Prediction of Life Cycle Inventories for Industrial Waste Treatment Processes from Historical Data using Machine Learning and Physical Models

Sabine Hallamasek,a Doğancan Karan,b Frank Roschangar,c Fabian Stiefel,d Judith Fellner,e Natalie Egreteau,e Moritz Mayer,e Hubert Weisere and Alexei Lapkina,b\*

aDepartment of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, UK.

bCambridge Centre for Advanced Research and Education in Singapore, CARES Ltd. 1 CREATE Way, CREATE Tower #05-05, 138602 Singapore.

cDevelopment Sciences, Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Road, Ridgefield, Connecticut 06877, USA.

dDevelopment Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach, Germany.

eBoehringer Ingelheim RCV GmbH & Co KG, Doktor-Boehringer-Gasse 5-11, A-1121 Vienna, Austria.

aal35@cam.ac.uk

Abstract

The treatment of waste effluents from industrial activity is a mandatory task to protect the ecosphere from pollutants. However, the operation of waste treatment systems itself causes environmental impacts, which must be accounted for in life cycle assessment (LCA) studies of industrial processes. The magnitude of the environmental impacts caused by a particular waste treatment process is dictated by the characteristics (composition, physical properties) of the treated waste, since these affect the quantity of consumed process auxiliaries. In this work we leveraged machine learning algorithms and historical process data of industrial waste treatment systems to identify relationships between waste characteristics and process performance. The obtained correlations are shown to enable the calculation of waste-specific impacts of a given treatment system.

**Keywords**: predictive LCI, machine learning, data-driven regression, waste treatment, green chemistry.

* 1. Introduction

Every year, enormous amounts of waste are generated by industrial activities, with the chemical industries acting as one of the key players. In 2020, the European Union reported the production of 55 million tons of waste for this sector alone (Eurostat, 2023). Since the generated waste streams are often heavily contaminated with organic and in-organic process residues, it is imperative that effluents are treated appropriately. Thereby resources can be reused within the technosphere or returned safely to the ecosphere. However, while the treatment of hazardous wastes serves to mitigate the industry’s environmental footprint, it must be recognized that waste treatment processes themselves consume resources and release emissions, generating an inherent environmental impact which is attributable to the production process that generated the waste.

The impact incurred by the treatment of a specific waste is dependent on the waste category, which determines the type of required treatment processes, as well as on the exact waste composition, since attributes such as pollutant concentration determine the operational response of a given treatment system in terms of consumed auxiliaries and utilities. Consequently, LCAs of industrial waste treatment systems should consider waste-specific life cycle inventories (LCIs) to obtain reliable impact estimations. However, waste-specific process data for waste treatment systems is not easily obtainable, particularly during early-stage process design, and the availability of waste-specific LCI estimation tools is rather low (Köhler et al., 2007; Seyler et al., 2005; Struijs, 2014). In the absence of waste-specific data, waste treatment LCIs are often estimated using generic datasets from commercial databases (Moreno Ruiz et al., 2022), calculated through time-consuming process modelling, or omitted from the study scope, all of which can introduce significant skew into LCA results. In this work we aimed to close this knowledge gap by developing waste-specific predictive LCI tools for a given waste treatment process from historical process datasets through the application of machine learning techniques.

* 1. Methods

A data-driven approach was employed to quantify the relationships between waste characteristics and the responses of the investigated waste treatment system. The data was sourced from industry logs of real-time in/on-line monitoring and control measurements, which provide a transparent account of system operation over extended time periods. Logged characteristics of the waste feed (quantity, physical properties) were identified as system inputs, while records of the system’s resource and utility consumptions were classified as system outputs. Correlating the system outputs to the variation in waste characteristics was treated as a machine learning regression problem, with the system inputs serving as numerical features. The presented method builds on the approach taken by Seyler et al. (2005) and Köhler et al. (2007) for the development of waste-specific LCI prediction tools, who conceptualised that every consumption of a waste treatment process can be attributed to one or more specific waste characteristics. However, while previous tools relied on averaged annual production data and required fundamental hypotheses on input-output relationships, this work benefits from access to more granular datasets and the leveraging of machine learning to more accurately model the true impact of multiple waste features on the system’s outputs.

The hypothesis was conceptualised on the case study of an industrial wastewater neutralisation system employed by Boehringer Ingelheim. For this purpose, process data logged by the system’s control instruments were extracted for the years 2021 and 2022 at 30 second intervals. Data curation and analysis, as well as model screening, training and tuning was implemented in Python. The available dataset was supplemented with artificial data generated by a physical model emulator to explore its benefits in situations of data scarcity.

* 1. Case study results

The investigated waste treatment system consisted of unit processes for mixing, pH neutralization and cooling. Accordingly, the waste characteristics that determine the system’s operational response were established as waste pH, temperature, and volume. As such, measured values of these quantities were curated to make up the feature space of the regression problem. Equally, the process consumptions that constitute the gate-to-gate LCI for the investigated process were identified as the consumption of neutralizing agents (HCl and NaOH), cooling power (in the form of cooling water and chilling power) and electricity, all of which were regarded as the outputs of the desired machine learning models. Data curation was found to be a crucial stage of the project, revealing limitations and challenges of working with real industrial data.

To determine the relationships between the waste characteristics and process consumptions, the importance of each feature on the magnitude of each consumption was identified via multi-feature single-output regression models on the industrial process data. The obtained fits were used to generate waste-specific LCIs for a range of scenarios ranging from extremely acidic to extremely basic pH, and considering various temperatures and process volumes. For comparison, a non-specific “baseline” LCI dataset was calculated as the process consumptions per m3 of wastewater treated across the entire measurement period of two years. Benchmarking the waste-specific LCIs against the non-specific baseline inventory, significant differences in predicted process consumptions were observed. This was especially apparent in LCIs generated for wastes of different pH values, where the multi-feature regression models could differentiate waste-specific acid and base consumptions, and thus allocate different environmental impacts for different wastes processed by the system, while also capturing non-intuitive consumptions such as compensation for overshoots in neutralization, which was not captured to the same granularity in the baseline LCI.

* 1. Conclusions

The symbiotic use of large historical datasets, statistical and physical models yielded a predictive tool that supports the LCA practitioner in the rapid assessment of waste-related emissions. Specifically, the models were able to generate waste-specific LCIs that allow to differentiate between the impacts caused by wastes of different compositions and physical properties. Understanding the influence of waste characteristics on the environmental process impacts is crucial to enable accurate and fair allocation of waste-related process impacts to the waste-generating production processes.

References

Eurostat, 2023. Generation of waste by waste category, hazardousness and NACE Rev. 2 activity. URL https://ec.europa.eu/eurostat/databrowser/view/ENV\_WASGEN\_\_custom\_2246895/ default/table?lang=en (accessed 11.30.23).

Köhler, A., Hellweg, S., Recan, E., Hungerbühler, K., 2007. Input-dependent life-cycle inventory model of industrial wastewater-treatment processes in the chemical sector. Environ. Sci. Technol. 41, 5515–5522. https://doi.org/10.1021/es0617284

Moreno Ruiz, E., FitzGerald, D., Bourgault, G., Vadenbo, C., Ioannidou, D., Symeonidis, A., Sonderegger, T., Müller, J., Dellenbach, D., Valsasina, L., Minas, N., Baumann, D., 2022. Documentation of changes implemented in the ecoinvent database v3.9. Zürich, Switzerland.

Seyler, C., Hofstetter, T.B., Hungerbühler, K., 2005. Life cycle inventory for thermal treatment of waste solvent from chemical industry: A multi-input allocation model. J. Clean. Prod. 13, 1211–1224. https://doi.org/10.1016/j.jclepro.2005.05.009

Struijs, J., 2014. SimpleTreat 4.0 : a model to predict fate and emission of chemicals in wastewater treatment plants. RIVM Rep. 601353005.