

Approaches to fill data gaps and evaluate process completeness in LCA—perspectives from solid waste management systems

Henriksen, Trine; Levis, James W.; Barlaz, Morton A.; Damgaard, Anders

Published in: International Journal of Life Cycle Assessment

Link to article, DOI: 10.1007/s11367-019-01592-z

Publication date: 2019

Document Version Peer reviewed version

Link back to DTU Orbit

Citation (APA):

Henriksen, T., Levis, J. W., Barlaz, M. A., & Damgaard, A. (2019). Approaches to fill data gaps and evaluate process completeness in LCA—perspectives from solid waste management systems. *International Journal of Life Cycle Assessment*, 24(9), 1587-1601. https://doi.org/10.1007/s11367-019-01592-z

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.

- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Approaches to fill data gaps and evaluate process completeness in LCA -Perspectives from solid waste management systems

Trine Henriksen^{a*}, James W. Levis^b, Morton A. Barlaz^b, Anders Damgaard^a

^a Technical University of Denmark, Department of Environmental Engineering, Bygningstorvet 115, 2800 Kongens Lyngby, Denmark

^bNorth Carolina State University, Department of Civil, Construction, and Environmental Engineering, Campus Box 7908, Raleigh, NC 27695-7908, USA

*Corresponding author. Phone: +45 4525 2169; E-mail: trinhen@env.dtu.dk

Acknowledgements

The research was financially supported by the Technical University of Denmark through a scholarship from the R98 foundation.

1 Abstract

Purpose Large data amounts are required in an LCA, but often, site-specific data are missing and less representative
surrogate data must be used to fill data gaps. No standardized rules exist on how to adress data gaps and process
completeness. We suggest a systematic evaluation of process completeness, identification of data gaps, and application
of surrogate values to fill the gaps. The study focus on foreground process data.

6 Methods A solid waste management (SWM) scenario was used to illustrate the suggested method. The expected input 7 and output flows in a waste incineration model were identified based on legislation and expert judgement, after which 8 process completeness scores were calculated and missing flows identified. To illustrate the use of different types of 9 surrogate data to fill data gaps, data gaps were selected for 16 different parameters in five SWM processes. We 10 compared the global warming potential (GWP) from using surrogate data, and from leaving the gap, to identify the data

11 gaps where representative surrogate data should be used.

Results and discussion The completeness score for the material inputs to waste incineration was 78%, and the missing flows were auxiliary fuels and precipitation chemicals. The completeness score for air emissions were between 38% and 50% with and without expert judgement. If only greenhouse gases were considered (CO₂, CH₄, and N₂O), the completeness score would be 67%. Applying weighting factors according to the greenhouse gas contribution in the US gave a completeness score of 94%. The system-wide data gaps, where representative surrogate data should be applied, were the CH₄ release from composting; electricity generation efficiency of incineration; recovery efficiencies at a material recovery facility; and composition of the plastic, metal, and paper fractions in the household waste; in these

19 cases, leaving the gap changed the GWP results by >5%.

20 Conclusions Completeness evaluation should take into account the relevance and importance of flows; relevance

21 depends on the considered life cycle impact methods and importance depends on the weighting of the different flows.

22 The set of expected flows and evaluation of relevance and importance must be documented in a transparent manner.

23 The choice of surrogate values to fill data gaps depends on the availability of secondary data and on whether the data

24 gap matters, i.e. significantly affects the LCA results. The suggested method can be used to properly document the

25 identification of missing flows, and to select and apply surrogate values to fill the data gaps.

26

27 Keywords: Completeness, data gaps, surrogate values, representativeness, waste management, LCA

28 **1** Introduction

29 There is an increasing need for predictive models to support environmental policy and decision making. The validity of these models is dependent on the availability of reliable data generated in a valid manner and relevant to the study 30 context. Life cycle assessment (LCA) is an example of a data intensive modeling framework that is often used for 31 environmental policy analysis. It is an analytical framework used to estimate and compare the potential environmental 32 33 impacts of products, processes, and systems. LCAs cover multiple impact categories (e.g., global warming, resource 34 use, eutrophication) as well as upstream and downstream processes to assess and explore trade-offs between different 35 impacts and life cycle stages (e.g., use and end-of-life). The comprehensive nature of LCA implies the need for large 36 amounts of data representing both foreground and background processes. As a result, missing data (i.e. data gaps) are 37 common. Data gaps can be defined as values that are unavailable but would improve the accuracy and reliability of 38 LCA results if they were included (Little et al., 2012). Data gaps that are ignored, i.e. not filled by some alternative 39 data, may bias the results of an LCA. As described by Zhou et al. (2014), even when data are missing, the goal remains to make conclusions about the facilities and processes covered by the goal and scope of the analysis. Thus, to ensure 40 41 that reliable conclusions are drawn from an LCA study, it is important that the practitioner identifies data gaps and 42 applies methods to fill them.

Data gaps may occur in the life cycle inventories (LCIs) or life cycle impact assessment (LCIA) methods. The focus of 43 44 this study is on data gaps in LCIs, which include entire processes or single data points within a process (Hischier et al., 45 2001). For example, an LCA practitioner may neglect the emissions and material use related to the process of 46 constructing a garage when modeling waste collection. A single data point, that is typically unknown and likely neglected, is the benzene emission associated with waste collection vehicles. There are principally two types of missing 47 48 data: complete lack of data and lack of site-specific (i.e. technologically, temporally, and/or geographically 49 representative) data (Huijbregts et al., 2001). In the case of a complete lack of data, no secondary data are available to 50 fill the data gap, while in the case of a lack of site-specific data, it is possible to apply secondary data as surrogates. A 51 common situation in which a lack of data occurs is when novel or future technologies are assessed or when there is a 52 need to include emissions that are not monitored; complete lack of data can therefore include 'unknown unknowns', i.e. data that are missing without the awareness of the practitioner, due to ignorance about the processes under study (e.g., a 53 54 practitioner may be unaware of the potential presence of chlorofluorocarbons (CFCs) in landfill gas [Hodson et al., 55 2010]). Lack of site-specific data, e.g. measured at the operated process, is frequent in LCA and can include 'known 56 unknowns', as the data need is known but only secondary, potentially non-representative, data are available (e.g., a

57 practitioner may be aware of potential CFC emissions in landfill gas, but have no data on the CFC emissions from the 58 landfill under study). Scarce or missing data in LCAs of waste management systems include complete lack of data, such 59 as emissions from wastewater treatment plants due to limited substance coverage of monitoring programs (Yoshida et 60 al., 2014), and lack of site-specific data, such as the physico-chemical composition of local household waste material fractions (Götze et al., 2016). Thus, missing data is a challenge when modeling waste management systems, particularly 61 62 because of the local specificities and the linked nature of the waste composition, collection, and treatment processes. 63 The ISO 14044 standard requires that missing data is treated by a justified application of a non-zero value, a zero value, or a calculated value (ISO, 2006). A number of publications have demonstrated methods to fill data gaps, including the 64 65 use of proxies, extrapolation methods, regression models, expert elicitation, and input-output data (Canals et al., 2011; Majeau-Bettez et al., 2011; Moreau et al., 2012; Steinmann et al., 2014; Subramanian and Golden, 2016). These 66 67 methods differ in level of sophistication, and previous studies have not yet compared different approaches to fill the 68 same data gap.

69 The occurrence of data gaps is closely linked with data completeness. Completeness is a data quality requirement in 70 ISO 14044, defined as the 'percentage of flow that is measured or estimated' (ISO, 2006), which allows for multiple 71 interpretations. The data quality assessment scheme used for the ecoinvent database evaluates completeness as the 72 quality of a single flow in terms of whether the data sampling represents variations over time and space (Weidema et 73 al., 2013). For example, NO_x emissions from waste collection fluctuate depending on the type of truck, duty cycle, and 74 collection system for a specific city, which should ideally be estimated using a sufficient sampling size and period. The 75 ILCD Handbook (JRC, 2010) and the US Environmental Protection Agency (US EPA) (Edelen and Ingwersen, 2016) define completeness at the process level as the number of included substances to model a process relative to the number 76 77 of expected substances for a process. In this study, we will focus on completeness at the process level. Few publications 78 describe how to calculate completeness scores for process inventories in practice. The approach by Edelen and 79 Ingwersen (2016) includes different types of data used to model a unit process (e.g. emissions, raw inputs, and reference 80 product). It provides a generic method for weighting completeness scores that can be adjusted to consider the relevant 81 priorities of a project stakeholder (e.g. a study may weigh the completeness of air emissions stronger than the 82 completeness of intermediate inputs). However, Edelen and Ingwersen (2016) do not offer an approach to identify the expected flows in the calculation of the completeness score. Other authors provide examples of using the method by 83 84 JRC (2010) to calculate the completeness of fuel and electricity datasets regarding substance flows to and from the 85 environment (Fazio et al., 2015; Garraín et al., 2015a, 2015b). The authors first conducted a sectorial analysis to

identify the substances that should be included in the process inventory, they then listed the substances actually
included according to how many impact categories they cover, and finally they compared the number of included
substances relative to the number of expected substances per impact category (they did not include weighting). While
publications on process completeness base the estimation of expected substances on expert judgement, substances
identified by environmental regulation and monitoring programs are also relevant. In this study, we suggest a systematic
approach to evaluate completeness that considers current and potential regulations, existing literature, and expert
judgement.

93 This study addresses foreground process data, and site-specific data gaps that can be identified and filled using 94 secondary surrogate data to build more complete lifecycle inventories. The aim of this study is to suggest a combined 95 systematic approach to calculate process completeness, identify missing flows (data gaps), and apply surrogate values to 96 fill the data gaps. A solid waste management (SWM) scenario is used as case study to illustrate the method. The 97 specific objectives are to 1) describe methods to fill data gaps and calculate completeness scores, 2) calculate process 98 completeness scores and identify missing flows for the waste incineration process as an example, and 3) suggest 99 methods and surrogate values to fill the data gaps for 16 parameters in five SWM processes. The next section describes 100 methods to fill data gaps and calculate completeness scores, followed by a description of the case study, and a 101 presentation and discussion of the study results.

102 **2 Methods**

103 2.1 Published methods to fill data gaps

Methods used to approximate the missing data in the SWM case study include deterministic approaches based on single-point values and intervals as well as stochastic approaches using statistical distributions. Table 1 shows a glossary of terms for the approaches used to fill data gaps in this study and each is described in this section.

Direct and average proxies require relatively little data (Canals et al., 2011). A direct proxy consists of using another value from a different process or system in place of the missing value, while an average proxy uses some value of central tendency (e.g., mean, median, mode) of a set of direct proxies. If more than one proxy representing the same process is available, an interval can be obtained (e.g. using the standard deviation). Scaled proxies are typically used for product portfolios, production volumes, and other compositional data. Compositional data contains relative information about the parts of a whole, given as proportions and percentages. In the SWM field, scaled proxies have been used to construct food waste composition by defining a set of food item proxies per food category. For example, Tonini et al.

- 114 (2017) used beef, pork, and chicken as proxies for the entire fraction of meat and meat products, where other meats
- 115 were considered negligible. Regression techniques can also be used to derive proxies, e.g. Moreau et al. (2012)
- predicted material flows for power plants from relationships between properties of different plants.
- 117 Min-max intervals cover the extreme observations of a sample. For example, min-max intervals were used to establish
- 118 landfill LCIs representing the expected technological and geographical variability within a given scope of study
- 119 (Henriksen et al., 2017). Alternatively, statistical distributions may be developed for the input values. When data are
- 120 limited, uniform or triangular distributions are often used to create approximate distributions (Bisinella et al., 2017;
- 121 Clavreul et al., 2013; Laner et al., 2015). If the sample is sufficiently large, fitting a statistical distribution to the
- 122 available data provides an estimate of the uncertainty of the value.
- 123 **Table 1** Glossary of terms for the methods used to fill data gaps in the case study

Proxy	Available surrogate data intended to reasonably represent missing data based on an underlying understanding of the process, material, or system
Direct proxy	Use of surrogate data without any modification of its original value(s)
Average proxy	Use of a measure of central tendency of multiple surrogate data values (i.e. arithmetic mean, median, mode)
Scaled proxy	Linear scaling of known part of some compositional data to cover an unknown part of the data
Min-max interval	Extremes of a data interval based on measured data, literature, or estimates
Statistical distribution	Fitting a statistical distribution to a data sample (normal, lognormal, triangular, etc.)
Leaving a gap	A gap in data is neglected because the practitioner is unable or unwilling to use any surrogate data to fill a data gap

124

125	In addition, models can be used to approximate missing data by describing the relationship between independent and
126	dependent variables. For example, the energy use at material recycling facilities (MRFs) was estimated from mass flows
127	to each piece of equipment and energy use per unit mass of material processed (Pressley et al., 2015). A limitation of
128	using models is that detailed information about sub-processes in a facility is required, which - if not available - creates
129	even more data gaps. Input-output data have been used to address background data gaps (Moberg et al. 2014), but may
130	overestimate the actual impact when including emissions from other sectorial processes. Finally, leaving a data gap is
131	an alternative to filling the gap, which, however, may favor the least documented process in an LCA (Moreau et al.,
132	2012).

133 2.2 Method to calculate process completeness scores

134 The applied method to calculate completeness scores follows the principles reported by the US EPA (Edelen and Ingwersen, 2016) and ILCD Handbook (JRC, 2010). The basic principle is to calculate process completeness as the 135 number of included flows relative to the number of expected flows. Flows here are distinguished as elementary 136 137 (exchanges with the environment) and intermediate (between industrial processes). In this study, elementary and 138 intermediate flows are assessed separately, because the former targets the impact coverage of the included data. 139 Equation 1 is used to calculate the completeness score for process p, and Eq. 2 calculates the system completeness by 140 dividing the total number of included flows for all P processes by the total number of expected flows for all P processes. 141 Process and system completeness can be calculated using either elementary or intermediate flows either as a whole or

142 for specific impact categories.

$$Process Completeness_{p} = \frac{Number of Included Flows_{p}}{Number of Expected Flows_{p}}$$
(1)

System Completeness =
$$\frac{\sum_{1}^{P} \text{Number of Included Flows}_{p}}{\sum_{1}^{P} \text{Number of Expected Flows}_{p}}$$
(2)

Depending on the type of LCI model, Eq. 1 and 2 may also be used to assess the completeness of flows modeled indirectly, i.e. flows that are a function of parameters such as transfer coefficients and other efficiencies. The set of expected flows should be identified in a transparent manner and be based on published sources and expert judgement. The next section describes a decision-support framework for evaluating process or system completeness that includes the identification of expected flows.

148 **2.3 Decision-support framework**

149 The decision-support framework shown in Figure 1 illustrates the suggested method to evaluate process completeness, 150 identify missing flows (data gaps), and apply surrogate values to fill the data gaps. The method consists of four steps, and we recommend that a user sets up a similar diagram when conducting an LCA. Step 1 is the initial modeling of the 151 152 processes under study, where the available input and output data are collected. The required data depend on the goal and 153 scope of the study, e.g., in a carbon footprint study only elementary exchanges of greenhouse gases are required. Step 2, 154 the identification of the expected flows for a process p, help to establish the required data. Legislative documents 155 contain lists of monitored substances, which are relevant to the process and should be included in the model, and other 156 relevant substances can be found in literature and based on expert judgement. In step 3, the included flows and set of 157 expected flows are used to calculate a process completeness score and identify the missing flows (i.e. data gaps). In step 158 4, the data gap is characterized as either a complete gap or a site-specific data gap, as this influences the ability to apply

159	secondary data as surrogates	. If it is a complete data gap,	a rough estimate can be used. I	f it is a site-specific data gap,
-----	------------------------------	---------------------------------	---------------------------------	-----------------------------------

- secondary surrogate data can be collected; starting with, potentially lower-quality, easily available data. The user should
- 161 then further evaluate the data gaps that significantly change the results (e.g., >0.1, 1 or 5% change) if not filled by
- surrogate data. Final surrogate data values should be selected for the important data gaps based on a comparison of the
- representativeness of the possible surrogate values (site-specific gap) or by refining the rough estimate, e.g. by
- 164 consulting an expert (complete gap). For the data gaps not considered important (<0.1, 1 or 5% change), it is sufficient
- to use lower quality surrogate data (or to leave the gap un-filled), as this will not significantly alter the LCA results.

166

167



169 Fig. 1 Decision-support framework diagram with the method applied in this study

170 **3** Case study

171 3.1 Solid waste management scenario

A SWM scenario (Figure 2) was modeled to illustrate the application of the suggested method in Figure 1. The scenario 172 173 represents the management of household waste in the US, with the functional unit being the collection, transport, and treatment of 1 metric ton of household waste including disposal of final residues. The included processes are waste 174 175 generation and source separation, collection of waste fractions, sorting at a MRF, composting of yard and food waste, 176 remanufacturing of recyclable materials, and incineration of the residuals with energy recovery. The LCI is documented 177 in the Supplementary Information (SI). The SI also contains an analysis of an additional SWM scenario where the 178 residual waste is landfilled. The second scenario was added to ensure generalizability for the proposed approach, and 179 the SI (section 3) includes a comparison of both scenarios. A consequential modeling approach was applied, hence, 180 multifunctionality was addressed by system expansion to account for avoided products. Mainly North American data 181 were used as a baseline, but the outcomes related to filling data gaps and calculating process completeness are generally

applicable to the LCA field. The LCA model EASETECH was applied (Clavreul et al., 2014).



183

Fig. 2 Flow diagram for the SWM scenario used as case study. Squares with solid lines indicate waste treatment
 processes and ovals indicate material flows. Dotted lines indicate the substitution of marginal energy, virgin materials,
 or fertilizer by recovered energy or secondary materials from the waste treatment. Arrows indicate collection from
 households and inter-facility transportation. Data gaps are selected for the five processes with bold edges

188

189 The four steps of the decision-support framework shown in Figure 1 was demonstrated as follows: The scenario was

190 modeled with the included data stemming from available secondary sources (step 1). Next, the waste incineration

191 process was used as an example to estimate the set of expected flows, calculate the completeness score, and identify the

192 missing flows (step 2 and 3). Then, 16 different parameters were selected as site-specific data gaps in five processes in

- the scenario to provide multiple examples of applying different approaches and surrogate data to fill data gaps (step 4).
- 194 Lastly, final surrogate values were selected based on their representativeness and other data quality aspects (step 4).
- 195 To present the results from using surrogate data to fill the data gaps, we applied the global warming potential (GWP)
- 196 metric. The purpose of this study was to compare approaches to fill the data gaps, rather than to present a full LCIA.
- 197 Thus, a carbon footprint study was adequate to illustrate the methodology. We quantified the GWP using the method
- developed by the Intergovernmental Panel on Climate Change as recommended in Hauschild et al. (2012).

3.2 Included and expected flows for the waste incineration process

To identify the missing flows and evaluate the completeness for the waste incineration process, the included and expected intermediate and elementary flows were identified (Table 2); this corresponds to step 1 and 2 in Figure 1. The intermediate flows were material inputs to the waste incineration plant, and the elementary flows were air emissions via

- the flue gas. Relevant expected flows were identified based on legislative documents.
- The reference for the intermediate flows was the former Danish register in which Danish waste incineration facilities were legally obliged to report annual consumption data (Danish EPA, 2010). The set of expected intermediate flows was identified based on an internal review of data reported by 10 Danish incineration plants from 1995 to 2016.
- 207 Multiple references were used to obtain an aggregated set of expected elementary flows, with and without additions
- based on expert judgement. Under the Clean Air Act, the US EPA has developed New Source Performance Standards
- for large municipal waste combustors (US EPA, 2006), named 'US standards' in Table 2. Within Europe, the Integrated
- 210 Pollution Prevention and Control (IPPC) directive regulates emissions from industrial facilities, including dedicated
- 211 waste incineration plants (European Parliament and Council, 2010), which contain emission limit values. Also in
- Europe, waste incineration plants must comply with the European Pollutant Release and Transfer Register (PRTR)
- (European Commission, 2006), which requires reporting of emissions that may or may not be regulated by the IPPC.
- 214 The additional flows from expert judgement were provided by Michael Van Brunt (Sr. Director of Sustainability,
- 215 Covanta Energy Corporation). The additional flows included emissions that may be regulated in local/regional
- 216 jurisdictions in the US (e.g. states and air quality districts), due to specific conditions. For example, individual states
- 217 may regulate emissions that are not regulated at the national level. Flows were therefore aggregated into two sets; one
- 218 with and one without flows based on expert judgement.

Table 2 Included input-output flows in the waste incineration model as well sets of expected intermediate and elementary
 flows. The full names/chemical formulas are explained in the SI

Included flows					
Intermediate (material use at plant)		Elementary (emissions to air)			
Activated carbon, calc electricity, heat, NH ₃ ,	ium agent (CaCO ₃ , Ca(OH) ₂), scrubber (NaOH, HCl), water	As, Cd, CH ₄ , CO, CO ₂ , Cr, Cu, dioxins, particles, HCl, HF, Hg, NH ₃ , Ni, NO ₂ , Pb, PM, Sb, SO ₂ , TOC			
Expected flows					
Intermediate (material	use at plant)				
Danish register	Activated carbon, calcium agent (CaCO ₃ , CaO, Ca(OH) ₂), electricity, heat, fuel (gas, oil, biomass), NH ₃ , precipitation chemicals (TMT15, FeCl ₃ , FeSO ₄), scrubber (NaOH, NaHCO ₃ , HCl), water				
Elementary (emissions	s to air)				
US standards	Dioxins/furans, Cd, Pb, Hg, particles, HCl, SO ₂ , NO _x , CO, CO ₂ , N ₂ O, CH ₄				
IPPC directive	As, Cd, Co, Cr, Cu, dioxins, furans, particles, HCl, HF, Hg, Mn, Ni, NO ₂ , Pb, Sb, SO ₂ , Tl, TOC, V				
PRTR register	As, C ₆ H ₆ , Cd, CH ₄ , CO, CO ₂ , Cr, Cu, dioxins/furans, HCB, HCl, HF, Hg, NH ₃ , Ni, NMVOC, N ₂ O, NO ₂ /NOx, PAH, PM10, Pb, PCB, PeCB, PER, SO ₂ /SOx, C ₂ H ₃ Cl, Zn				
Expert judgement	'US standards' + Be, Cr, Cr VI, As, Ni, NH ₃ , HF, Zn, Sb, Se, V, Cu, Mn, C ₂₀ H ₁₂ , PAH,				
	VOC, F ⁻ , HBr, Al, P, PCBs, Sn, Si, Ba, Tl, C ₂ Cl ₄ , CCl ₄ , C ₂ H ₃ Cl, C ₂ HCl ₃				
Sum without	Dioxins/furans, Cd, Pb, Hg, particles, HCl, SO ₂ , NO ₂ , PAH, PCB, PeCB, PER, NO _x , CO,				
experts addition	CO ₂ , N ₂ O, CH ₄ , As, Cd, Co, Cr, Cu, HF, Hg, Mn, Ni, NO2, Sb, Tl, TOC, V, C ₆ H ₆ , NH ₃ ,				
	HCB, Ni, NMVOC, C ₂ H ₃ Cl, Zn				
Sum with experts	'Sum without' + Be, Se, C ₂₀ H ₁₂ , F ⁻ , HBr, Al, P, Sn, Si, Ba, C ₂ Cl ₄ , CCl ₄ , C ₂ HCl ₃				
addition					

221 IPPC=Integrated Pollution Prevention and Control, PRTR=European Pollutant Release and Transfer Register,

3.3 Data gaps in the case study

223 We selected hypothetical data gaps for 16 parameters in five processes in the SWM scenario (highlighted in Figure 2).

224 The data gaps are examples of missing site-specific data in the modeling of SWM processes. Some parameters are

included as gaps (e.g. CH₄ emissions from composting and MRF recovery rates) to highlight the importance of using

226 representative surrogates, as SWM models are often sensitive to these parameters. The data gaps are described in this

227 section.

228 The waste generation process describes the household waste composition, which influences the environmental impacts

of waste treatment and disposal (Bisinella et al., 2017). The waste composition was developed from a residual waste

230 sorting study for Wake County, North Carolina, and data on recyclable waste composition for municipalities within the

county (SCS Engineers, 2014). Municipal solid waste composition data are available in the literature, e.g. the World

Bank includes six fractions (Hoornweg and Bhada-Tata, 2012) and other studies include more fractions (Burnley, 2007;

- Edjabou et al., 2014, US EPA, 2016). However, detailed waste composition data, e.g. the proportion of different plastic
- polymers, is limited, which leads to a frequent use of surrogate data to fill gaps. Consequently, we selected data gaps for
- the composition of the food, paper, plastic, ferrous metal, aluminum, and glass fractions. Specific considerations on
- how to approach the data gaps were developed for the waste composition. As the data gap is the composition of the

waste fractions, the proportions of the overall fractions were known, i.e. food (9%), paper (17%), plastic (12%), metal
(4%), and glass (7%). The unknowns were the proportions of the *constituents* of the fractions; e.g. for food waste, the
proportion of non-meat versus meat.

240 The MRF process controls the difference between the amount of materials collected for remanufacturing and the 241 amount sent to remanufacturing. MRF types are single (receiving a commingled recyclables stream), mixed (receiving 242 mixed waste containing recyclables and non-recyclable wastes), dual (receiving two streams of recyclables [e.g., mixed 243 containers and presorted paper] for separate treatment), and presorted MRFs (receiving source-separated streams that go 244 through final sorting and bailing) (Pressley et al., 2015). The performance of MRFs depends on individual design and 245 operational configurations governed by the waste to be treated, which is why there is no single industry-average dataset 246 (Pressley et al., 2015). This suggests the need for site-specific data that are often not available. To assess the influence 247 of MRF data gaps, we selected data gaps for material recovery efficiencies as well as diesel and electricity use.

The paper, plastic, glass, and metal remanufacturing processes in the system are based on relatively old data representing the US (RTI International, 2003). Data for the conversion of waste materials to secondary products are generally limited and variable (Brogaard et al., 2014). Hence, we selected data gaps for direct CO₂ emissions as well as electricity and natural gas use in remanufacturing. Instead of analyzing all remanufacturing processes, we used cardboard remanufacturing as an illustrative example. The influence of using remanufacturing surrogate data for other types of fiber (newsprint, office paper, etc.) was assessed.

The composting process included food and yard waste from households. Over 60% of yard waste is composted in the US (US EPA, 2016). Still, LCI data for composting is limited, e.g. material use, energy use, and input-specific emissions. Furthermore, composting technologies, e.g. open and enclosed, differ in the level of environmental exchanges (Boldrin et al., 2009). Thus, site-specific composting data is relevant but its availability is often limited. We selected data gaps for diesel and electricity use as well as input-specific greenhouse gas emissions. The land application of the finished compost was included in the model and was assumed to substitute commercial fertilizers, similar to other studies (e.g. Yoshida et al., 2016).

The waste incineration process recovers electricity and metals from the bottom ash. Relevant incineration data are material consumption, energy recovery, and flue gas emissions. Flue gas emissions are regulated and therefore not expected to exceed threshold values, e.g. European or US limits (European Parliament and Council, 2010; US EPA, 264 2006); thus, in the case of data gaps, emission limits can be used as proxies. The technology and efficiency of energy

265 recovery is less regulated so a regulation-based proxy is not possible. Energy recovery depends on site-specific

266 configurations, e.g. combined heat and power production or only heat or power production (Astrup et al., 2014).

- 267 Empirical data for the electricity efficiency of 58 incineration facilities in the US of which two produce combined heat
- and power ranged from 7 to 22% (US EPA, 2017). While data are available for waste incineration, it is challenging to
- 269 get site-specific data, which is why we selected electricity recovery as an illustrative data gap.

270 **3.4** Methods to fill the data gaps in the case study

271 We applied approaches to fill the data gaps based on the type of data gap and availability of surrogate values. The data gaps described in the previous section were missing site-specific data, thus, secondary data were identified for use as 272 273 surrogate values. As shown in Table 3, not all approaches matched with all data gaps. Direct and average proxies 274 require single surrogate values, and could be paired with the data gaps in all five SWM processes. Scaled proxies are 275 suitable for compositional data and were therefore only paired with the waste composition data gaps. Min-max intervals 276 were used when multiple surrogate values were available for the same data gap, enabling an interval consisting of the 277 extremes of the multiple surrogate values. A sufficiently large number of sampled data were available for the 278 incineration electricity recovery, which enabled the development of a statistical distribution. Furthermore, arithmetic 279 mean values and standard deviations for the CO₂ emissions from cardboard remanufacturing are available in literature, 280 and were used for uncertainty propagation of the LCA results (see documentation in the SI).

281 **Table 3** Approaches used to fill the data gaps in the SWM scenario

Solid waste management process	Approach to fill data gaps			
Waste generation	Direct, average, and scaled proxy			
Cardboard remanufacturing; Incineration	Direct and average proxy, min-max interval, statistical distribution			
Material recovery facility; Composting	Direct and average proxy, min-max interval			

282

Table 4 shows the applied surrogate data. For 'leaving gap', we inserted a zero value in all processes except for the MRF process, to illustrate the importance of filling a data gap even though it might not be realistic to use a value of zero in an actual study. In the MRF process, we interpreted leaving the gap as setting the material recovery efficiency to 100%. This interpretation is based on the fact that any input varying from 1 to 0 can be reversed to make 0 or 1 the

appropriate leave-gap value (e.g., if MRF recovery rates are 100%, then residual rates are 0%).

288	The waste composition data gaps regarded the constituents, i.e. the sub-fractions of the waste fractions (e.g., meat and				
289	non-meat were constituents of food waste). For waste fractions with two constituents (i.e. food: non-meat and meat;				
290	aluminum: cans and others; ferrous metal: cans and others) or three constituents (i.e. paper: mixed paper, office paper,				
291	and newsprint; glass: brown, green, and clear glass),				
292	a. Leave $gap = one constituent$ set to 100%, the other(s) set to 0%				
293	b. Direct proxy ('default') = Proportional split based on US county data (SCS Engineers, 2014)				
294	c. Direct proxy ('equal') = equally split between all constituents.				
295	For waste fractions with five constituents (plastic: LDPE, HDPE, PET, mixed, and non-recyclable plastic),				
296	a. Leave $gap = one$ constituent set to 100%, the others set to 0%				
297	b. Direct proxy ('default') = Proportional split based on US county data (SCS Engineers, 2014)				
298	c. Direct proxy ('equal') = equally split between all constituents				
299	d. Scaled proxy = unknown proportion of two constituents; known proportion of the other three constituents				
300	scaled to 100%.				
301	The applied method for 'leave gap' reflected that the functional unit of 1000 kg was maintained to obtain equivalent				
302	functionalities. Furthermore, it reflected that we chose the data gaps at the constituent level, e.g. the unknown				
303	proportion of meat and non-meat. Hence, 'leave gap' was interpreted as setting a single constituent to 100% and the				
304	others set to 0%, and the 'scaling method' was interpreted as extending the proportions of the known constituents to all				
305	the constituents of a fraction, e.g. the known proportions of PET, Mixed, and Non-recyclable plastic were extended to				
306	the entire plastic fraction. Accuracy of the fractional waste composition is important as it influences the physio-				
307	chemical composition of the waste (e.g., more plastic enhances the fossil carbon content).				

308 Table 4 Applied surrogate values to fill the selected data gaps. Bold values are inserted as default values when the data gap is not analyzed. The square brackets in the

SWM process	Data gap	Applied surrogate data values			
Material recovery		Direct proxies	Average proxies	Min-max intervals	Statistical distr.
facility (MRF)	Diesel use [kg/kg waste]	0.0007 ^{a,b} , 0.0008 ^b	0.0008 ^b	0.0007-0.001°	-
	Electricity use [kWh/kg input]	0.0047/0.0060/ 0.0062 /0.0078 ^a , 0.0115/0.0138 ^b	$0.0088^{a,b}, 0.020^{c}, 0.016^{d}, 0.01735^{a,b,c,d}$	0.009-0.03 ^c , 0.012-0.02 ^d , 0.0047-0.03 ^{a,b,c,d}	-
	Material recovery rates:				-
	• Cardboard [% input]	• Cardboard: 54 /76/99 ^a	• Cardboard: 77/88 ^a	• Cardboard: 54-99 ^a	
	• Paper [% input]	• Paper: 39/ 77 /98 ^a	• Paper: 69 ^a	• Paper: 39-98 ^a	
	• Al [% input]	• Al: 87/96/ 97 ^a	• Al: 92 ^a	• Al: 87-100 ^a	
	• Fe [% input]	• Fe: 88/97/ 98 ^a	• Fe: 93 ^a	• Fe: 87-100 ^a	
	• Glass [% input]	• Glass: 69/93/ 95 ^a	• Glass: 81 ^a	• Glass: 69-95 ^a	
	• HDPE-PET [% input]	• HDPE-PET: 83/97/ 98 ^a	• HDPE-PET: 90 ^a	• HDPE-PET: 83-100 ^a	
	• LDPE [% input]	• LDPE: 77/ 98 ^a	• LDPE: N.A.	• LDPE: 77-98 ^a	
Cardboard	Electricity use [kWh/kg input]	0.462/ 0.589 /0.620/1.258 ^e	0.73 ^e	0.589-1.258 ^e	-
remanufacturing	Natural gas use [kg/kg input]	0.0243 /0.080/0.107/0.175 ^e	0.097 ^e	0.0243-0.175 ^e	-
	Direct CO ₂ emissions [kg/kg input]	1.04 /1.06/1.4 ^e	1.2 ^e , 0.73/0.82/0.89 ^f	0.61-1.4 ^{e,f} , 0.38-1.56 ^f ,	Norm (0.82, 0.31 ²) ^f
				0.22-1.86 ^f , 0.31-1.26 ^f	Norm (0.73, 0.41 ²) ^f
					Norm (0.89, 0.60 ²) ^f
Composting	Diesel use [kg/kg input]	0.00265 ^g	0.003 ^h	0.0004-0.006 ^h	-
	CH ₄ emission [% kg C transformed]	1.7 ^g	1.7/2.7/7.2/9.3 ^h	2.4-3, 5.1-13.5, 0.8-2.5 ^h	-
	N ₂ O emission [% kg N transformed]	0.4 ^g , 1.8 ^h	0.4 ^h	0.1-0.7 ^h	-
Incineration	Electricity recovery [% input energy]	18.2/ 20.9/ 24.4 ⁱ	17.2 ^j	7.7-22.2 ^j , 18.2-24.4 ⁱ	Norm (17.2, 2.70 ²) ^j

table note contain the type of technologies and data in the literature references. N.A.=not available in literature

a: Pressley et al. (2015) [single, dual, presorted, and mixed MRF], **b**: Fitzgerald et al. (2012) [presorted MRF], **c**: Merrild et al. (2009) [paper MRF], **d**:Larsen et al.

311 (2009) [glass MRF], e: RTI International (2003) [newsprint, magazines, office paper, and corrugated cardboard], f: Brogaard et al. (2014) [copy paper, newsprint,

312 corrugated cardboard, and cardboard], g: Levis and Barlaz (2013) [SWOLF model default values], h: Boldrin et al. (2009) [values from literature review], i: Hodge et

al. (2016) [US estimates], **j**: US EPA (2017) [US incineration facility data], **k**: De la Cruz and Barlaz (2010) [default, low, medium, and high decay rates], **l**: IPCC

314 (2006) [decay rates for temperate, dry, and wet climate]

315

316

317 4 Results and Discussion

4.1 Completeness scores and missing flows for the incineration process

- 319 Table 5 shows the completeness scores calculated for the waste incineration process, which corresponds to step 3 in
- 320 Figure 1. A single completeness score was obtained for the intermediate flows, while a score range was obtained for the
- 321 elementary flows (with and without expert judgement) as well as an example of a weighted score for the GWP metric.

322 **Table 5** Completeness scores for the waste incineration process in the SWM scenario

Reference of expected flows	Completeness score [%] (included no. flows / expected no. flows)			
Intermediate flows completeness				
Danish register	78 (7/9)			
Elementary flows completeness				
Sum without experts addition	51 (19/37)			
Sum with experts addition	38 (19/50)			
Example of weighted score				
GWP metric	94%			

³²³

324 The completeness score for the intermediate flows was 78%, as 7 out of 9 flows were included. From Table 2, the 325 missing intermediate flows were auxiliary fuel and precipitation chemicals which are common material inputs at 326 incineration plants. Secondary fuel is typically oil with natural gas becoming more common and precipitation chemicals 327 are e.g. TMT15, FeCl₃, and FeSO₄. Generic secondary data are available for the production of both fossil oil and FeCl₃. 328 Thus, they are site-specific data gaps for which process data exist; flow amounts can be estimated by expert judgement. 329 The completeness score for the total elementary flows were between 38% and 51% depending on whether expert 330 judgement was applied and how many flows were added by the expert. In the case of the score of 38%, all the possible 331 additional flows by the expert were added (see flows in Table 2). The relatively low score range suggests that the included air emissions do not sufficiently represent the true number of air emissions from waste incineration plants. 332 333 However, the unweighted completeness score does not say anything about the amounts emitted nor the importance or 334 relevance of the flows. Relevance of the flows can relate to the LCIA methods being investigated in an LCA study. For example, in a carbon footprint study only greenhouse gases are relevant. The main greenhouse gas emissions from 335 336 waste incineration are CO₂, CH₄, and N₂O (Astrup et al., 2009). As CO₂ and CH₄ are included in the LCI of the waste incineration process, but N_2O is not; this gives a completeness score of 67%. Furthermore, the importance can be 337 338 considered by adding weighting factors to the equation. Weighting factors for individual flows may be based on the 339 contribution of emissions to global warming within the geographical area of interest. For example, in the US, 97.4% of

340 the contribution to global warming stems from the emission of $CO_2(81.6\%)$, $CH_4(10.1\%)$, and $N_2O(5.7\%)$ given as 341 percentages based on CO₂ eq (US EPA, 2018). Using these numbers as weighting factors gives a completeness score of 94% (i.e. $\frac{0.816 \times 1 \text{ CO}_2 + 0.101 \times 1 \text{ CH}_4}{0.816 \times 1 \text{ CO}_2 + 0.101 \times 1 \text{ CH}_4 + 0.057 \times 1 \text{ N}_20}$), shown in Table 5. This notably higher weighted completeness score seems 342 343 appropriate for the waste incineration process, because - despite N₂O being a very potent greenhouse gas - the emitted amounts are significantly smaller compared to CO₂ (Astrup et al., 2009; US EPA, 2017), and thus the impact 344 345 contribution from CO_2 is expected to be larger. We therefore suggest the addition of weighting factors to obtain a more 346 valid completeness score, which is in alignment with the guidelines suggested by the US EPA (Edelen and Ingwersen, 2016). 347

348 **4.1.1** Is the subjectivity of the completeness metric a problem?

349 In the context of data quality assessment, the process completeness scores are usually converted to a quality rating 1-5 based on defined criteria, e.g. 1 (≥95%), 2 (85-95%), 3 (75-85%), 4 (50-75%), and 5 (<50%) (JRC, 2010). Applying 350 these criteria to the completeness scores in Table 5 give ratings of 3 (intermediate flows) and 2, 4 or 5 (elementary 351 352 flows). The completeness metric is influenced by the subjective judgement of the user, which is why we suggest a more 353 systematic method to estimate the set of expected flows (Figure 1). Furthermore, transparency in the choice of 354 weighting factors is crucial, e.g. by documentation of references as in the above example with greenhouse gas 355 contributions in the US. Due to the subjectivity of the completeness metric, it can be argued that a qualitative 356 assessment of process completeness is just as useful, e.g. scoring completeness as Good, Average, or Poor based on 357 whether the missing flows were from a priority list or not (van den Berg et al., 1999). However, authors have argued 358 that data quality criteria do not need to be objective but must be inter-subjective, meaning that multiple individuals 359 should be able to come up with similar data quality scores based on the same criteria (Funtowicz and Ravetz, 1990). As the data quality criteria for process completeness equals the set of expected flows, this highlights the importance of 360 361 validation and transparency of the identification of the expected flows.

362 4.2 Carbon footprint results

This section corresponds to the part of step 4 in Figure 1 that evaluates whether the data gaps matter, i.e. if they will affect the final GWP results if not filled by surrogate data. We have evaluated this by comparing the GWP when applying a surrogate value with the GWP when leaving the gap for each of the 16 parameters in Table 4; if the difference between the GWP is larger than 5% then the data gap matters (i.e. is considered important). Section 4.3 then targets the part of step 4 in Figure 1, where the final surrogate values are selected for the data gaps identified as being important. Figures 3 and 4 illustrate the GWP as a function of the data gap and surrogate value/method. In addition, Figure 5

370 illustrates the summary data of the results and illustrates the variability of the applied surrogate data from literature. The

applied surrogate data are shown in Table 4, however, it is not possible to identify each specific value in the plots,

which are meant to show the overall trends in the results. It should be mentioned that the GWP results from use of the

373 statistical data are not shown in the plots, instead they are described in the text below for comparison.



Fig. 3 Global warming potential (per functional unit) from filling or leaving the data gaps for the 16 parameters in Table
4. The results are shown at scenario level. The symbol shape denotes the method (i.e. type of surrogate data) and the color
denotes the parameter. Horizontal placement of the values is irrelevant

378



Fig. 4 Global warming potential results (per functional unit) from filling or leaving the data gaps for the 16 parameters in Table 4. The results are shown at process level. The symbol shape denotes the method (i.e. type of surrogate data) and the color denotes the parameter. Horizontal placement of the values is irrelevant. Note that the y scale of the plots differs

383 Figure 3 illustrates the results at the scenario level, i.e. the net scenario results. Regarding the composting process, only 384 the data gap representing the CH_4 emission factor deviated >5% to leaving the gap, when applying maximum and 385 average proxy values from Boldrin et al. (2009). Thus, the potential anaerobic CH₄ generation, occurring locally in the 386 piled waste, should be modeled accurately. Regarding the incineration process, there is clearly >5% difference between leaving the gap and applying proxy values, thus, electricity efficiency is an important parameter that should also be 387 388 modeled accurately. The outlier in Figure 5 (scenario level) were associated with the minimum value for electricity 389 recovery from the US EPA (2017). To compare, fitting a normal distribution to the empirical energy recovery data for 390 the US plants, resulted in -151 ± 45 kg CO₂ eq, which corresponds to the average proxy result in Figure 3 with a standard 391 deviation of 45 kg CO₂ eq around it. Regarding the MRF process, the GWP from using the material recovery proxies 392 for cardboard, paper, and ferrous metal from Pressley et al. (2015) differed >5% to leaving the gaps. This is due to the 393 linked nature of the processes and data, as differences are only seen for the net scenario (Figure 3) where the values 394 control how much material can be finally recycled, but the material recovery efficiencies are insignificant when looking 395 at the MRF process independently (Figure 4.) Another outcome is that, for cardboard, there were less environmental 396 savings with 100% recovery (leaving the gap), because the savings from the subsequent cardboard recycling (-0.09 kg 397 CO_2 eq per kg) were lower than the savings from incineration of the non-recovered cardboard (-1.1 kg CO_2 eq per kg). 398 However, this is not a general conclusion but is based on the data for secondary and primary material production used in 399 this study (RTI International, 2003) as well as the electricity efficiency and marginal electricity fuel (natural gas) 400 assumed for incineration in the US. In the cardboard remanufacturing process, the difference between using surrogate values and leaving the gap was >5% for the electricity use and CO₂ emissions, with proxies from RTI International 401 402 (2003) and Brogaard et al. (2014); using these proxies caused minor outliers (Figure 5). The natural gas data gap gave 403 differences <5%, thus, an easily available surrogate value is sufficient for this parameter. To compare, the use of 404 statistical data for the CO₂ emissions from three different fiber proxies (Brogaard et al., 2014) resulted in -216±7, -405 219 ± 9 , and -215 ± 13 kg CO₂ eq , which are close to the GWP results in Figure 3.

406 The identified important data gaps in the waste composition (>5% difference) were related to the plastic, metal

407 (aluminum), and paper fraction in the waste, which were also those proxies causing outliers (Figure 5). For plastic, first,

- 408 leaving the gap (i.e. assuming 100% LDPE) reduced the environmental benefits from the scenarios because the GWP
- 409 savings per kg recycled material for PET (-2.4 kg CO₂ eq/kg) and HDPE (-1.8 kg CO₂ eq/kg) were larger than for
- 410 LDPE (-0.2 kg CO₂ eq/kg). Second, the equal split between the plastic constituents decreased the environmental savings
- 411 for the same reason, but to a smaller extent. Third, using the scaled proxy (i.e. assuming unknown proportion of HDPE

and LDPE) led to larger environmental savings than the default settings, as more PET was recovered giving larger savings from recycling. For aluminum, the equal split between the constituents differed >5% to leaving the gap (i.e. assuming only aluminum cans). Leaving the gap gave a biased results in the form of enhanced environmental savings because aluminum cans had larger recycling benefits than aluminum foil in the model. Leaving the gap for the paper fraction (i.e. assuming only office paper) gave a bias in the form of reduced environmental savings, because the assumed energy content in office paper was lower compared to the other paper fractions (larger amount of mineral fillers), thus, less electricity was generated at the incineration plant.

419 To summarize, the data gaps identified as important in this case study were the CH₄ emission factor in composting;

420 electricity efficiency in incineration; recovery efficiencies at the MRF; electricity use and CO₂ emissions in

421 remanufacturing; and composition of the plastic, metal, and paper fraction in the household waste. Final surrogate

422 values must be selected for these parameters, which is done in section 4.3.





424Fig. 5 Boxplots summarizing the GWP results per functional unit for the SWM processes, at process and scenario level.425The 50^{th} (median), 25^{th} and 75^{th} percentile are shown. Outliers (dots) are the values beyond 75^{th} percentile+1.5×IQR and426below 25^{th} percentile-1.5×IQR. Interquartile range (IQR) = 75^{th} percentile- 25^{th} percentile

427

428 Figure 4 illustrates the GWP at the process level, i.e. how the data gaps influenced the net results for the processes.

429 These outcomes are relevant if the user is modeling a single process thoroughly instead of an entire SWM system. The

- 430 data gaps with >5% difference between using surrogate data and leaving the gap were the CH₄ and N₂O emissions as
- 431 well as diesel use in the composting process (differs to scenario level results), and the electricity efficiency in the

incineration process. The latter was important at process level because our incineration module included the avoided
production of marginal electricity. For the MRF process, there was a shift of the important data gaps from the recovery
efficiencies to the electricity use, because the recovery efficiencies only affect the subsequent remanufacturing and
avoided primary production. Finally, for the cardboard remanufacturing process, all the parameters differed >5%, i.e.
also natural gas in contrary to the scenario level results. This highlights the need for assessing the importance of data
gaps with regard to the entire system, and not just aggregated important data gaps at the process level, when modeling a
SWM scenario.

The threshold of 5% was a type of cut-off rule for contribution to impacts, which affected how many data gaps were considered important to the LCA results. In a comparative LCA study, the threshold level should be decided taking into account the difference between the scenarios, because the threshold should be smaller than the difference between the scenarios' potential impacts. Furthermore, we evaluated the threshold relative to the net impacts, but it could also be relative to the absolute impacts from a unit process.

444 **4.3** Selection of final surrogate values by evaluation of representativeness

445 The final surrogate values were selected among the possible values in Table 4, and were those with the best match of the temporal, geographical, and technological scope of the case study, being 2018-2020 (temporal scope), the US 446 447 (geographical scope), and US average (technological scope). Table 6 shows the final surrogate values. In some cases, the reason for not selecting the other surrogate data in Table 4 was discrepancies with the technological scope of the 448 449 case study. For example, the material recovery efficiencies of the other MRF surrogate data represented mixed waste 450 stream, dual stream, and presorted stream MRFs (Pressley et al., 2015), whereas the case study targeted single stream MRFs. Also, it was assumed that the composting facility was an aerated pile (open technology) receiving food and yard 451 452 waste, while the other surrogate data in Table 4 represented enclosed, home, or pure garden composting (Boldrin et al., 453 2009). For the incineration process and waste composition data, the selection of final surrogate values was a matter of representativeness as well as data reliability. The electricity efficiencies for 'state-of-the-art', 'average' or 'worse' 454 455 incineration plants are based on expert judgement (Hodge et al., 2016), while the selected final surrogate data are based 456 on actual measurements (US EPA, 2017). Similarly, the waste composition from SCS Engineers (2014) are based partly 457 on site-specific data for the Wake County as well as literature data from Riber et al. (2009), which is considered more 458 valid than the direct and scaled proxies that we assumed for the waste composition. Finally, in the selection of the final surrogate values for the cardboard remanufacturing process, we prioritized data that were generated to represent 459 460 cardboard remanufacturing in the USA, despite the relatively outdated data (from 1990-1992).

Process	Parameter	Final surroga	te values	Temporal coverage	Geogr covera	aphical age	Technological coverage
Composting	CH ₄ emission	g/h:0.8-2.5 1.7)	(average	2000-2020	No area /	specific generic	Open technologies
Incineration	Electricity efficiency	j: 17.2±2.7		2016	USA		Average of 58 plants
Material recycling facility	Recovery efficiency	a: 100 (cardboard), 99 (paper), 97 (Al), 98 (Fe), 90 (LDPE), 98 (HDPE, PET), 95 (glass)		2014	USA		Single stream
Waste generation	Paper, plastic, and metal	k: composition	fractional	2008	Wake USA	County,	Household waste
Cardboard remanufacturing	Electricity use	e: 0.589		1990-1992	USA		Secondary cardboard, average of unknown
	CO ₂ emission	e: 1.04		1990-1992	USA		technology differences

461 **Table 2** Final surrogate values selected to fill the important data gaps in the case study

462 a: (Pressley et al., 2015), e: (RTI International, 2003), g: (Levis and Barlaz, 2013), h: (Boldrin et al., 2009), j: (US EPA,

463 2017), k: (Riber et al., 2009)

464 **5** Conclusions

465 In this study, we suggested a method for systematically estimating system and process completeness, identification of 466 missing flows (data gaps), and application of surrogate values to fill the data gaps. The study targeted foreground process data and used a solid waste management (SWM) scenario as case study. The expected input and output flows in 467 a waste incineration model were identified based on legislative documents and expert judgement, after which process 468 469 completeness scores were calculated and missing flows identified. The completeness score for the material inputs to 470 waste incineration was 78%, and the missing flows were auxiliary fuels and precipitation chemicals. The completeness 471 score for air emissions ranged from 38% and 50%, with and without expert judgement. If only greenhouse gases were 472 relevant (CO₂, CH₄, and N₂O), the completeness score increased from 38% or 50% to 67%. Further, when impact-based 473 weighting factors were applied the score increased to 94%, which provided a better reflection of the actual 474 completeness of the model concerning the coverage of the global warming potential. Thus, the evaluation of 475 completeness should consider the relevance and importance of flows, depending on the included life cycle impact 476 methods and the weighting of different flows, respectively. To illustrate the use of approaches and surrogate data to fill data gaps, data gaps were selected for 16 different parameters in five SWM processes. The global warming potential 477 478 was used as the metric to compare the results from the use of surrogate data, and from leaving the gap, to identify the 479 important data gaps that should be filled by representative surrogates. The important data gaps were the CH₄ emission factor in the composting process; electricity efficiency of incineration; recovery efficiencies at the MRF; electricity use 480 481 and CO₂ emissions in remanufacturing; and composition of the plastic, metal, and paper fraction in the household

- 482 waste. For these parameters, leaving the gap would change the LCA results >5%. For the other parameters, even the use
- 483 of min-max surrogates led to relatively small changes in the results, therefore lower quality surrogate values are
- 484 sufficient. It is recommended that practitioners follow the suggested method for a more systematic identification of
- 485 missing flows and application of surrogate values in the life cycle inventories process modeling.

486 **References**

- Astrup, T., Møller, J., Fruergaard, T., 2009. Incineration and co-combustion of waste: accounting of greenhouse gases
 and global warming contributions. Waste Manag. Res. 27, 789–799. doi:10.1177/0734242x09348529
- Astrup, T.F., Tonini, D., Turconi, R., Boldrin, A., 2014. Life cycle assessment of thermal Waste-to-Energy technologies:
 Review and recommendations. Waste Manag. 37, 104–115. doi:10.1016/j.wasman.2014.06.011
- Bisinella, V., Götze, R., Conradsen, K., Damgaard, A., Christensen, T.H., Astrup, T.F., 2017. Importance of waste composition for Life Cycle Assessment of waste management solutions. J. Clean. Prod. 164, 1180–1191. doi:10.1016/j.jclepro.2017.07.013
- Boldrin, A., Andersen, J.K.J.K., Møller, J., Christensen, T.H.T.H., Favoino, E., 2009. Composting and compost utilization: accounting of greenhouse gases and global warming contributions. Waste Manag. Res. 27, 800–812. doi:10.1177/0734242X09345275
- Brogaard, L.K., Damgaard, A., Jensen, M.B., Barlaz, M., Christensen, T.H., 2014. Evaluation of life cycle inventory data
 for recycling systems. Resour. Conserv. Recycl. 87, 30–45. doi:10.1016/j.resconrec.2014.03.011
- 499 Burnley, S.J., 2007. A review of municipal solid waste composition in the United Kingdom. Waste Manag. 1274–1285.
- Canals, L.M.I., Azapagic, A., Doka, G., Jefferies, D., King, H., Mutel, C., Nemecek, T., Roches, A., Sim, S., Stichnothe,
 H., Thoma, G., Williams, A., 2011. Approaches for addressing life cycle assessment data gaps for bio-based
 products. J. Ind. Ecol. 15, 707–725. doi:10.1111/j.1530-9290.2011.00369.x
- Clavreul, J., Baumeister, H., Christensen, T.H., Damgaard, A., 2014. An environmental assessment system for
 environmental technologies. Environ. Model. Softw. 60, 18–30. doi:10.1016/j.envsoft.2014.06.007
- Clavreul, J., Guyonnet, D., Tonini, D., Christensen, T.H., 2013. Stochastic and epistemic uncertainty propagation in LCA.
 Int. J. Life Cycle Assess. 18, 1393–1403. doi:10.1007/s11367-013-0572-6
- Danish EPA, 2015. Bekendtgørelse om et register over udledning og overførsel af forurenende stoffer (PRTR). BEK nr.
 1172 af 13/10/2015 (Gældende).
- 509 Danish EPA, 2010. Bekendtgørelse om visse virksomheders afgivelse af miljøoplysninger (historisk).
- Edelen, A., Ingwersen, W., 2016. Guidance on Data Quality Assessment for Life Cycle Inventory Data. US EPA,
 Cincinnati, Ohio, USA.
- Edjabou, M.E., Bang Jensen, M., Götze, R., Pivnenko, K., Petersen, C., Scheutz, C., Astrup, T.F., 2014. Municipal solid
 waste composition: Sampling methodology, statistical analyses, and case study evaluation. Waste Manag. 36, 12–
 23.
- European Commission, 2006. Vejledning for gennemførelse af det europæiske Pollutant Release and Transfer Register
 (PRTR).
- European Parliament and Council, 2010. DIRECTIVE 2010/75/EU OF THE EUROPEAN PARLIAMENT AND OF
 THE COUNCIL of 24 November 2010 on industrial emissions (integrated pollution prevention and control)
 (Recast). doi:10.3000/17252555.L_2010.334.eng
- 520 Fazio, S., Garraín, D., Mathieux, F., De la Rúa, C., Recchioni, M., Lechón, Y., 2015. Method applied to the background

- analysis of energy data to be considered for the European Reference Life Cycle Database (ELCD). Springerplus 4.
 doi:10.1186/s40064-015-0914-x
- Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and quality in science for policy, in: Uncertainty and Quality in Science
 for Policy. Kluwer Academic Publishers, p. 229.
- Garraín, D., Fazio, S., de la Rúa, C., Recchioni, M., Lechón, Y., Mathieux, F., 2015. Background qualitative analysis of
 the European reference life cycle database (ELCD) energy datasets part II: electricity datasets. Springerplus 4,
 doi:10.1186/s40064-015-0812-2
- Götze, R., Boldrin, A., Scheutz, C., Astrup, T.F., 2016. Physico-chemical characterisation of material fractions in household waste: Overview of data in literature. Waste Manag. 49, 3–14. doi:10.1016/j.wasman.2016.01.008
- Hauschild, M.Z., Goedkoop, M., Guinée, J., Heijungs, R., Huijbregts, M., Jolliet, O., Margni, M., Schryver, A., Humbert,
 S., Laurent, A., Sala, S., Pant, R., Guinee, J., Heijungs, R., Huijbregts, M., Jolliet, O., Margni, M., De Schryver, A.,
 Humbert, S., Laurent, A., Sala, S., Pant, R., Guinée, J., Heijungs, R., Huijbregts, M., Jolliet, O., Margni, M.,
 Schryver, A., Humbert, S., Laurent, A., Sala, S., Pant, R., 2012. Identifying best existing practice for
 characterization modeling in life cycle impact assessment. Int. J. Life Cycle Assess. 18, 683–697.
- Henriksen, T., Astrup, T.F., Damgaard, A., 2017. Linking Data Choices and Context Specificity in Life Cycle Assessment
 of Waste Treatment Technologies: A Landfill Case Study. J. Ind. Ecol. 00, 1–11. doi:10.1111/jiec.12709
- Hischier, R., Baitz, M., Bretz, R., Frischknecht, R., Jungbluth, N., Marheineke, T., Mckeown, P., Oele, M., Osset, P.,
 Renner, I., Skone, T., Wessman, H., Beaufort, A.S.H. De, 2001. Guidelines for Consistent Reporting of Exchanges
 from / to Nature within Life Cycle Inventories (LCI). Int. J. Li 6, 192–198. doi:10.1007/BF02979374
- Hodge, K.L., Levis, J.W., DeCarolis, J.F., Barlaz, M.A., 2016. Systematic Evaluation of Industrial, Commercial, and
 Institutional Food Waste Management Strategies in the United States. Environ. Sci. Technol. 50, 8444–8452.
 doi:10.1021/acs.est.6b00893
- Hodson, E.L., Martin, D., Prinn, R.G., 2010. The municipal solid waste landfill as a source of ozone-depleting substances
 in the United States and United Kingdom. Atmos. Chem. Phys. 10.
- Hoornweg, D., Bhada-Tata, P., 2012. Urban development series Knowledge papers Waste composition. World Bank,
 Washington DC, USA.
- Huijbregts, M.A.J., Norris, G., Bretz, R., Ciroth, A., Maurice, B., von Bahr, B., Weidema, B., Beaufort, A.S.H. De, 2001.
 Framework for Modelling Data Uncertainty in Life Cycle Inventories. Int. J. 6, 127–132. doi:10.1007/BF02978728
- ISO, 2006. ISO 14044 Environmental management Life cycle assessment Requirements and guidelines. Geneva,
 Switzerland.
- JRC, 2010. ILCD Handbook International Reference Life Cycle Data System General guide for Life Cycle Assessment
 Detailed guidance, International Reference Life Cycle Data System (ILCD) Handbook. Ispra, Italy.
 doi:10.2788/38479
- Laner, D., Rechberger, H., Astrup, T., 2015. Applying Fuzzy and Probabilistic Uncertainty Concepts to the Material Flow
 Analysis of Palladium in Austria. J. Ind. Ecol. 19, n/a-n/a. doi:10.1111/jiec.12235
- Little, R.J., D'Aostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W.,
 Molenberghs, G., Murphy, S.A., Neaton, J.D., Rotnitzky, A., Scharfstein, D., Shih, W.J., Siegel, J.P., Stern, H.,
 2012. The prevention and treatment of missing data in clinical trials. N. Engl. J. Med. 1355–1360.
- Majeau-Bettez, G., Strømman, A.H., Hertwich, E.G., 2011. Evaluation of process- and Input-Output-based Life Cycle
 Inventory Databases with Regards to Truncation and Aggregation Issues. Environ. Sci. Technol. 45, 10170–10177.
 doi:10.1021/es201308x
- Moberg, A., Borggren, C., Ambell, C., Finnveden, G., Guldbrandsson, F., Bondesson, A., Malmodin, J., Bergmark, P.,
 2014. Simplifying a life cycle assessment of a mobile phone. Int. J. Life Cycle Assess. 19, 979–993. doi:DOI
 10.1007/s11367-014-0721-6

- Moreau, V., Bage, G., Marcotte, D., Samson, R., 2012. Statistical estimation of missing data in life cycle inventory: An
 application to hydroelectric power plants. J. Clean. Prod. 37, 335–341. doi:10.1016/j.jclepro.2012.07.036
- Pressley, P.N., Levis, J.W., Damgaard, A., Barlaz, M.A., DeCarolis, J.F., 2015. Analysis of material recovery facilities
 for use in life-cycle assessment. Waste Manag. 35, 307–317. doi:10.1016/j.wasman.2014.09.012
- Riber, C., Petersen, C., Christensen, T.H., 2009. Chemical composition of material fractions in Danish household waste.
 Waste Manag. 29, 1251–1257. doi:10.1016/j.wasman.2008.09.013
- RTI International, 2003. Life-Cycle Inventory Data Sets for Material Production of Aluminium, Glass, Paper, Plastic, and
 Steel in North America. Research Triangle Park, North Carolina, USA.
- SCS Engineers, 2014. Wake County, North Carolina. Waste Characterization Study: Summary of Results. Wake County,
 North Carolina, USA.
- Steinmann, Z.J.N., Venkatesh, A., Hauck, M., Schipper, A.M., Karuppiah, R., Laurenzi, I.J., Huijbregts, M.A.J., 2014.
 How to address data gaps in life cycle inventories: A case study on estimating CO2 emissions from coal-fired electricity plants on a global scale. Environ. Sci. Technol. 48, 5282–5289. doi:10.1021/es500757p
- Subramanian, V., Golden, J.S., 2016. Patching Life Cycle Inventory (LCI) data gaps through expert elicitation: Case
 study of laundry detergents. J. Clean. Prod. 115, 354–361. doi:10.1016/j.jclepro.2015.11.098
- 580 US EPA, 2018. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2016. US EPA, USA.
- US EPA, 2017. eGRID2014 Data File v2 (Version 2.0) [WWW Document]. URL https://www.epa.gov/energy/emissions generation-resource-integrated-database-egrid (accessed 5.21.18).
- US EPA, 2016. Advancing Sustainable Materials Management: 2014 Fact Sheet Assessing Trends in Material
 Generation, Recycling, Composting, Combustion with Energy Recovery and Landfilling in the United States. US
 EPA, USA.
- US EPA, 2006. Part II 40 CFR Part 60 Standards of Performance for New Stationary Sources and Emission Guidelines
 for Existing Sources: Large Municipal Waste Combustors; Final Rule.
- van den Berg, N.W., Huppes, G., Lindeijer, E., Ven, B.L. Van Der, Wrisberg, N.M., 1999. Quality Assessment for LCA.
- Weidema, B.P., Bauer, C., Hischier, R., Mutel, C., Nemecek, T., Reinhard, J., Vadenbo, C.O., Wenet, G., 2013. Overview
 and methodology. Data quality guideline for the ecoinvent database version 3. St. Gallen, Switzerland.
- Yoshida, H., Clavreul, J., Scheutz, C., Christensen, T.H., 2014. Influence of data collection schemes on the Life Cycle
 Assessment of a municipal wastewater treatment plant. Water Res. 56, 292–303. doi:10.1016/j.watres.2014.03.014
- Yoshida, H., Nielsen, M.P., Scheutz, C., Jensen, L.S., Bruun, S., Christensen, T.H., 2016. Long-Term Emission Factors
 for Land Application of Treated Organic Municipal Waste. Environ. Model. Assess. 21, 111–124.
 doi:10.1007/s10666-015-9471-5
- Zhou, X., Zhou, C., Liu, D., Ding, X., 2014. Applied Missing Data Analysis in the Health Sciences: Chapter 1. John
 Wiley and Sons Inc., Hoboken, New Jersey.