



Contents lists available at ScienceDirect

Resources, Environment and Sustainability

journal homepage: www.elsevier.com/locate/resenv

Perspective

A pragmatic and industry-oriented framework for data quality assessment of environmental footprint tools

Ramy Salemdeeb ^{a,*}, Ruth Saint ^b, William Clark ^a, Michael Lenaghan ^a, Kimberley Pratt ^a, Fraser Millar ^a^a Zero Waste Scotland, Forthside Way, Stirling FK8 1QZ, United Kingdom^b Edinburgh Napier University, 10 Bankhead Terrace, Edinburgh EH11 4DY, United Kingdom

ARTICLE INFO

Keywords:

Environmental policy
Data quality
Environmental modelling
Life Cycle Assessment
Life cycle sustainability assessment

ABSTRACT

Environmental footprinting, underpinned by systematic life-cycle thinking, is increasingly seen as a critical concept for designing policies to fight climate change. The holistic nature of a life-cycle approach, built using the principles of Life Cycle Assessment (LCA), enables policy makers to understand the potential impacts and benefits of policy options. Although LCA is a widely used and well-established method, methodological aspects such as the quality of background data, model uncertainty, and comparison against existing literature are not usually communicated effectively to wider audiences, in particular policy makers. This paper introduces a novel hybrid data quality assessment method in the context of a case study based on the Scottish Waste Environmental Footprint Tool, a newly developed environmental life-cycle thinking tool. It offers an accessible method to present results of the data quality assessment of environmental models to policy makers and helps identify areas of improvements in future upgrades.

1. Introduction

Climate change is universally recognised as a global issue with increasing impacts expected due to rapid population growth and urbanisation (United Nations, 2018; Sarkodie et al., 2020). The global climate crisis has prompted governments to develop strategic carbon reduction policies to tackle the devastating consequences of climate change. To monitor the effectiveness of new policies, several environmental modelling tools have been created by governmental and international organisations such as Scotland's Carbon Metric (Zero Waste Scotland, 2020) and US EPA's Waste Reduction Model (US Environmental Protection Agency, 2019). Built using the principles of life-cycle thinking, these data-intensive tools are designed to provide policy makers, who do not necessarily have technical competence in Life Cycle Assessment (LCA), with insights to evaluate and design environmental policies.

A review of several environmental modelling tools built by non-academic organisations reveals the majority of developers behind these tools do not publish a comprehensive and easy-to-understand data quality assessment (Bicalho et al., 2017; Henriksen et al., 2020). However, it is important to inform prospective users of these tools about the quality of background data and consistency with other similar tools. There are a number of methods available to assess the quality of these data and assumptions (Henriksen et al., 2020; Woodall et al., 2013), though they are often targeted at stakeholders with technical LCA knowledge and

expertise (Batini et al., 2009; May and Brennan, 2003). Therefore, there is a need to develop a method to effectively communicate the quality of data used in environmental tools to wider audiences, in particular policy makers.

In addition to supporting policy makers in understanding uncertainty associated with these tools, data quality assessment helps developers to identify priority components of the model that need further development. The developers of environmental modelling tools tend to design and deploy policy tools as efficiently as possible by first developing a Minimal Viable Product, a version with just enough features to meet essential requirements and provide feedback for future product development (Ang et al., 2020). Establishing robust understanding of the quality of data will enable developers to design a model-upgrade strategy that targets priority areas, while considering available resources and time constraints.

This introductory overview paper presents a hybrid approach to data quality assessment designed to: (1) provide an overview of the quality of data used in the development of assessment tools, and (2) identify areas for improvement in their future upgrades. The approach integrates a widely-used, semi-quantitative approach that uses a pedigree matrix, into a qualitative analysis, and presents the results in an easy to use and understand ranking system that can be communicated to a broad audience. To demonstrate the benefits of this hybrid Data Quality Assessment Method (DQAM), a case study based on the Scottish

* Corresponding author.

E-mail address: ramy.salemdeeb@zerowastescotland.org.uk (R. Salemdeeb).<https://doi.org/10.1016/j.resenv.2021.100019>

Received 10 December 2020; Received in revised form 5 March 2021; Accepted 5 March 2021

Available online 8 March 2021

2666-9161/© 2021 The Authors. Published by Elsevier B.V. on behalf of Lishui Institute of Ecology and Environment, Nanjing University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Waste Environmental Footprint Tool (SWEFT) is presented. SWEFT is based on life-cycle thinking principles and was built to quantify the environmental impacts of waste generated and managed in Scotland. The overarching purpose of SWEFT is to provide policy makers and industry actors with insights into the actual environmental impacts of waste.

2. Method

2.1. Data quality assessment method

The proposed assessment method scores data through two stages: (i) data robustness and evidence, and (ii) the level of confidence and consistency with existing literature. This two-stage evaluation is important as it allows the accuracy and reliability of the data to be quantified alongside a qualification of the level of confidence in a given dataset. For example, if data receives a high score in Stage (i) but a low score in Stage (ii), the validity of existing literature can be questioned. The following sections describe these two stages in general terms and how results are presented. Section 3 present a case study where the data quality assessment method is applied to SWEFT data.

2.1.1. Stage (i): data robustness and evidence

To assess the robustness of the underlying data, an ISO 14040 complaint data quality rating method was employed. This approach is a semi-quantitative assessment of the quality criteria of datasets based on a pedigree matrix of five independent characteristics; reliability, completeness, technological representativeness, geographical representativeness, and time-related representativeness (Weidema et al., 2013). This methodology has been used widely in academic literature and industry publications (European Union, 2016; Passarini et al., 2014; Turner, 2016; Laurent et al., 2014). Moreover, it has been recommended by the European Commission to assess the quality of data in the Product Environmental Footprint and the Organisation Environmental Footprint (Zampori and Pant, 2019a,b).

From the five independent characteristics mentioned above, reliability and completeness indicate the way data is derived and the related level of uncertainty, while the representativeness (i.e. technological, geographical and time-related) characterises the degree to which the processes and products selected are depicting the analysed system. The overall Data Quality Score (DQS) for Stage (i) is calculated using Equation (United Nations, 2018):

$$DQS_{stage(i)} = \frac{R + C + TeR + GeR + TiR}{5} \quad (1)$$

Where R is reliability, C is completeness, TeR is technological representativeness, GeR is geographical representativeness, and TiR is time-related representativeness. The Pedigree matrix used in assessing each characteristic is provided in the Supplementary Material.

2.1.2. Stage (ii): confidence and consistency

The second stage of the DQAM aims to check the consistency of the model factors with existing literature and whether the new factors align with current understanding. This additional assessment stage aims to evaluate the 'level of confidence' in a given dataset, for example as used by WRAP (2019) to assess their modelling work to quantify glass packaging.

This analysis works as follows: the modeller is expected to use the responses and associated scores listed in Table 1 to answer the questions below. The overall DQS for the confidence and consistency stage can then be calculated based on the average overall score, as shown in Eq. (2). To be consistent with the scoring scale in Stage (i) and avoid confusion, 5 points are split between available responses, thus, the DQS of Stage (ii) will range from 1 (the highest level of confidence) to 5 (the lowest level of confidence).

Question 1: Do one or more data sources confirm the findings (within +/- 10%)?

Table 1

Rating scores for the confidence and consistency questions.

Response	Score
Yes	1 (highest quality)
Yes, with some reservations	2.5
More yes than no, but equivocal	4
No	5 (lowest quality)

Question 2: Do the key stakeholders/experts actively agree with the findings?

Question 3: Has feedback from the key stakeholders/experts been incorporated?

$$DQS_{stage(ii)} = \frac{Q_1 + Q_2 + Q_3}{3} \quad (2)$$

The weighting score shown in Table 1 is applicable to SWEFT, but can be modified if necessary, based on the scope and objective of the model in question. The method used to interpret these results alongside Stage (i) results is explained in the following section and demonstrated using SWEFT as a case study in Section 3.

2.2. DQAM results presentation and interpretation

Using the hybrid approach, the two stages presented above are combined to provide a simple visual presentation of the results which enables easy interpretation. The scoring scale for the DQS has three quality levels that can be achieved for both Stage (i) and (ii): Excellent = $DQS \leq 2$; Good = $2 < DQS \leq 4$; and Poor = $4 < DQS \leq 5$. The tabulated results are colour coded using a RAG (red–amber–green) system, where red represents the poorest data quality and green the best. This DQS is also graphically presented, as shown in Fig. 1, and provides a more detailed interpretation of the results. The horizontal axis refers to DQSs derived from the data robustness and evidence (Stage i), while the vertical axis shows the DQSs associated with the level of confidence and consistency (Stage ii). The results in this graph follow the same RAG system. An objective of this DQAM is to identify areas of improvement so any data that fall in the 4th quarter of Fig. 1 should be targeted first, with the goal of shifting to a higher quality score, ideally the 1st quarter, as indicated by the grey arrows.

3. Results and discussion: SWEFT case study

SWEFT reviews the whole life-cycle environmental impacts of waste generated in Scotland (Eurostat, 2010), from resource extraction and manufacturing emissions (i.e. embodied impacts), right through to waste management emissions, regardless of where in the world these impacts occur. The life-cycle stages of SWEFT are broken down into 'generated' (i.e. embodied impacts) and the different waste management activities: 'recycled', 'incinerated', 'landfilled' and 'other diversion' (SEPA, 2020). Within each of these stages there are underlying data about the composition of each waste category and the life-cycle inventory data that are assigned to each material and activity. The data used in SWEFT comes from robust, internationally renowned datasets and assumptions based on the best available data; the data quality assessment results are not intended for use outside this specific analysis. Assessing this data is essential to make any claim about the validity of SWEFT's results. Based on the DQAM described above, Table 2 provides the DQSs for Stage (I) and (ii) of the SWEFT case study for household waste categories.

The benefit of this assessment is to provide LCA practitioners with insights into areas for improvements in future updates to improve the DQS, and to provide policy makers with an overview of the quality of environmental modelling tools. The results of the DQAM for SWEFT show excellent or good data quality for all waste categories in terms of confidence and consistency with current understanding and existing literature, i.e. Stage (ii). In terms of data robustness and evidence,

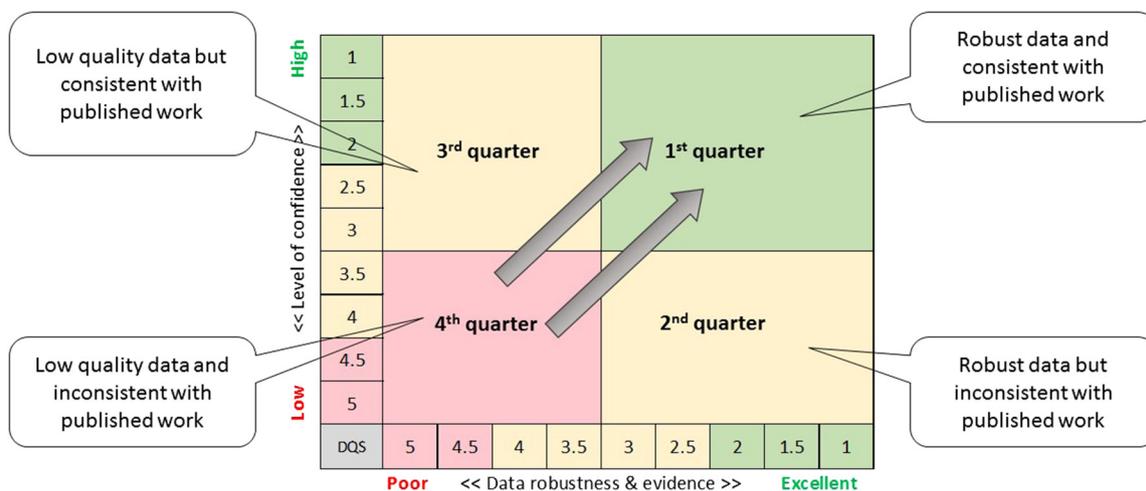


Fig. 1. Data quality assessment visualisation method with the DQS for Stage (i) and (ii) shown along the bottom and left-hand side of the figure, respectively.

i.e. Stage (i), there are areas in need of immediate improvement, and a significant number of waste categories only achieve a ‘good’ DQS. These results highlight the need for this two-pronged approach; although the DQS is high for Stage (ii), the low scores in Stage (i) suggest that the data reported in existing literature is not necessarily robust. Looking deeper into each of the two stages will help to understand where these low scores are coming from.

A more detailed view of these DQSs is provided through the graphical presentation, which can be produced for each of the household waste categories. Using the animal and mixed food wastes (household) category as an example in Fig. 2, the graph can be interpreted as follows: ‘R’, ‘I’ and ‘L’ are located in the green zone which indicates that the overall quality of data used in modelling these stages is robust and consistent with published work: results of the first stage of the data quality assessment show that data used to model the ‘Recycling’, ‘Incineration’ and ‘Landfilled’ activities are robust across all five indicators listed in the pedigree matrix (see Table SM1 in Supplementary Material). Modelled stages achieve a moderate to high score for the level of confidence – the second stage of the data quality assessment – which means that the new calculated emissions factors align well with existing literature and the existing factors in Zero Waste Scotland’s Carbon Metric (Zero Waste Scotland, 2020). Contrastingly, the graph shows that ‘O’ is in the red zone meaning that ‘Other Diversion’ is modelled using datasets that have poorer data robustness. The vertical axis also shows that the new factor for other methods of managing food waste does not align well with current understanding and existing literature (for example, the difference between the new and old factors is more than 10%). As discussed above, this would not necessarily be a negative result as it could suggest current understanding needs to be improved. However, in combination with the low robustness score, this factor has a poor overall DQS and should be developed. The ‘Generated’ factor is located in the amber zone (3rd quarter); the new factor for ‘G’ suggests poor data robustness but high confidence and consistency.

4. Conclusion

This introductory overview paper introduces a pragmatic approach that can be used to present results of the data quality assessment of environmental models to policy makers and help identify areas of improvements in future upgrades. Additionally, it can be used to increase transparency by sharing the results of the assessment method. This will enable benchmarking exercises to take place as well as reduce costs and resources spent by focusing on high-priority areas in need of development. This paper helps to highlight the practical challenges

Waste category: Animal and mixed food waste
 Waste source: Household Waste

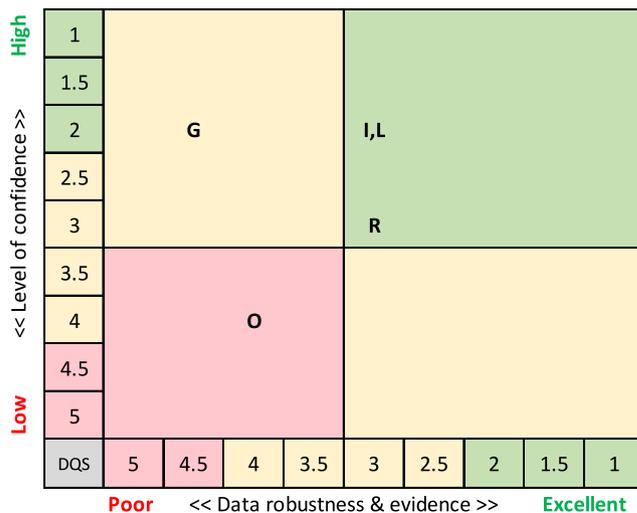


Fig. 2. Data quality assessment results for the Animal and mixed food waste category (Household Waste). G = generated, R = Recycling, I = Incineration, L = Landfilled, and O = Other Diversion.

faced by industry professionals so that academic expertise can help to address them.

As highlighted throughout this paper, the two-pronged approach aims to reduce uncertainty both quantitatively and qualitatively. However, there are limitations; a high DQS might mask a very low score for one of the Stage (i) indicators or Stage (ii) questions, e.g. the data may come from an out-dated source or the data cannot be well corroborated. Thus future updates to this DQAM might require a weighting system as opposed to averaging the scores. Additionally, there is a risk that policy makers might be overly reassured by a high overall DQS and assume that no further improvements are necessary. However, model development is an iterative and evolving process, as is the quality assessment of the underlying data.

To précis, the main objective of writing this introductory overview paper is to introduce a two-tier data assessment approach and invite experts in both academia and industry to work together to refine this method and explore ways to address gaps while maintaining its accessibility to lay audiences, in particular policy makers.

Table 2
Overall data quality scores of household waste by category (Stage i and ii), following the RAG system.

Waste type	Stage (i)	Stage (ii)
Acid, alkaline or saline	3.7	3.0
Animal and mixed food	3.4	3.0
Batteries and accumulators	3.6	3.0
Chemical	3.4	2.0
Combustion	3.6	1.0
Discarded electronic equipment	3.4	4.0
Discarded vehicles	2.8	2.0
Glass	2.6	3.0
Health care and biological	4.7	4.0
Household and similar	3.8	3.0
Metallic (ferrous)	2.9	3.0
Metallic (mixed)	3.3	4.0
Metallic (non-ferrous)	2.9	3.0
Mineral waste from construction and demolition	3.2	4.0
Mixed materials	4.3	4.0
Paper and cardboard	2.9	2.0
Plastic	2.9	2.0
Rubber	3.3	2.0
Soils	2.9	2.0
Sorting residues	4.0	3.0
Spent solvents	3.7	3.0
Textile	3.0	2.0
Used oils	4.0	4.0
Vegetal	2.6	1.0
Wood	2.7	2.0

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This paper is an output of a research project funded by Zero Waste Scotland, United Kingdom to develop the Scottish Waste Environmental Footprint Tool (SWEFT).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.resenv.2021.100019>.

References

- Ang, Y.Q., Berzolla, Z.M., Reinhart, C.F., 2020. From concept to application: A review of use cases in urban building energy modeling. *Appl. Energy* 279, 115738. <https://doi.org/10.1016/j.apenergy.2020.115738>.
- Batini, C., Cappiello, C., Francalanci, C., Maurino, A., 2009. Methodologies for data quality assessment and improvement. *ACM Comput. Surv.* 41, <https://doi.org/10.1145/1541880.1541883>.
- Bicalho, T., Sauer, I., Rambaud, A., Altukhova, Y., 2017. LCA data quality: A management science perspective. *J. Clean. Prod.* 156, 888–898. <https://doi.org/10.1016/j.jclepro.2017.03.229>.
- European Union, 2016. *Methodology for Evaluating LCC*, Brussels.
- Eurostat, 2010. Guidance on classification of waste according to EWC-Stat categories. <https://ec.europa.eu/eurostat/documents/342366/351806/Guidance-on-EWCStat-categories-2010.pdf/0e7cd3fc-c05c-47a7-818f-1c2421e55604>.

- Henriksen, T., Astrup, T.F., Damgaard, A., 2020. Data representativeness in LCA: A framework for the systematic assessment of data quality relative to technology characteristics. *J. Ind. Ecol.* 1–16. <https://doi.org/10.1111/jiec.13048>.
- Laurent, A., Christensen, T., Bakas, I., 2014. Review of LCA studies of solid waste management systems - Part II: Methodological guidance for a better practice. *Waste Manag.* 34, 589–606. <https://doi.org/10.1016/j.wasman.2013.12.004>.
- May, J.R., Brennan, D.J., 2003. Application of data quality assessment methods to an LCA of electricity generation. *Int. J. Life Cycle Assess.* 8, 215–225. <https://doi.org/10.1007/BF02978474>.
- Passarini, F., Nicoletti, M., Ciacci, L., Vassura, I., Morselli, L., 2014. Environmental impact assessment of a wte plant after structural upgrade measures. *Waste Manag.* 34, 753–762. <https://doi.org/10.1016/j.wasman.2013.12.022>.
- Sarkodie, S.A., Owusu, P.A., Leirvik, T., 2020. Global effect of urban sprawl, industrialization, trade and economic development on carbon dioxide emissions. *Environ. Res. Lett.* 15, <https://doi.org/10.1088/1748-9326/ab7640>.
- SEPA, 2020. Waste from all sources - summary data 2018. <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>.
- Turner, D., 2016. *Evaluating the Life Cycle Climate Impacts of Solid Waste Management*. University of Southampton.
- United Nations, 2018. World urbanization prospects 2018 revision: Key facts. <https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf>.
- US Environmental Protection Agency, 2019. ICF, Documentation for Greenhouse Gas Emission and Energy Factors Used in the Waste Reduction Model (WARM): Background Chapters. https://www.epa.gov/sites/production/files/2019-06/documents/warm_v15_background.pdf.
- Weidema, B., Vadenbo, C., Bauer, C., Hischier, R., Nemecek, T., 2013. Overview and methodology (final) acknowledgements v3.
- Woodall, P., Borek, A., Parlikad, A.K., 2013. Data quality assessment: The Hybrid Approach. *Inf. Manag.* 50, 369–382. <https://doi.org/10.1016/j.im.2013.05.009>.
- WRAP, 2019. *GlassFlow 2025: A Review of the Quantity of Glass Packaging Being Placed on the Market (POM) and Recycled in 2017 Project*, Banbury, Oxon.
- Zampori, L., Pant, R., 2019a. Suggestions for updating the organisation environmental footprint (OEF) method. <https://doi.org/10.2760/577225>.
- Zampori, L., Pant, R., 2019b. Suggestions for updating the product environmental footprint (PEF) method. <https://doi.org/10.2760/424613>.
- Zero Waste Scotland, 2020. *The Carbon Footprint of Scotland's Waste: Carbon Metric Technical Report (2017 & 2018)*.