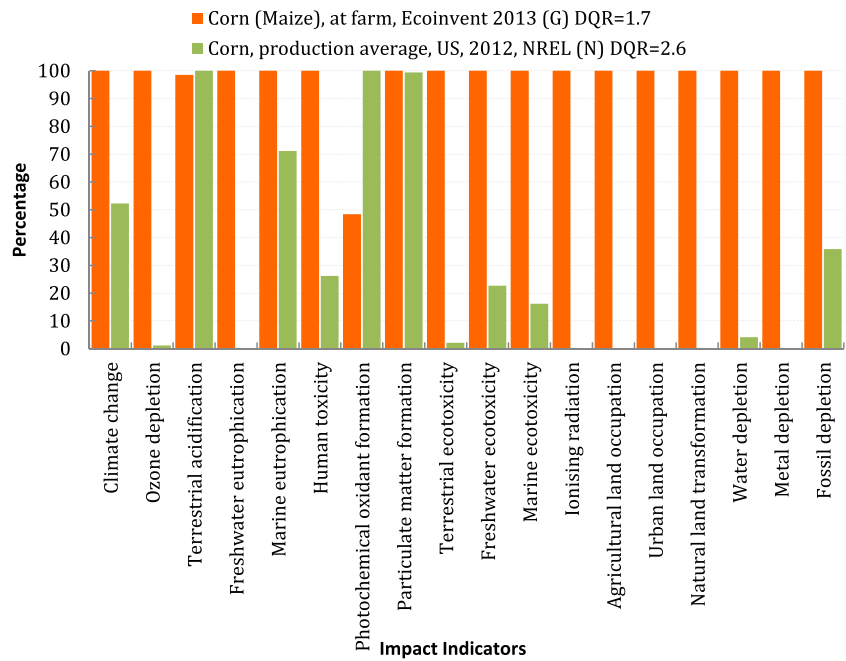
Soy dataset name	Identifier
Soy beans organic, at farm/CH (Ecoinvent 2013)	X
Soy beans IP, at farm/CH (Ecoinvent 2013)	Y
Soybeans; at harvest in (year); at farm; 85–92 % moisture (state) (LCA Digital Commons 2013)	Z
Life cycle assessment of soybean based biodiesel production in Argentina for export (Panichelli et al 2008)	AA
Soybeans, at farm/US (Ecoinvent 2013)	BB
Substitutable biodiesel feedstocks for the UK: a review of sustainability issues with reference to the UK RTFO (Upham et al. 2009)	CC
Soybean grains, at field (1998–2000)/kg/US (NREL 2012)	DD
Soybeans, at farm/BR (Ecoinvent 2013)	EE
Soybean grains, at field/kg/US (NREL 2012)	FF
Biogenic greenhouse gas emissions linked to the life cycles of biodiesel derived from European rapeseed and Brazilian soybeans (Reijnders and Huijbregts 2008)	GG
Soy bean, from farm (LCA Food 2007)	HH

The dataset “soy bean, from farm,” available from the LCA Food DK database, was the lowest scoring of all the datasets assessed with a DQR of 4.5. This rating is outside the range defined by the ILCD handbook for the lowest quality level of estimate (3.0 to 4.0). The low DQR is a product of receiving the undesirable score of 5 in the categories of both completeness and temporal representativeness. The temporal score was given because the age of the data is unknown, and the completeness score reflects the fact that very few flows are present in the dataset and major omissions were found in both the inputs and the emissions. Another selective comparison was conducted for two soybean datasets (i.e., soybean, from farm, LCA Food 2007 (HH) DQR=4.5 and soybeans, at farm/US (Ecoinvent 2013) (BB) DQR=2.3) using the ReCiPe Midpoint (H) V1.09/World ReCiPe H impact assessment method. Figure 7 shows that most of the ReCiPe midpoint indicator values are lower for the LCA Food DK dataset, or missing entirely. Climate change and ozone depletion are exceptions. Additional comparisons are presented in the online available Electronic Supplementary Material (Tables S1 to S3).

3.2.1 Technological representativeness

The datasets for all three feedstocks consistently scored the best in the technological representativeness category. This means that data from a different process than the one under study rarely had to be used as a substitute to fill a data gap. A technology aspect of corn and soybean cultivation that warrants attention is tilling practices. The method of tilling can have a large effect on the environmental impacts of a system as it affects the use of fossil fuels and soil degradation, among other issues (EPA 2013). Conventional till, reduced/

Fig. 6 Comparison of two corn datasets with DQR 2.6 and 1.7 using ReCiPe Midpoint (H) V1.09/World Recipe H impact assessment method

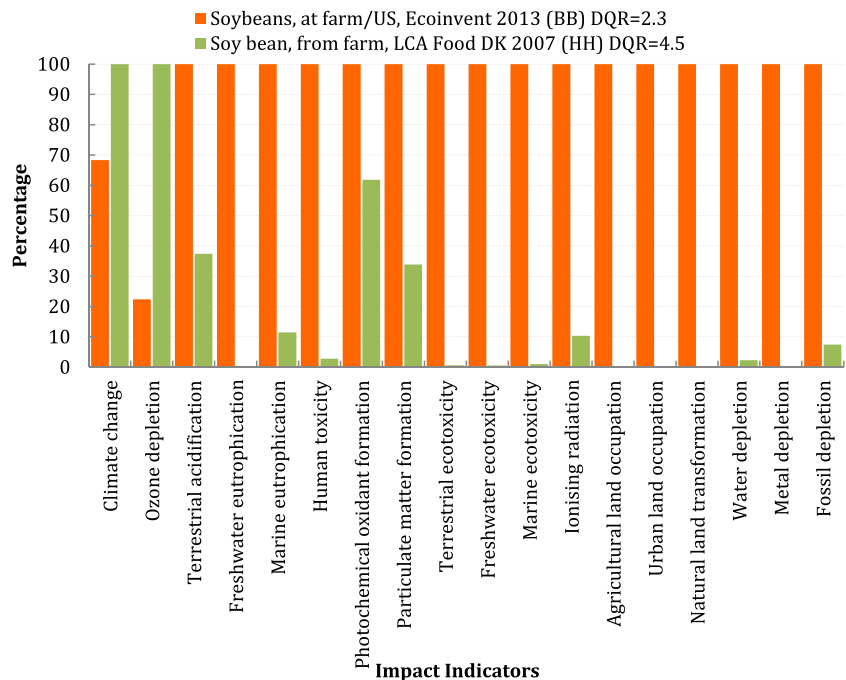


conservation till, and no-till are the basic methods used in corn and soy cultivation. In reduced and conservation till, the amount of tilling is decreased in relation to the conventional method, and a different type of plough is usually employed, while no-till uses a different planting technology to eliminate the need to till altogether. A 2010 USDA report estimates that 28.8 % of corn in the US was grown using conventional till practices in 2005 compared to 47.5 % grown with reduced or

conservation tillage, and 23.5 % which used no-till technology (Horowitz et al. 2010).

Ecoinvent uses a weighted average of conventional and conservation till practices for both corn and soy in their files. Of the 13 datasets evaluated for corn, seven (including Ecoinvent) reflected average tilling practices at the national level. Two represent state level averages, and one was varied by county. One dataset is explicitly for no-till technology.

Fig. 7 Comparison of two soybean datasets with DQR 2.3 and 4.5 using ReCiPe Midpoint (H) V1.09/World Recipe H impact assessment method



There is no corn dataset specifically representing conservation or reduced tillage practices, despite the fact that this is the most common tillage practice in the USA.

A higher percentage of soy was produced using no-till in 2005 than corn (45.3 %), while 43.2 % used reduced or conservation till technology and the remaining 11.6 % was conventional till (Horowitz et al. 2010). Unlike corn, a dataset modeling conservation till is available for soy. However, the majority of the soy files (seven of eleven) use a weighted average of conventional and conservation till similar to that used by Ecoinvent.

This weighted average data is not ideal for an LCA study that seeks to model a specific cultivation system because it does not accurately model any one tillage method. Additionally, the state averages vary significantly from the national averages. Texas, for example, produced 68.4 % of corn in 2005 using conventional till while Nebraska used conventional till for just 5.7 % of its corn production. Therefore, depending on what state is under study, the national average could either over- or underestimate the amount of inputs and emissions associated with tilling by a wide margin (Horowitz et al. 2010).

Sugarcane has other technological issues worth noting, particularly relating to the method of harvesting. Sometimes sugarcane is burned before harvest, and sometimes it is harvested green. This difference has an effect on emissions to air, and also on the amount of process water used during milling (Renouf et al. 2010). The burned sugarcane becomes sticky from the release of juices during burning and therefore generally has a large amount of debris mixed in with the harvested cane. As a result, a more vigorous washing process is required during processing. The sugarcane datasets vary in the percentage of cane that is burned before harvest. The Australian datasets reflect the national average of around 40 % burned and 60 % green at time of harvest, while a Brazilian dataset has the opposite ratio of 60 % burned and 40 % green harvest (Renouf et al.; 2010; Souza et al. 2012).

Additionally, there is a significant difference between manual and mechanical harvest. The Ecoinvent file that models sugarcane production in Brazil assumes that 80 % of the harvesting is done manually. In contrast, Australia uses dominantly mechanical harvesting, which is reflected in the Australian datasets. Harvesting sugarcane is the part of cultivation that contributes most significantly to global warming (Roberto Ometto et al. 2009). Therefore, data that accurately represents the method used to harvest sugarcane is important for accuracy of the final results of an LCA. São Paulo, the Brazilian state where much of the country's sugarcane is grown, adopted a law that will phase out the burning of sugarcane by 2017 (Manzatto 2009). This means that the harvesting datasets analyzed here will be significantly less representative as Brazil shifts to mechanical harvest and phases out burning of cane fields.

3.2.2 Geographical representativeness

Evidence suggests that variation between geographical areas can be complex and difficult to model. One study that tracked water consumption in ethanol production from corn in 81 different watersheds in Minnesota found significant variation in the range of water consumption between watersheds. The study included the use of both irrigation water and process water. The range in the amount of water used to produce 1-L of ethanol was found to be 3–181 L in a watershed in central Minnesota. Contrastingly, ethanol produced in a watershed in the south of the state had a much smaller range in the amount of water used. Farms and production facilities based in the southern watershed used only 3–8 L of water to produce 1-L ethanol from corn (Chiu et al. 2012). This demonstrates that the issue of geographical correlation is not as simple as one might assume. LCA practitioners are often compelled to use data from a different region than the one under study. Generally, an effort is made to use data from a region with similar geography and practices, but in this case large variation was found within a single region that was utilizing relatively uniform technology. This suggests not only that it is quite difficult to predict a correlation between regions, but also that even data from a slightly different part of the same region may not be an accurate substitute.

In the category of geographic representativeness, the corn datasets scored fairly well. All but one dataset either represents the area under study or is from a larger area that includes the study area. The exception is a dataset based on a single-site field study conducted in central Italy. The publication associated with this data was intended to model corn production in the “Mediterranean region,” so the dataset was classified as from a smaller area within the Mediterranean and therefore was scored at a value of 3 (Goglio et al. 2012).

Four of the corn datasets are based on US average values. This is problematic because the US is quite large, and therefore subject to regional variations not only in weather and conditions, but also in technology. For example, the USDA reports that about 15 % of corn acres in 2007 in the US were irrigated (USDA 2012). However, this 15 % is not evenly scattered over the entire growing region, but concentrated in certain areas. Nebraska, for example, has an irrigation rate of 60.6 %, much higher than the national average irrigated (Univ. Nebraska-Lincoln 2012). A dataset based on national averages would therefore likely underestimate the amount of water used if the system under study was in Nebraska.

More geographically specific data has the potential to solve this problem, and some of it is available. Three datasets model corn cultivation at the county level. Of the three, one is only for water in Minnesota and one is specifically for Scott County, Iowa. These two datasets will therefore be of limited use to practitioners because of their geographic and technological restraints. The third dataset, titled “Annual Report: Life Cycle

Assessment to Improve the Sustainability and Competitive Position of Biobased Chemicals: A Local Approach,” by Kim and Dale (2009a) is far more versatile in scope. It includes detailed data for several counties across corn growing states, and also has the highest DQR (1.4) of all the datasets evaluated. This level of geographic detail is unique to corn cultivation. State level data is the most geographically specific level of information available for all other crops in this report. Only the LCA Commons data is available by state for soy cultivation. The rest of the soy datasets reflect national averages.

This type of regional variation is not unique to the USA. Four of the sugarcane datasets represent Brazilian production, with two specifically for the state of São Paulo. Three datasets are for Queensland, Australia. This region accounts for 98 % of sugarcane production in Australia, so the lack of data from other areas is not a particularly important gap (Renouf et al. 2010). However, there is significant variation in the growing conditions and intensity of inputs within this region. The study by Renouf et al. (2010) includes datasets for the two areas of Queensland with the most disparate growing conditions.

As illustrated by the discussion above, geographical and technological differences are often strongly related, with quality issues bridging both categories. These types of quality issues call into question the usefulness of datasets based on national averages when there is significant variation within regions.

3.2.3 Temporal representativeness

Of the three crops evaluated, sugarcane scored the worst in the category of temporal representativeness with five of the nine sets receiving a rating of 3 or above. Two soy datasets have a score of 5 in this category; one is based on primary data of an unknown age, and the other uses a significant amount of data from 1979. Even the Ecoinvent dataset for sugarcane relies on older data than the Ecoinvent files for corn and soy, using data for agronomic inputs from 1988. This is uncharacteristic of the Ecoinvent database, which collected most of their data between the late 1990s and the mid-2000s.

Crop yield data is particularly sensitive to age because of advances in yields over recent years. A yield increase basically has the effect of diluting the environmental impacts of a cropping system by spreading them over more outputs. For this reason, the results of LCA studies involving crops tend to be quite sensitive to yield changes (Fukushima and Chen 2009). Therefore, when new data is collected, a high priority should be given to the collection of updated crop yields.

3.2.4 Completeness

Land use change (LUC), and specifically indirect land use change, is the largest problem in the category of completeness. Four of the corn datasets do not include any type of LUC data, and only two of the datasets explicitly include indirect LUC. It

should be noted that the database files (Ecoinvent, USLCA, and LCA Commons) do not distinguish between the types of LUC at the flow level. Therefore, all that could be determined about these files is that they include some land use change data, but it was not possible to distinguish between direct LUC and indirect LUC.

This data gap is even more pronounced for sugarcane. Five of the ten datasets evaluated do not account for LUC, and three of those five also do not include land occupation. The Taiwanese sugarcane dataset does not include LUC despite the fact that the article itself states that sugarcane production is expected to expand in the region and that fallow land will likely be converted for cultivation (Fukushima and Chen 2009). Additionally, three of the studies that include LUC also state that their information on this input is not complete. One of the Australian studies that include LUC presents the data with the qualifier that the methods used to evaluate both LUC and water impacts have significant limitations (Renouf et al. 2013). Another study echoes this sentiment when it states that the LUC emissions are uncertain due to lack of uniform methods (Liptow and Tillman 2012). The Ecoinvent file also follows this pattern as there is a high degree of uncertainty associated with the LUC flows. The soy datasets also deal inconsistently with this impact, although it is less pronounced than in the sugarcane data. Four of the eleven soy files do not include indirect LUC, and two of these also exclude direct LUC.

The significance of this omission varies depending on the system under study. If the study is for an established growing system that is not expanding, then it would not be highly important information to include. However, for most of these crops, production is expanding in response to increased demand for bio-products. This is true in South America, where sugarcane production has expanded in Brazil at an average rate of approximately 85,000 ha/year since 1990 and soy production in Argentina has gone from less than a million hectares to 13 million since 1970 (Martinelli and Filoso 2008; Grau et al. 2005). Similarly, corn in the US has also been expanding in recent years. From 2007 to 2013, 11.17 million acres of land that was formerly in the conservation reserve program has been converted to crop cultivation, most of it for corn production (Farm Service 2007, 2013).

3.2.5 Uncertainty

As stated in the “Methods” section, an uncertainty evaluation was only done for the Ecoinvent data. In general, these datasets had low uncertainty since data is mostly based on verified measurements. The exception is that all transport distances in the Ecoinvent datasets are estimates. The Ecoinvent sugarcane datasets also have a greater degree of uncertainty than the soy or corn Ecoinvent datasets because the sugarcane energy and carbon dioxide data are partly based on qualified estimates.

Additionally, an uncertainty score was given to one USLCI file: “Corn, production average, US, 2022.” It was possible to score this dataset because it is a qualified estimate for corn production in the future and was therefore given the standard uncertainty score of 4 for qualified estimates.

As discussed above, many studies that included data for LUC expressed concerns of uncertainty along with that data. It was not possible to give an overall uncertainty score to these datasets, however, because these concerns were generally expressed only in qualitative terms and uncertainty information was not available for the other inputs in the datasets.

4 Conclusions

Significant data gaps exist in the availability of life cycle inventory data for bio-based polymers. These gaps occur geographically, technologically, and temporally. In addition, gaps exist for certain inputs, like land use change, independent of those qualifying factors. National averages are unlikely to adequately represent either technology used to cultivate a crop in any specific region or the growing conditions in that region. There is therefore a need for more regionally explicit data that accurately models the technology and conditions of a specific system under study. Land use change is often not accounted for in otherwise relatively complete datasets, which is a significant quality issue because it can have a large influence on the overall impacts of a system. More data for land use change is needed, and standardized methods for collecting and incorporating such data into LCA studies are also necessary. Newer data for the cultivation of feedstocks, especially crop yields, would also be beneficial. Finally, the currently available data is skewed heavily to Europe and the USA, leaving a significant portion of the globe with very few datasets available. In conclusion, understanding the impacts caused by the production of bioplastics is the first step on the path to more sustainable bioplastics, and in order to accurately evaluate these impacts the data gaps described above must be resolved.

Acknowledgements The authors wish to thank the Center for Packaging Innovation and Sustainability for the financial support for this project, Prof. Bruce Dale for providing information, undergraduate packaging student Alison Schuitema for her contributions to the literature survey, and graduate student Woranit Muangmala for her contributions to the Electronic Supplementary Material. RA acknowledges the financial support by the USDA National Institute of Food and Agriculture and Michigan AgBioResearch, Hatch project R. Auras.

References

Álvarez-Chávez CR, Edwards S, Moure-Eraso R, Geiser K (2012) Sustainability of bio-based plastics: general comparative analysis and recommendations for improvement. *J Clean Prod* 23:47–56

- Arvidsson R, Persson S, Fröling M, Svanström M (2011) Life cycle assessment of hydrotreated vegetable oil from rape, oil palm and Jatropha. *J Clean Prod* 19:129–137
- Canola Council of Canada (2011) Industry overview. <http://www.canolacouncil.org/markets-stats/industry-overview/>. Accessed 14 April 2013
- Chiu YW, Suh S, Pfister S, Hellweg S, Koehler A (2012) Measuring ecological impact of water consumption by bioethanol using life cycle impact assessment. *Int J of Life Cycle Assess* 17:16–24
- EC JRC (2010) ILCD handbook: general guide for life cycle assessment—detailed guidance, European Commission Joint Research Center. <http://eplca.jrc.ec.europa.eu/uploads/2014/01/ILCD-Handbook-General-guide-for-LCA-DETAILED-GUIDANCE-12March2010-ISBN-fin-v1.0-EN.pdf>. Accessed 17 April 2013
- Ecoinvent (2007) Overview and methodology, Swiss Center for Life Cycle Inventories. http://www.ecoinvent.org/fileadmin/documents/en/01_OverviewAndMethodology.pdf. Accessed 17 April 2013
- Ecoinvent (2013) software program, Swiss Center for Life Cycle Inventories
- EPA (2013) Soil preparation. <http://www.epa.gov/agriculture/ag101/cropsoil.html>. Accessed 17 April 2013
- Evans J (2010) Bioplastics get growing. *Plastics Engineering*, Feb., pp. 15–19
- Farm Service Agency (2007) USDA Conservation Reserve Program Summary and Enrollment Statistics FY 2007. USDA http://www.fsa.usda.gov/Internet/FSA_File/annual_consv_2007.pdf. Accessed 30 May 2014
- Farm Service Agency (2013) USDA Conservation Reserve Program Monthly Summary December 2013. USDA http://www.fsa.usda.gov/Internet/FSA_File/dec2013asummary.pdf. Accessed 30 May 2014
- Fukushima Y, Chen SP (2009) A decision support tool for modifications in crop cultivation method based on life cycle assessment: a case study on greenhouse gas emission reduction in Taiwanese sugarcane cultivation. *Int J Life Cycle Assess* 14:639–655
- Goglio P, Bonari E, Mazzoncini M (2012) LCA of cropping systems with different external input levels for energetic purposes. *Biomass Bioenerg* 42:33–42
- Grau HR, Aide TM, Gasparri NI (2005) Globalization and soybean expansion into semiarid ecosystems of Argentina. *Ambio* 34:265–266
- Horowitz J, Ebel R, Ueda K (2010) “No-Till” farming is a growing practice. United States Department of Agriculture. <http://www.ers.usda.gov/ersDownloadHandler.aspx?file=/media/135329/eib70.pdf>. Accessed 17 April 2013
- ISO 14040 (2006) Environmental management in life cycle assessment—principles and framework. International Organization for Standardization, Geneva
- ISO 14044 (2006) Life cycle assessment—requirements and guidelines. International Organization for Standardization, Geneva
- Kim S, Dale BE (2005) Life cycle assessment of various cropping systems utilized for producing biofuels: bioethanol and biodiesel. *Biomass Bioenerg* 29:426–439
- Kim S, Dale BE (2009a) Life cycle assessment to improve the sustainability and competitive position of biobased chemicals: a local approach. Annual Report, Grant Agreement No. 68-3A75-6-509. Michigan State University, East Lansing, MI
- Kim S, Dale BE (2009b) Regional variations in greenhouse gas emissions of biobased products in the United States—corn-based ethanol and soybean oil. *Int J Life Cycle Assess* 14:540–546
- LCA Digital Commons (2013) National Agricultural Library, <https://www.lcacommons.gov/>. Accessed various dates in 2013
- LCA Food DK (2007) LCA Food Database. Downloaded from <http://www.lcafood.dk/>
- Liptow C, Tillman AM (2012) A comparative life cycle assessment study of polyethylene based on sugarcane and crude oil. *J Ind Ecol* 16(3): 420–435

- Liska AJ et al (2009) Improvements in life cycle energy efficiency and greenhouse gas emissions of corn-ethanol. *J Ind Ecol* 13:58–74
- Manzatto CV (2009) Sugarcane agroecological zoning for ethanol and sugar production in Brazil. *Embrapa Environment*
- Martinelli LA, Filoso S (2008) Expansion of sugarcane ethanol production in Brazil: environmental and social challenges. *Ecol Appl* 18: 885–898
- Nemecek T et al (2012) Estimation of the variability in global warming potential of worldwide crop production using a modular extrapolation approach. *J Clean Prod* 31:106–117
- NREL (2012) U.S. Life Cycle Inventory Database. National Renewable Energy Laboratory. <http://www.nrel.gov/lci/> Accessed various dates in 2013
- Panichelli L, Dauriat A, Gnansounou E (2008) Life cycle assessment of soybean-based biodiesel in Argentina for export. *Int J Life Cycle Assess* 14:144–159
- Patel M, Narayan R, Selke S, Auras R, Grabowski A, Rahimiashtiyani S (2014) LCA studies for bio-based plastics: closing critical data gaps and agreeing on representative generic datasets. Center for Packaging Innovation and Sustainability, Michigan State University, East Lansing, MI, available at http://www.cpis.msu.edu/research/the_impacts_of_the_centers_research
- Reijnders L, Huijbregts MAJ (2008) Biogenic greenhouse gas emissions linked to the life cycles of biodiesel derived from European rapeseed and Brazilian soybeans. *J Clean Prod* 16:1943–1948
- Renouf MA, Wegener MK, Nielsen LK (2008) An environmental life cycle assessment comparing Australian sugarcane with US corn and UK sugar beet as producers of sugars for fermentation. *Biomass Bioenergy* 32:1144–1155
- Renouf MA, Wegener MK, Pagan RJ (2010) Life cycle assessment of Australian sugarcane production with a focus on sugarcane growing. *Int J Life Cycle Assess* 15:927–937
- Renouf MA, Pagan RJ, Wegener MK (2013) Bio-production from Australian sugarcane: an environmental investigation of product diversification in an agro-industry. *J Clean Prod* 39:87–96
- Roberto Ometto A, Zwicky Hauschild M, Nelson Lopes Roma W (2009) Life cycle assessment of fuel ethanol from sugarcane in Brazil. *Int J Life Cycle Assess* 14:236–247
- Smith PB (2013) Trends in biobased chemistry. *Plastics Engineering*. June, pp 44–48
- Souza SP, de Ávila MT, Pacca S (2012) Life cycle assessment of sugarcane ethanol and palm oil biodiesel joint production. *Biomass Bioenergy* 44:70–79
- Univ. Nebraska-Lincoln (2012) Irrigation and water management for corn. Irrigation and water management resources for Nebraska producers 2012. <http://cropwatch.unl.edu/web/corn/water>. Accessed 16 April 2013
- Upham P, Thornley P, Tornei J, Boucher P (2009) Substitutable biodiesel feedstocks for the UK: a review of sustainability issues with reference to the UK RTFO. *J Clean Prod* 17:S37–S45
- USDA (2012) 2007 Census Ag Atlas Maps, US Dept. of Agriculture, http://www.agcensus.usda.gov/Publications/2007/Online_Highlights/Ag_Atlas_Maps/Crops_and_Plants/. Accessed 2 Jun 2014
- USDA (2014) Grain: world markets and trade. <http://usda.mannlib.cornell.edu/usda/current/grain-market/grain-market-05-09-2014.pdf>. Accessed 30 May 2014
- van den Berg N, Huppel G, Lindeijer EW, van der Vaen BL, Wrisberg MN (1999) Quality assessment for LCA. CML Report 152. Centre of Environmental Science, Leiden, Netherlands
- van der Voet E, Lifset RJ, Luo L (2010) Life-cycle assessment of biofuels, convergence and divergence. *Biofuels* 1:435–449
- Yuttitham M, Gheewala SH, Chidthaisong A (2011) Carbon footprint of sugar produced from sugarcane in eastern Thailand. *J Clean Prod* 19:2119–2127