DATA AVAILABILITY, DATA QUALITY

Using national inventories for estimating environmental impacts of products from industrial sectors: a case study of ethanol and gasoline

Debalina Sengupta · Troy R. Hawkins ·
Raymond L. Smith

Abstract

Purpose In order to understand the environmental outcomes associated with the life cycle of a product, to compare these outcomes across products, or to design more sustainable supply chains, it is often desirable to estimate results for a reference supply chain representative of the conditions for a sector in a specific region. This paper, by examining ethanol and gasoline production processes, explains how choices made in the calculation of sector-representative emission factors can have a significant effect on the emission estimates used in life cycle assessments.

Methods This study estimates reference emission factors for United States dry-grind corn ethanol production and gasoline production processes suitable for use in baseline life cycle assessment unit processes. Based on facility-specific emissions and activity rates from the United States National Emissions

D. Sengupta
Oak Ridge Institute for Science and Education (ORISE),
Oak-Ridge, TN 37830, USA

D. Sengupta · T. R. Hawkins · R. L. Smith (✉)
Sustainable Technology Division, National Risk Management
Research Laboratory, U.S. Environmental Protection Agency,
Cincinnati, OH 45268, USA
e-mail: smith.raymond@epa.gov

Present Address:
D. Sengupta
Artie McFerrin Department of Chemical Engineering, Texas A&M
University, College Station, TX 77843, USA

Present Address:
T. R. Hawkins
Business Intelligence, Enviance, Carlsbad, California 92008, USA

Inventory, the Energy Information Administration, and an ethanol industry trade publication, the average emissions per unit energy content of fuel are computed using three different approaches. The Tool for the Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI) characterization factors are used to estimate impact potentials for six environmental and three human health categories. Sector-specific direct emissions and impact potentials are compared across the three approaches and between the two sectors. The system boundary for this analysis is limited to the fuel production stage of these transportation fuel lifecycles.

Results and discussion Findings from this work suggest that average emission factors based on total emissions and total production may significantly under estimate actual process emissions due to reporting thresholds and otherwise unreported emissions.

Conclusions Because of the potential for unreported emissions in regional inventories, it is more appropriate to estimate sector reference emission factors based on matched sets of facility or process level emissions and activity rates than to use aggregated totals. This study demonstrates a method which can be used for inventory development in cases where multiple facilities producing the same product are involved.

Keywords Biofuel · Biorefinery · Emission factors · Ethanol · Gasoline · National Emissions Inventory (NEI) · Toxics Release Inventory (TRI) · Uncertainty

1 Introduction

Life cycle assessment (LCA) is gaining importance as a tool for design of sustainable supply chains of product systems (Young et al. 2012; Ingwersen et al. 2013; You

et al. 2012). The life cycle stages of a product resemble the stages in a supply chain network, and combining these two methods will help in the development of sustainable supply chains. When LCA is used as a tool for sustainable supply chain design of products, arriving at the right level of detail to model a unit process is a challenging task (Hagelaar and van der Vorst 2001). When the objective is to represent a reference supply chain or when the details regarding a particular process are unknown, it is common practice to create a reference process to represent a sector average. In this case, national average input ratios and release factors are usually calculated using top-down data representing national total inputs, releases, and activity rates. This practice is particularly prevalent in the creation of input–output LCA models. The attractiveness of such an approach is its simplicity. Unfortunately, when certain releases are not reported by some facilities, the calculation of an average in this way may yield misleading results. In this study, air emission factors for the United States petroleum refining and dry-grind corn ethanol production sectors are compared using different approaches and assumptions. These differences arise from instances where emission data might be expected but no data are available. This study was conducted in order to understand the bias introduced using top down values in estimating sector average emissions.

This work was part of a larger project tasked with developing approaches for designing sustainable supply chains (Young et al. 2012) with a specific focus on the replacement of conventional fuels with biofuel alternatives. The project began with the development of models of the life cycles of gasoline and dry-grind corn ethanol. In particular, within the larger project, the work described here contributed to improving the understanding of the petroleum refining and dry-grind corn to ethanol sectors. The initial thinking in modeling these sectors was they could be represented as reference processes within supply chains representing United States average conditions.

An important aspect in the inventory development stage of the life cycle of a product is the sourcing of data for each stage in the supply chain (Curran 2012). In the early stage of the project, there were numerous approaches proposed by a diverse team of chemical engineers and LCA practitioners regarding how best to represent reference supply chains and the associated data quality requirements. One option was to use data collected from a specific facility selected to serve as a reference. This option presented challenges in obtaining data for a specific facility and the obstacles associated with the desire of those responsible for the facility to manage publicly communicated information regarding their practices. A second option was to attempt to infer unit process information from optimized process-design models built in process simulation software. This option proved untenable as the process-design software was helpful in obtaining conversion of the primary products, but did not include sufficient information to infer credible emission factors for most emitted pollutants.

A third option was to use national totals to calculate an average. This option was feasible and attractive due to its transparency and ability to represent a generic case. However, the drawback in this method was that the national total *sector average* facility is conceptual and cannot be directly validated to represent the sector totals, nor does it provide insight into variability within the sector. Thus, a fourth option was pursued where a correspondence between national datasets containing emissions and activity rates was developed, and this was used to estimate emission factors for a large number of facilities by aggregating to a sector average.

Typically in an LCA, the inventory of inputs and outputs from a process is associated with a reference flow. The focus of discussion in this paper is on the differences that arise in inventory development and impact assessment when data on emissions and the reference flow (fuel capacity or production rates) from various data sources are used. The data can be interpreted in numerous ways, three of which based on the activity rate are presented in this paper. The method is demonstrated with ethanol and gasoline production stages, but it is widely applicable to any sustainable supply chain design of products from the manufacturing sector for use by LCA practitioners.

The results from this paper provide users with a robust method to calculate emissions for the purpose of LCA and help users to compare results when emission scenario assumptions from manufacturing facilities are considered. This is crucial for LCA as the emission factors developed using the method affect the results of analysis. For an analyst, this presents a way to differentiate between these assumptions and accurately connect the source and flow of information throughout the life cycle.

2 Methods

Generally, detailed process models are effective in modeling primary emissions from the conversion of raw material to a product and emissions from associated energy use, thus providing a nominal set of process emissions. However, in an operating process plant, various other incidents can give rise to emissions, which are harder to predict through models. Examples of these actual emissions include side reactions that occur during a process which depend on the operating conditions, plant process upset conditions such as leaks and spills, etc. To confirm the importance of the various release types, a comparison of ratios of natural gas boiler emissions (U.S. EPA 2013a) and ratios of actual inventory emissions showed very large differences for greenhouse gases and criteria pollutants. This confirms that the method needs all of the release types; that natural gas boiler emission factors are insufficient to describe the complex actual emissions. The LCA stage for product manufacturing often provides a high-level view and

assumes a fixed supply chain structure and operation (Nwe et al. 2010), where information on actual process emissions are not accounted for. Moreover, a complete speciation of emissions is not available from LCA databases, which represent a generic facility producing the product. The method given in this paper addresses these limitations by developing a detailed inventory from a data driven approach.

Three national average emission factors were computed based on data from publicly available emission and production data. These emission factors accounted for the emission scenario assumptions related to the activity rate in the denominator of the emission factor. To characterize these environmental stressors a commonly used LCA impact assessment method, Tool for the Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI; Bare 2002), was used to compute the impacts from the two fuel production stages in six environmental and three human health categories.

2.1 Data sources

2.1.1 Emission data

National Emissions Inventory Under the Clean Air Act (CAA), EPA establishes air quality standards to protect public health, including the health of people with asthma, children, and older adults (U.S. EPA 2013b). EPA sets limits to protect public welfare including protection of ecosystems, damage to crops, vegetation, and buildings. To achieve the standards of the CAA, EPA has regulatory and voluntary programs in place to reduce the amount of air pollutants emitted from a wide range of sources. To keep track of these emissions, the EPA maintains the National Emissions Inventory (NEI), a national database of air pollutant emission information. Point source emission is a classification among the different types of emissions inventoried in NEI (U.S. EPA 2013c). All industries having point source emissions and meeting certain reporting criteria are required to report their annual air emission data to State, Local, or Tribal (S/L/T) air agencies. The emission estimates are supplemented by data developed by the EPA and compiled in the NEI database (U.S. EPA 2013c). Thus, data provided by individual industries are verified by the respective agency prior to entry into the national inventory. The NEI is compiled every 3 years, and air emissions of facilities are reported with identifiers of the facility and chemical compound.

Toxics Release Inventory The Emergency Planning and Community Right-to-Know Act (EPCRA) establishes requirements for Federal, state and local governments, Indian Tribes, and industry regarding emergency planning and "Community Right-to-Know" reporting on hazardous and toxic chemicals (U.S. EPA 2013d). EPCRA Section 313 requires EPA and the States to collect data annually on releases and transfers of over

650 toxic chemicals from industrial facilities and make the data available to the public through the Toxics Release Inventory (TRI; U.S. EPA 2013e). The TRI contains information about how facilities manage toxic chemicals through recycling, energy recovery, and treatment, collected under the Pollution Prevention Act. Facilities are required to report to TRI, but the data does not go through the same level of review as NEI.

Though the two databases satisfy different needs, NEI and TRI are good places to start an investigation of the release of a particular air pollutant in a certain geographical location for a particular sector. The two databases have many chemical compounds in common. In this paper, the NEI is used as the primary data source of the chemicals common to both databases, primarily because of the rigorous review it undergoes before it is reported in the national level database. TRI is used to supplement the data for those emissions, which are not reported to NEI. The 2008 NEI dataset was used in this analysis. Since TRI is available on an annual basis, a 3-year average over 2007–2009 was used so the median year was 2008 and comparable to the temporal choice for NEI. The 3-year average was chosen to include the variability in the emissions over the years. Kim et al. (2013) provide a useful comparison of the NEI and TRI, where TRI was identified to have potential data gaps due to reporting requirements which do not require reporting by a number of facilities in certain sectors (Miranda et al. 2008).

2.1.2 Activity data

The denominators for the emission factors presented here are based on the production of saleable products. Table 1 presents the products from a biorefinery (ethanol and dry distiller grains with solubles (DDGS)) and a petroleum refinery (eighteen refinery products). National average emission factors were calculated based on national ethanol and gasoline production statistics obtained from the Energy Information Administration (EIA) for 2007–2009 (U.S. EIA 2011). The list of pollutants included in the inventory for these sectors is available in the Supplementary Information. Because facility-specific actual production information is not readily available, the production of facilities is based on their capacity, assuming equivalence, although one could apply a capacity utilization factor to the total calculated capacity when available. For instance, EIA reports over 2007–2009 a capacity utilization for petroleum refineries of 85.6 %, which would have increased sector emission factors by 17 %. Ethanol capacities were obtained from Ethanol Producer Magazine (EPM) (EPM 2011). Production capacities of petroleum refineries for this work were estimated based on their crude oil processing capacity. Yield data from EIA (U.S. EIA 2011) were used to translate refinery crude oil processing capacity into gasoline production capacity.

Table 1 Economic allocation of products from corn biorefineries and petroleum refineries (italicized values in the table are used in determining the allocation of emissions to ethanol and gasoline)

Products	Economic allocation
Corn biorefinery products of economic value (Based on a 40-MMgpy ethanol plant)	
Ethanol, denatured, E98	*83 %*
Distillers dried grains with solubles (DDGS)	17 %
Petroleum refinery products of economic value (sector totals)	
Gasoline	*27.79 %*
Gasoline blending components	*21.33 %*
Diesel	26.35 %
Kerosene-type jet fuel	9.70 %
Lubricants	3.54 %
Residual fuel oil	2.73 %
Liquefied petroleum gases	2.15 %
Asphalt and road oil	1.54 %
Other oils for petrochemical feedstock use	1.29 %
Naphtha for petrochemical feedstock use	1.25 %
Petroleum coke	0.87 %
Miscellaneous petroleum products	0.68 %
Special naphthas	0.33 %
Kerosene	0.21 %
Aviation gasoline	0.13 %
Waxes	0.12 %
Aviation gasoline blending components	0.01 %

2.2 Allocation method

Economic allocation was chosen to allocate the emissions to the products from the two fuel production stages. Ethanol production by dry-grind technology produces dried distillers grain with solubles (DDGS) as a co-product. A USDA generic model for the production of ethanol was used to estimate the production rates and economic values associated with the ethanol and DDGS production (Kwiatkowski et al. 2006). Similarly, for petroleum refineries, varieties of products are produced in addition to finished motor gasoline (see Table 1). EIA provides data on the economic values and quantities of these products from which the economic allocation was computed (U.S. EIA 2011). Table 1 shows the economic allocation for various products of the corn biorefinery and petroleum refinery. Of all the emissions produced, 83 % from the biorefinery are allocated to ethanol and 49 % from petroleum refineries to gasoline. It should be understood that individual processes in a sector can have different product slates and thus different allocated emissions. For example, see the Supplementary Information, where regional gasoline production is shown to vary between 15 and 47 % as compared to the sector average of 28 % used in Table 1.

2.3 Harmonizing data

In order to connect the data on emissions, production capacities, and production rates and allocate emissions from the two fuels for all of the facilities in the USA, a harmonization effort was undertaken in this research. This effort has set up the groundwork and identified the needs for the development of the Life Cycle Assessment Harmonization Tool (Hawkins et al. 2013). The primary goal of the effort was to standardize the unique key that connects the five (NEI, TRI, EPM, EIA, and TRACI) data sources using different nomenclature to denote the pollutants. There was a need for harmonization of the facility names so that facility specific emission factors could be computed. It is important to note here that several facilities changed names over the years (often as the ownership changed), and proper accounting of this needed to be reflected in the computation of emission factors from TRI data. The harmonization involved the following steps: (1) connecting the facility emission dataset to the facility activity (capacity in this case study) dataset, (2) allocating the emissions to the fuel, and (3) connecting the emission data to the impact categories for the pollutants. A relational database was created in Microsoft Access® to organize the data, find the number of matches in the databases, and compute the emission factors. The results of this effort are summarized in Supplementary Information which includes the emission factors, potential environmental impacts, and the statistical summary of the emission of chemicals and facilities used in this paper.

2.4 Emission factors and reference flow

An emission factor is a representative value that relates the quantity of a pollutant released to the atmosphere with an activity associated with the release of that pollutant. EPA gives the general method for estimating emissions from a source in the AP 42 standards (U.S. EPA 2013a). These factors are usually expressed as the mass of pollutant divided by a unit mass, volume, distance, or duration of the activity emitting the pollutant (e.g., kilograms of particulate emitted per megagram of coal burned). Various methods of emission estimation exist, such as continuous emission monitoring, process equipment emission factor ratings, plant site-specific emission factors, and others. Using these methods, manufacturing facilities are required to estimate and report emissions under the requirements of the CAA and EPCRA. These emission estimates are then compiled in databases and made publicly available by the EPA.

In this paper, these national databases serve as the sources of historical emission data from all manufacturing facilities in the United States producing ethanol and gasoline. The basis of emission estimation for individual facilities varies, and the computation method takes an arithmetic average to find the

national average emissions, accounting for the differences in estimation methods.

For the purpose of LCA, the emissions have been expressed as a function of the reference flow, one megajoule of energy, based on the lower heating value of ethanol or gasoline produced in the USA. Thus, all the fuel production capacities and actual production rates from the data sources were converted to the reference flow unit and used as the activity rate in computing emission factors, expressed as kilograms of pollutant per megajoule of energy.

After a closer inspection of facilities reporting to the NEI and TRI databases, it was found that the total number of facilities operating in the USA is different from the number of facilities reporting their emissions. There may be several reasons for this according to the reporting requirements of facilities (Miranda et al. 2008), or the environmental rules guiding these databases (U.S. EPA 2013f). As an example, there were 149 operating corn ethanol facilities reported in the Ethanol Producer Magazine in 2008, but only 76 facilities reporting in the NEI database for that year. Of these, not all of the 76 facilities reported all the pollutants. For example, only 11 facilities reported the pollutant 1,3-butadiene. Thus, calculations are based on assumptions of the smaller number of facilities that report any emission to the databases, and an even smaller number, which report a particular pollutant. Hence, N^* is the number of facilities operating in the USA, N is the total number of facilities reporting to NEI or TRI, and n is a subset of these N facilities which report pollutant i. From observing data reported to NEI and TRI, one can conclude that $n \leq N \leq N^*$ To account for these differences in reporting of pollutants, three different emission factors were considered, as denoted in Table 2. These emission factors are highly dependent on the quality of data, and the correctness of the reported emissions, capacities, and production volumes were assumed as a prerequisite for analysis. The first emission factor is based on the n

facilities reporting pollutant i in the NEI or TRI databases, the second is based on the N facilities reporting any pollutant to the NEI or TRI databases, and the third is based on the actual production of the fuel from the N^* facilities in the United States. The production from N^* facilities is represented as P, the total production of a given fuel in the country as reported in EIA. If a particular technology is used, as in this case dry grind for corn ethanol, the fraction of technology is expressed as f_T. For ethanol, 86 % of the total production uses dry-grind technology (Mueller 2010).

Emission factor based on production of facilities reporting emissions of a given pollutant (F_{C_n}) This emission factor represents the emission of a particular pollutant, i, from the n facilities that report that pollutant. As represented in Table 2, the emission factor is constructed from the emissions from n facilities having a total capacity of C_n. For low reporting of pollutants, i.e., $n \ll N$, this assumption gives an upper estimate of the value for the emission factor based on the values available in the emission inventories.

Emission factor based on production of facilities reporting emissions of any pollutant (F_{C_N}) This emission factor represents the emission of a particular pollutant, i, from the N facilities reporting any pollutant in the dataset of NEI and TRI. This assumes that only n facilities emit the chemical i, and the $N-n$ facilities do not report the chemical because the amount produced is either zero, or negligible. In this case the EF denominator should still include the capacities of the $N-n$ facilities. For low reporting of pollutants, i.e., $n \ll N$, this assumption gives a lower representative estimate of the value for the emission factor.

Emission factor based on an aggregate reported production of the fuel ($F_{P_{N^*}}$) This emission factor represents the emission of

Table 2　Emission and activity assumptions for calculation of emission factors

Emission factor calculation options $\left(F = \frac{M}{A}\right)$	Calculation of emission	Calculation of activity	Assumption
Emission factor based on production of facilities reporting emissions of a given pollutant $F_{C_n,i} = \frac{\text{Emission of Pollutant}}{\text{Production from facilities reporting a given pollutant}}$	$M_i = \sum\limits_{j}^{n} m_{i,j}$	$A = C_n = \sum\limits_{j}^{n} c_j$	$c_j = 0$ if $m_{i,j} = 0$ If $m_{i,j}$ is not available in dataset, then the capacity, c_j is not included in calculating C_n
Emission factor based on production of facilities reporting emissions of any pollutant $F_{C_N,i} = \frac{\text{Emission of Pollutant}}{\text{Production from facilities reporting any pollutant}}$		$A = C_N = \sum\limits_{j}^{N} c_j$	$c_j = c_j$ if $m_{i,j} = 0$ If $m_{i,j}$ is not available in dataset, then the capacity, c_j is included in calculating C_N
Emission factor based on an aggregate reported production of the fuel $F_{P_{N^*},i} = \frac{\text{Emission of Pollutant}}{\text{Production of fuel}}$		$A = f_T \times P_{N^*}$	If $m_{i,j}$ not available in the dataset, then, $m_{i,j} = 0$, P_{N^*} = total production from N^* operating facilities

F emission factor, i pollutant, j facility, $m_{i,j}$ mass of emission of pollutant i from facility j, M total mass of pollutants, A activity rate in terms of reference flow (capacity C or production P), f_T fraction of reference flow produced using a particular technology, n number of facilities reporting pollutant i to NEI or TRI, N number of facilities reporting any pollutant to NEI or TRI, N^* number of facilities operating in the USA

a particular pollutant i, from the $N*$ facilities for which a total annual production rate is available. This assumption takes into account a frequently used method for computing emission factors, where the available emission data from one source is divided by any available production data from another source, often in an aggregated form to obtain an average value.

2.5 Impact assessment

The Tool for the Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI) is a widely used tool for life cycle impact assessment (Bare 2013, 2002). The three average emission factors described above were used to compute a suite of impacts commonly reported in LCA studies using the TRACI characterization factors.

3 Results

Two sets of findings from this research are presented. The first compares total sector emissions and emission factors for two fuels. The second is to use the emissions to calculate characterized impact potentials of the emissions using the TRACI characterization method.

In the first finding, the sector emissions from corn biorefining and petroleum refining are compared to determine which sector has higher emissions. Then, the allocated emission factors for the products from the two sectors, ethanol and gasoline, are compared to determine which fuel has higher emissions based on a megajoule of the fuel produced. Finally, the reporting of a particular chemical in the NEI database is discussed for its effect on the emission factor. This reporting statistic is a key for a user to decide on the choice of an emission factor for a given pollutant as will be established from the results.

In the second finding, the impact potentials using TRACI are calculated from the emission factors. Here, ratios have been used to compare the impact potentials to determine whether the use of an emission factor gives a higher impact for a particular fuel. In addition, ratios of impact potentials have been used to compare between the two fuels using a particular set of emission factors.

3.1 Comparison of total emissions and allocated emission factors

There were 92 air pollutants common to the corn and petroleum refining sectors in the NEI database. A subset of greenhouse gases (GHGs) and criteria air pollutants (CAPs) is presented in Fig. 1. A comparison of all 92-air pollutants common to the two sectors is provided in the Supplementary Information (see Table SI1 and Fig. SI1, Electronic Supplementary Material).

The left axis represents the ratios of the two fuels from the total sector emissions ($M_{Corn}/M_{Petroleum}$) and the emission factors ($F_{C_n,E}/F_{C_n,G}$, $F_{C_N,E}/F_{C_N,G}$ and $F_{P_{N*},E}/F_{P_{N*},G}$) for the ith chemical. The ratios are designed so that a value less than 1 denotes lower sector emissions and emission factors for the corn biorefining and ethanol respectively. From Fig. 1, it can be seen that $M_{Corn}/M_{Petroleum}<1$ for all of the pollutants shown. From the corresponding Fig. SI1 (Electronic Supplementary Material) for all 92 pollutants, $M_{Corn}/M_{Petroleum}<1$ for 85 of them. From Fig. 1, for the selected GHGs and CAPs, the ratios of emission factors are all greater than one. This suggests that the allocated emissions from ethanol are higher than gasoline per megajoule of the fuel produced. From the corresponding Fig. SI1 (see ESM), $F_{C_n,E}/F_{C_n,G} < 1$ for 25 of the pollutants, $F_{C_N,E}/F_{C_N,G} < 1$ for 29 of the pollutants, and $F_{P_{N*},E}/F_{P_{N*},G} < 1$ for 39 of the pollutants. Thus, one can conclude that sector wide emissions for corn biorefining are less than petroleum refining for more than 90 % of the common pollutants. However, when considering the emission factors of the individual fuels, more than 50 % of the pollutants (including CAPs and GHGs) are higher for the production of ethanol.

In Fig. 1, the circle (gasoline) and triangle (ethanol) data points follow the right axis which shows the percentage of facilities reporting the particular pollutant (n/N). For certain greenhouse gases like carbon dioxide and methane, the percentage of facilities reporting is quite low (below 50 %). It is not mandatory to report GHGs, and hence the percentage reflects all the facilities that have voluntarily reported their GHG emissions. For other pollutants like VOC, nitrogen oxides, sulfur dioxide, etc., a high percentage of reporting reflects mandatory reporting requirements. In addition, certain pollutants like ammonia, lead, etc., have high reporting percentages, ~80 %. The percentage reporting for all the 92 common pollutants is obtained from Fig. SI1 (see ESM). This value is useful for deciding which emission factor to select for a certain pollutant. When $n<<N$, F_{C_n} gives a higher estimate of emission per megajoule of fuel produced, and F_{C_N} gives a lower estimate of emission per megajoule of fuel produced. When $n\approx N$ either F_{C_n} or F_{C_N} can be used as an emission factor. $F_{P_{N*}}$ should be used only when $N\approx N*$. In such a case, it is advisable to use $F_{P_{N*}}$, where the actual production of the fuels is used instead of capacity. Details on the number of facilities reporting and operating in the USA are given in the Supplementary Information Sheet "SI5 Facilities Reporting."

3.2 Computing environmental and human health impacts using emission factors

Comparing the common pollutants using the method in the previous section is sufficient to conclude which pollutants are released in greater quantity; however, different pollutants have
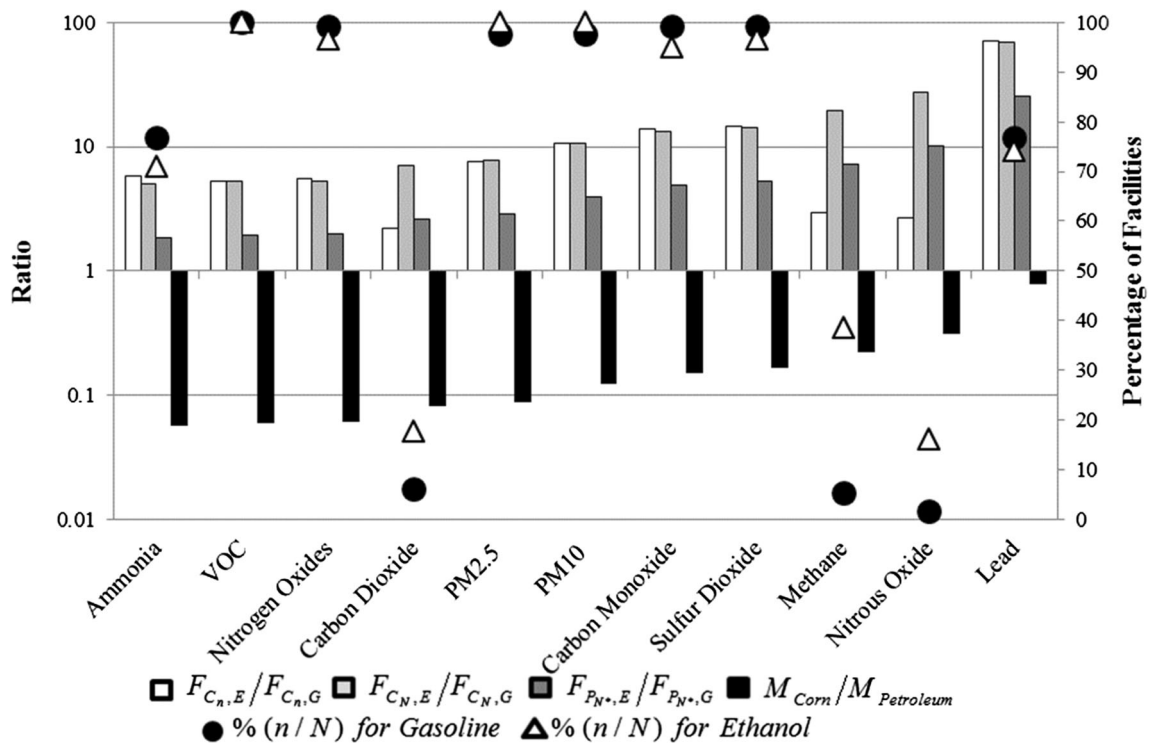
Fig. 1 Comparison of select GHG and CAPs from NEI for ethanol (E) and gasoline (G) production. The *left axis* shows the ratios of the two fuels as represented by the *bars*. The *right axis* shows the percentage of facilities reporting the particular pollutant. See Table 2 for notation

varying impacts, and it is important to consider their relative and cumulative environmental and human health impact potentials. In addition, one needs to consider the impacts from the pollutants which are not common to the two fuel production stages in NEI, as well as the pollutants listed in the TRI. For this, the TRACI impact characterization method was used to

determine the impacts from ethanol and gasoline. This comparison is shown in Figs. 2 and 3. In the figures, the impacts are expressed as a percentage of the impacts calculated as $I_{F_{C_n,E}}$. From the figures, one sees the impacts vary widely depending on the choice of the emission factor in the calculation of impact. Two studies are conducted from these results, the first is on the
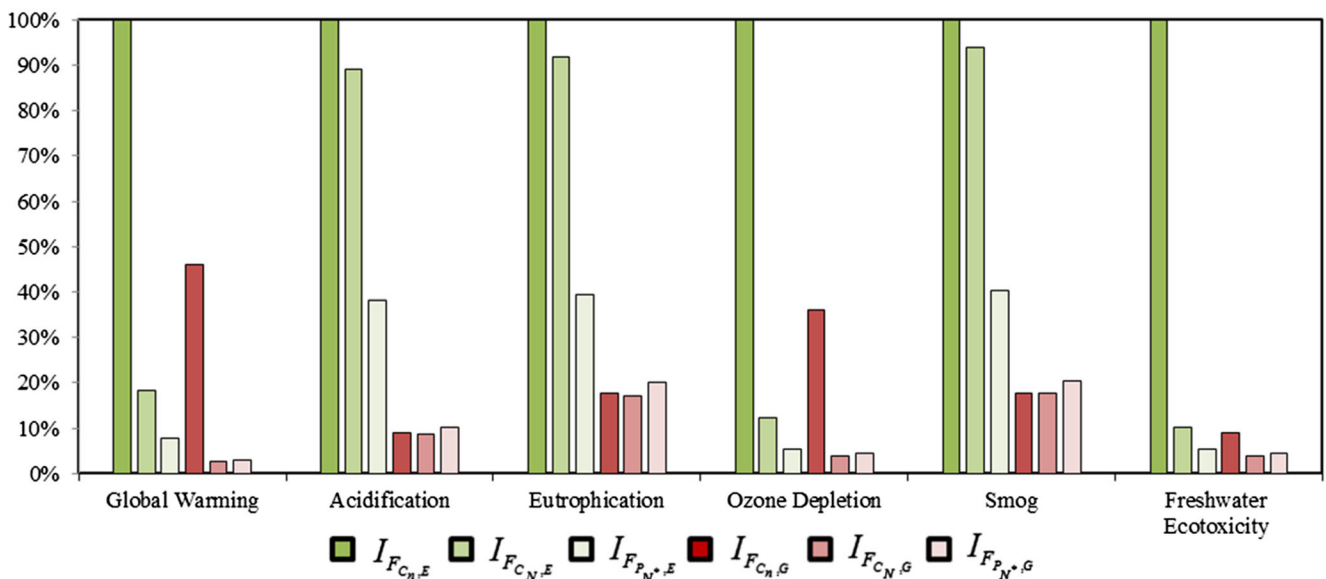

Fig. 2 Environmental impact potentials (I) calculated from ethanol (E) and gasoline (G) emission factors (F) for the three activity-based methods (designated by C_n, C_N, and P_{N*}) expressed as a percentage of $I_{F_{C_n,E}}$
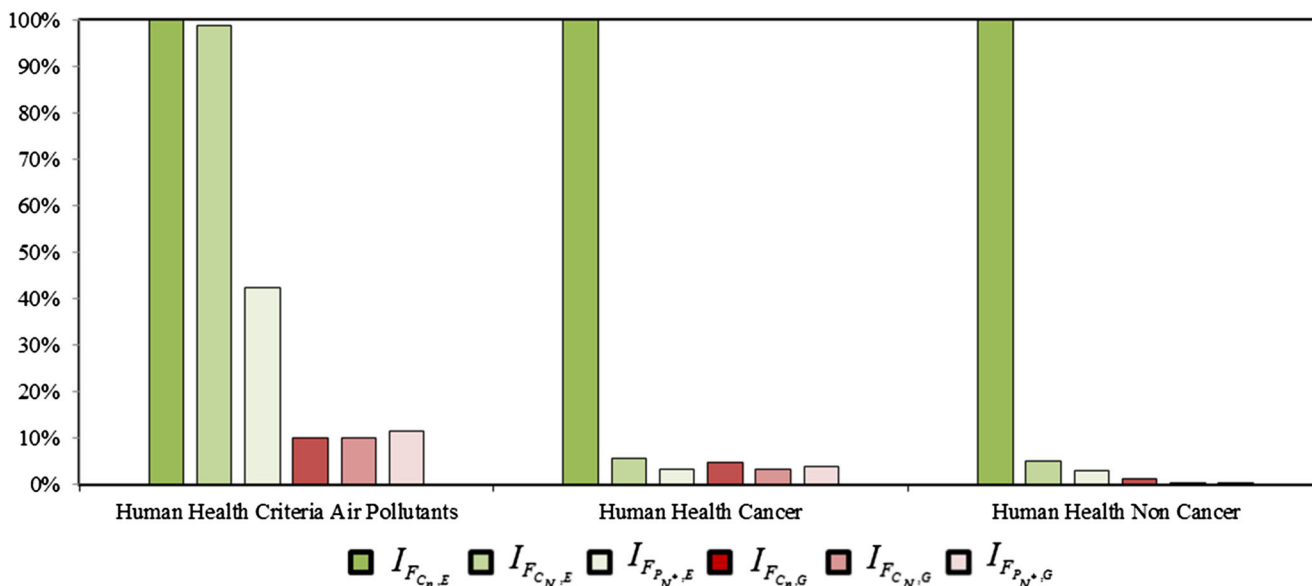
Fig. 3 Human health impact potentials (I) calculated from ethanol (E) and gasoline (G) emission factors (F) for the three activity-based methods (designated by C_n, C_N, and P_{N*}) expressed as a percentage of $I_{F_{C_n,E}}$

effect of choice of activity rate of a particular sector on the impact, and the second is on the comparison of impacts of gate-to-gate production-related emissions for the two fuels.

3.2.1 Effect of activity rate assumptions in sector-average emission factors on impact assessment

The choice of calculation method for determining emission factors can have a significant effect on results. Calculating an average emission factor based on the sum of emissions in an inventory including many facilities within a sector divided by the total production of that sector could lead to significantly under estimating actual values. While further work is needed to explore specific instances of variability across facilities, findings from this work indicate that analysts should take care in reporting results based on emission totals divided by production totals.

The approach presented here provides a simplified means of bounding the range of the average value without the use of more sophisticated statistical analysis. Estimates obtained using the three calculation options by using the potential impact ratios $I_{F_{C_N,E}}/I_{F_{C_n,E}}$ and $I_{F_{P_{N*},E}}/I_{F_{C_n,E}}$ for ethanol and $I_{F_{C_N,G}}/I_{F_{C_n,G}}$ and $I_{F_{P_{N*},G}}/I_{F_{C_n,G}}$ for gasoline are compared. In Table 3 it can be seen that for ethanol production, the ratio of the impact value calculated using the total capacity of all reporting facilities to the impact value calculated using only the capacities of facilities reporting emissions of a particular chemical, $I_{F_{C_N,E}}/I_{F_{C_n,E}}$, is 18 % for global warming, 12 % for ozone depletion, 10 % for freshwater ecotoxicity, and 6 and 5 % for human health cancer and non-cancer, respectively. For gasoline, the corresponding ratio $I_{F_{C_N,G}}/I_{F_{C_n,G}}$ is 68 and 25 % for human health cancer and non-cancer categories,

respectively. For the categories of acidification, eutrophication, smog, and human health criteria air pollutants $I_{F_{C_N,E}}/I_{F_{C_n,E}}$ is at least 89 % and $I_{F_{C_N,G}}/I_{F_{C_n,G}}$ is at least 98 %. In general, the ratios are closer to one for the impact categories of acidification, eutrophication, smog, and human health criteria air pollutants, all of which are closely related to criteria pollutant emissions, suggesting actual values likely more tightly bound by the range than is the case for other impact categories. This is not surprising given that criteria pollutant emissions have been the subject of much closer scrutiny over the years than the wide variety of other toxic species reported in the NEI and TRI. It is important to note that greenhouse gas reporting in the NEI is voluntary, and so one would expect a lower participation rate and perhaps lower data quality for this category.

In the absence of facility-specific data for C_n or C_N, the only possibility is to use aggregate production data, P_{N*}, to approximate the activity rate associated with facilities reporting emissions. Results from this study provide some insight into cases when this may or may not be appropriate. It was observed that the capacity of gasoline production is more than the actual quantity of gasoline produced in the USA, i.e., $C_{N,G} > P_{N*,G}$. Thus, the ratios in Table 3 comparing impacts for gasoline (both with a denominator of $F_{C_n,G}$) indicate that $I_{F_{C_N,G}} < I_{F_{P_{N*},G}}$. However, for ethanol, $N \ll N^*$, leading to $C_{N,E} < P_{N*,E}$, and then the ratios in Table 3 comparing impacts for ethanol (both with a denominator of $F_{C_n,E}$) show more complex behavior. Thus, based on the data, $F_{P_{N*}}$ should be used only: (1) when no emission and matching capacity data are available from specific facilities and (2) when the capacity of N facilities is comparable to the production from N^* facilities. In this example, it would be correct to use $F_{P_{N*},G}$ in place

Table 3 Ratios of impacts calculated using emission factors, expressed as percentages

Impact category	Comparing between impacts for ethanol		Comparing between impacts for gasoline		Comparing between ethanol and gasoline impacts	
	$I_{F_{C_N,E}}/I_{F_{C_n,E}}$	$I_{F_{P_{N*},E}}/I_{F_{C_n,E}}$	$I_{F_{C_N,G}}/I_{F_{C_n,G}}$	$I_{F_{P_{N*},G}}/I_{F_{C_n,G}}$	$I_{F_{C_N,G}}/I_{F_{C_n,E}}$	$I_{F_{C_N,G}}/I_{F_{C_N,E}}$
Global warming	18 %	43 %	6 %	116 %	46 %	14 %
Acidification	89 %	43 %	98 %	116 %	9 %	10 %
Eutrophication	92 %	43 %	98 %	116 %	18 %	19 %
Ozone depletion	12 %	43 %	11 %	117 %	36 %	32 %
Smog	94 %	43 %	99 %	116 %	18 %	19 %
Freshwater ecotoxicity	10 %	52 %	44 %	118 %	9 %	38 %
Human health criteria air pollutants	99 %	43 %	99 %	116 %	10 %	10 %
Human health cancer	6 %	59 %	68 %	117 %	5 %	57 %
Human health non cancer	5 %	62 %	25 %	118 %	1 %	6 %

of $F_{C_N,G}$ (for gasoline), but the impacts using $F_{P_{N*},E}$ in place of $F_{C_N,E}$ (for ethanol) will give erroneous results and should not be used.

3.2.2 Comparison of impacts from gate-to-gate production-related air emissions for ethanol and gasoline

For comparing the impacts between the two fuels, the ratios, $I_{F_{C_n,G}}/I_{F_{C_n,E}}$ and $I_{F_{C_N,G}}/I_{F_{C_N,E}}$, are used. From the results in Table 3, it is clear the choice of an emission factor can greatly affect the conclusions regarding the impacts in the categories of global warming, freshwater ecotoxicity, and human health cancer from a particular fuel. For example, when F_{C_n} is used, the freshwater ecotoxicity impacts from gasoline are only 9 % of that from ethanol, but when F_{C_N} is used, the impacts from gasoline are 38 % of that from ethanol. Revisiting the definitions of emission factors earlier in this paper, one can conclude the different assumptions related to the computation of the emission factors are reflected in the impact calculation. Thus, the user is presented with a range of possible impacts with the use of these different emission factors. From Table 3 one can conclude the gate-to-gate impact potentials due to gasoline production are less compared to ethanol production for all the impact categories when compared on a unit energy basis.

4 Discussion

From the above analysis, computing the different emission factors gives a method for estimating the national average emissions per unit energy of fuel produced. Using this method, more than 200 pollutants could be characterized, and about 170 of these have been characterized using the TRACI impact methods. From the cumulative impact potentials of the emissions in the two fuel production stages, gasoline has lesser impacts than ethanol when compared based on energy content

of the fuel produced. The following could be the reasons for this observation, and can be explored for future research.

First, an order of magnitude analysis on the energy content of the two fuels produced and the total sector emissions is done. The total energy content of gasoline produced is higher by two orders of magnitude compared to ethanol (3.75×10^{13} MJ of gasoline compared to 6.6×10^{11} MJ of ethanol produced in 2008). The total sector emissions for the petroleum sector are higher by two or more orders of magnitude for only 28 of the 92 pollutants in NEI compared to the corn biorefining sector. Thus, for the rest of the pollutants, the emission factor for ethanol is expected to be higher than gasoline and this is reflected in the results of Fig. SI1 (Electronic Supplementary Material).

Secondly, from the data analysis almost all the petroleum refineries operating in the USA report to NEI and TRI databases, while a smaller number of ethanol refineries report to the NEI database, making a comparison based on all operating facilities difficult. A comparable inclusion of nearly 100 % of facilities for the corn biorefining sector will help in accurately determining the total emissions and impacts due to the two fuels.

The maturity of petroleum refineries have resulted in stringent and tight control of their internal processes with emission control equipment to comply with Title V permitting standards (U.S. EPA 1998). The rules for Prevention of Significant Deterioration, Nonattainment New Source Review, and Title V to establish and operate chemical process plants were changed in 2007 to exclude facilities which produce ethanol through a natural fermentation process (U.S. EPA 2007). This rule holds regardless of whether the ethanol is produced for human consumption, fuel or for an industrial purpose, and this may be a possible reason for the high emission factors associated with ethanol production. One could account for the size of the petroleum refineries, which are large and concentrated facilities with different rules guiding their operations compared to smaller, more distributed corn biorefineries.

While the emissions per unit production are larger in corn biorefineries, the facilities are more distributed and so have less environmental justice related concerns. They may be located in lower population areas or areas with less pronounced environmental impacts at present and so may be of less concern overall. Many air emission sources may not be covered or are exempt from various emission controls, reporting, and other requirements. In some cases, the number or stringency of requirements is tiered according to source size or other criteria, and hence not found in either of the databases. Thus, the above factors can significantly affect the national average emission factor calculations. This is possibly the cause for the higher ethanol emission-factor environmental and human health impact potentials compared to gasoline.

Results found here suggest that coordinating the reporting of facility-level emissions and facility-level production would offer important advantages for quality assurance of emission inventories. These values would allow for the calculation of facility-specific, activity-based emission factors at the time of reporting. The results of this calculation could be used to identify instances where a facility's reported emissions highlight it as an outlier from other facilities reporting emissions of a certain chemical species in that or earlier years. While following up on differences across facilities in detail could potentially be time-consuming and therefore overly burdensome, the simple act of offering data providers automated feedback indicating the percent difference between their emission factors for the chemicals reported and sector benchmarks would help prevent instances where erroneous data would otherwise be unknowingly reported.

A key finding of this analysis is that national emissions calculated based on emission inventories and national production totals have the potential to significantly underestimate actual emission factors. The reason for this is bias related to misreported or not reported emissions in national inventories (Environmental Integrity Project 2004; Lombardi and Fuller 2013). This is especially important to consider in interpreting the results of streamlined life cycle assessment or similar studies, which rely on national average emission factors to characterize *reference* supply chains of products. This is relevant for the development of economy-wide LCA models based on combining input–output datasets with national emission inventories (Suh 2005; Hendrickson et al. 1998; Mazzanti and Montini 2010; Lenzen et al. 2013; Tukker et al. 2013; de Haan and Keuning 1996). The reasons for potentially *missing* emissions include facility-specific reporting requirements based on threshold production or emission levels as well as data quality issues including either intentional or unintentional misreporting. For example, Bennear and colleagues have demonstrated instances where industrial actors endeavor to remain just under thresholds in order to avoid reporting requirements (Bennear 2008). The analysis presented here offers one approach for addressing this problem, which can be utilized by LCA practitioners. The

benefit of this approach is that it provides insight into the potential for bias associated with the national average emission factors and allows for bounding the range of possible emission factors using two different emission factor samples, inclusive of all facilities for which matches were found (assuming zero emissions where the emission inventory does not provide data) or limited to only instances where explicit emission values can be associated with production information.

5 Conclusions

A rigorous method of calculation of national average emission factors for estimating environmental impacts of products from industrial sectors has been presented in this paper. The key finding from this work is the potential for underreporting in national emission inventories results in significant potential for underestimation of emission factors when calculated on a sector average basis. This problem is especially acute for input–output life cycle assessment (IO-LCA) techniques, where often national emission inventories are "summed up" and divided by national total activity rates to get average emission factors. This analysis demonstrates an approach, which could be used to improve the estimation of sector average emission factors for use in process and IO-LCA studies. This paper discusses when a particular emission factor should be used based on appropriate assumptions and the situations where a type of emission factor should be avoided.

The influence of emission factors on the computation of environmental and human health impact potentials is discussed, where the differences that arise from the data choices are explained. Two fuel production sectors, corn biorefining for ethanol using the dry grind technology and petroleum refining for gasoline, are compared as a case study to demonstrate the computation and use of the emission factors. More than 200 chemical species were available for characterization and more than 170 chemical species were characterized using the TRACI methodology, expanding the scope of research on environmental impact assessment for ethanol and gasoline. Using the TRACI characterization factors, this analysis helped to determine the impact in six environmental and three human health categories. Thus, this method of emission estimation can be used to develop the inventories for products from a manufacturing sector. Such results can be directly used in the LCA stage of production, especially for the design of sustainable supply chains of products.

Disclaimer The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

References

Bare JC (2002) TRACI: the tool for the reduction and assessment of chemical and other environmental impacts. J Ind Ecol 6:49–78

Bare JC (2013) Tool for the Reduction and Assessment of Chemical and other Environmental Impacts (TRACI). http://www.epa.gov/nrmrl/std/traci/traci.html. Accessed 2 July 2013

Bennear LS (2008) What do we really know? The effect of reporting thresholds on inferences using environmental right-to-know data. Regul Gov 2:293–315

Curran MA (2012) Sourcing life cycle inventory data. In: Curran MA (ed) Life cycle assessment handbook: a guide for environmentally sustainable products. Wiley, Hoboken, pp 105–141

de Haan M, Keuning S (1996) Taking the environment into account: the NAMEA approach. Rev Income Wealth 42:131–148

Environmental Integrity Project (2004) Gaming the system how off-the-books industrial upset emissions cheat the public out of clean air. http://www.environmentalintegrity.org/pdf/publications/eip_upsets_report_appendixa.pdf. Accessed 2 July 2013

EPM (2011) USA Plants. http://www.ethanolproducer.com/plants/listplants/USA/. Accessed 2 July 2013

Hagelaar G, van der Vorst J (2001) Environmental supply chain management: using life cycle assessment to structure supply chains. Int Food Agribus Man 4:399–412

Hawkins T, Ingwersen W, Srocka M, Transue T, Paulsen H, Ciroth A (2013) Development of tools to promote efficient life cycle assessment: OpenLCA and the LCA harmonization tool. Paper presented at the ICOSSE '13 - 3rd International Congress on Sustainability Science & Engineering, Cincinnati, OH

Hendrickson C, Horvath A, Joshi S, Lave L (1998) Economic input–output models for environmental life-cycle assessment. Environ Sci Technol 32:184A–191A

Ingwersen W et al. (2013) LCA models at the core of the sustainable supply chain toolkit: a paper towel case study. Paper presented at the ISSST Conference, Cincinnati, OH, 16th May 2013

Kim J, Yang Y, Bae J, Suh S (2013) The importance of normalization references in interpreting life cycle assessment results. J Ind Ecol 17:385–395

Kwiatkowski J, McAloon A, Taylor F, Johnston D (2006) Modeling the process and costs of fuel ethanol production by the corn dry-grind process. Ind Crop Prod 23:288–296

Lenzen M, Moran D, Kanemoto K, Geschke A (2013) Building EORA: a global multi-region input–output database at high country and sector resolution. Econ Syst Res 25:20–49

Lombardi K, Fuller A (2013) 'Upset' emissions: flares in the air, worry on the ground. http://www.publicintegrity.org/2013/05/21/12654/upset-emissions-flares-air-worry-ground. Accessed 2 July 2013

Mazzanti M, Montini A (2010) Embedding the drivers of emission efficiency at regional level — Analyses of NAMEA data. Ecol Econ 69:2457–2467

Miranda ML, Keating MH, Edwards SE (2008) Environmental justice implications of reduced reporting requirements of the toxics release inventory burden reduction rule. Environ Sci Technol 42:5407–5414

Mueller S (2010) 2008 National dry mill corn ethanol survey. Biotechnol Lett 32:1261–1264

Nwe E, Adhitya A, Halim I, Srinivasan R (2010) Green supply chain design and operation by integrating LCA and dynamic simulation. In: Pierucci S, Ferraris GB (eds) Computer aided chemical engineering, vol 28. Elsevier, pp 109–114

Suh S (2005) Developing a sectoral environmental database for input–output analysis: the comprehensive environmental data archive of the US. Econ Syst Res 17:449–469

Tukker A et al (2013) EXIOPOL – Development and illustrative analyses of a detailed global MR EE SUT/IOT. Econ Syst Res 25:50–70

U.S. EIA (2011) Petroleum & other liquids. http://www.eia.gov/petroleum/data.cfm. Accessed 1 March 2011

U.S. EPA (1998) Air pollution operating permit program update key features and benefits. United States Environmental Protection Agency, OAQPS, Research Triangle Park

U.S. EPA (2007) Prevention of significant deterioration, nonattainment new source review, and Title V: treatment of certain ethanol production facilities under the "Major Emitting Facility" definition; final rule vol 72. Federal Register

U.S. EPA (2013a) Emissions factors & AP 42, compilation of air pollutant emission factors http://www.epa.gov/ttn/chief/ap42/index.html. Accessed 2 July 2013

U.S. EPA (2013b) Air trends. http://www.epa.gov/airtrends/sixpoll.html. Accessed 2 July 2013

U.S. EPA (2013c) National Emissions Inventory (NEI) Air pollutant emissions trends data. http://www.epa.gov/ttnchie1/trends/. Accessed 2 July 2013

U.S. EPA (2013d) Laws and regulations. http://www.epa.gov/emergencies/lawsregs.htm. Accessed 2 July 2013

U.S. EPA (2013e) Toxics Release Inventory (TRI) Program. http://www.epa.gov/tri/. Accessed 2 July 2013

U.S. EPA (2013f) 2008 national emissions inventory, version 3 technical support document. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. http://www.epa.gov/ttn/chief/net/2008neiv3/2008_neiv3_tsd_draft.pdf

You F, Tao L, Graziano D, Snyder S (2012) Optimal design of sustainable cellulosic biofuel supply chains: multiobjective optimization coupled with life cycle assessment and input–output analysis. AIChE J 58:1157–1180

Young DM, Hawkins T, Ingwersen W, Lee S-J, Ruiz-Mercado G, Sengupta D, Smith RL (2012) Designing sustainable supply chains. Chem Eng Trans 29:253–258