CHEMICAL ENGINEERING TRANSACTIONS

VOL. 103, 2023

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it

Guest Editors: Petar S. Varbanov, Panos Seferlis, Yee Van Fan, Athanasios Papadopoulos

ISBN 979-12-81206-02-1; **ISSN** 2283-9216

DOI: 10.3303/CET23103151

Applications of Multivariate Statistics in the Context of Life Cycle Assessment

Bianca Köck*, Martin Elser, Bettina Mihalyi

Institute of Chemical, Environmental and Bioscience Engineering, TU Vienna, Getreidemarkt 9, 1060 Vienna, Austria
bianca.koeck@tuwien.ac.at

LCA is essential for achieving the European Green Deal's targets and supporting sustainable development. As LCA tools become more complex, multivariate statistics can offer valuable solutions. However, their potential benefits in LCA are yet to be fully explored. Therefore, this review evaluates scientific publications combining both, focusing on evaluating the fields of application and potential of multivariate statistics in LCAs. Key findings were the identified use cases, which can be categorised as "Grouping of Products/Systems", "Reduction of Parameters", "Evaluation of Parameters", and "Support for Decision Makers". Among these, the evaluation of parameters was the most commonly used category, with Principal Component Analysis (PCA) being the most frequently used statistical method. Although the publications showed high potential for gaining information, there is a lack of publications combining both topics. They were often used for high Technology Readiness Levels and often based on small datasets, although large data sets are desirable for gaining reliable and robust results. Therefore, future work will need to validate the data requirements and statistical methods recommended per use case.

1. Introduction

Multivariate statistics has emerged as a powerful extension of univariate statistics that enables the investigation of many variables together, leading to a new and improved quality of data evaluation. This approach has found widespread applications in diverse fields, including Chemometrics (Varmuza and Filzmoser, 2009), as well as in process engineering to evaluate the performance of different processes, for example, for residual biomass valorisation (Del Duca et al., 2022) or improvement seeking in the oil and gas industry (Foronda et al., 2023). Similar data complexity and uncertainties can also be observed when calculating LCAs, which could benefit from the use of multivariate statistical methods to provide additional context for the assessment. LCA is a standardised method used to evaluate the environmental impact of products or systems throughout their whole life cycle (ISO, 2006). LCA holds significant prominence as a pivotal tool within the European Union, garnering escalating recognition in recent legislative endeavours, notably exemplified by the Green Deal or the Circular Economy Action Plan (Sala et al., 2021). By integrating LCA into legislative frameworks, the European Union aims to advance its sustainability objectives, foster eco-friendly practices, and promote informed decision-making that aligns with environmental stewardship and the pursuit of a greener and more sustainable future for its constituents and the global community. This increasing adoption of LCA also accentuates the need for further standardisation and innovative approaches to address the complex and multifaceted challenges concerning the data uncertainty and interpretation of results (e.g. analyze multiple environmental and non-environmental impacts). Overcoming these challenges empowers decision-makers to base their policies and strategies on robust data-driven insights, ensuring a more effective and impactful formulation of environmental policies. However, as the results of this review show, only 17 LCAs utilising multivariate statistics between 2005-2023 were found, with four of them having been published in the last three years. To learn from their use cases, this paper aims to explore the potential applications of multivariate statistics in LCA methodology through a literature review of existing applications. This study summarises the current information in publications for the given research question, “What application possibilities do multivariate statistical methods offer LCA?”. Additionally, the statistical method and objective, as well as the software used in the literature, are evaluated.

Paper Received: 12 June 2023; Revised: 18 July 2023; Accepted: 25 July 2023
Please cite this article as: Köck B., Elser M., Mihalyi B., 2023, Applications of Multivariate Statistics in the Context of Life Cycle Assessment, Chemical Engineering Transactions, 103, 901-906 DOI:10.3303/CET23103151

Table 1: Overview of the reviewed studies combining LCA and multivariate methods. Other used statistical methods, such as bivariate and univariate, are also included in the section "Statistical Methods"

	Study	Topic	Statistical Methods	Data Amount
Agriculture & Food Production	Fraterrigo Garofalo et al. (2023)	Production optimization of omega-3 oil from tuna viscera	MLR, PCA	15 experiments
	González-Quintero et al. (2021)	Environmental impact assessment of dual-purpose farms in Colombia	PCA, HCPA, Kruskal-Wallis & Kruskal-Nemenyi	Data from 1,313 farms
	Michos et al. (2012)	LCA of peach orchard farming systems in Greece	PCA, ANOVA, MC, HCA, Mann-Whitney	16 farming systems
	Chen et al. (2015)	LCA of trout farms in France	PCA with non-parametric bootstrap	24 trout farms
	Bava et al. (2014)	Environmental impact analysis of dairy farms	PCA, CA, PC	28 dairy farms
	Mu et al. (2017)	Environmental performance of specialized dairy farms	MLR, CCA, Outlier detection	55 specialized dairy farms
	Grados and Schrevens (2019)	LCA of potato agricultural in the Central Peruvian Andes	CA, LDA, EFA	Data from 58 potato pilot plots
Material & Elements	Smith et al. (2021)	Chemical element sustainability index for piezoelectric materials	PCA, Trend analysis, Monte Carlo Analysis	59 chemicals
	Rydh and Sun (2005)	Simplification of LCAs at an early stage of product design	LRA, PCA (Sartorius AG, 2023)	data of 214 mechanical design materials
Supply Chain	Genovese et al. (2017)	Reducing redundancies and identifying relationships among indicators in LCA	PCA	LCA calculations on five random samples of 1,000 supply chains
	Pozo et al. (2012)	Dimensionality reduction of environmental metrics for multi-objective optimization	PCA, Pareto solutions	2 supply chain cases: power plant and 14 recyclable products
Waste Water Treatment	Flores-Alsina et al. (2010)	Incorporating multiple criteria into a common LCA for wastewater treatment	Normalisation techniques, HCA, PCA, FA, Discriminant Analyses	12 control strategies at plant level and environmental, legal, technical & economic indicators
	Guo et al. (2023)	Relationships between GWP of waste/sludge treatment facilities	PCA, uncertainty analysis	Data from 660 cities
Other	Rowley et al. (2015)	Unsupervised weighting algorithm using PCA for the Choquet integral in two case studies	PCA	135 Australian industry sectors and 8 biosolids management options
	Bersimis and Georgakellos (2013)	Assessing environmental performance of beverage packaging (aluminum, glass, and PET) in the Greek market	PC, PCA, consistency test	16 beverage containers and 5 impact categories
	Gutiérrez et al. (2010)	Relationships of impact categories in two LCAs: wastewater treatment plants and cultivation systems, and processing and consumption of mussels.	PC, PCA, Multidimensionale Scaling	7 impact category results from 13 wastewater treatment plants and 10 impact category results from 5 life stages of mussels.
	Basson and Petrie (2007)	Decision-making under uncertainty in LCA while comparing bed combustion and refurbished existing pulverized fuel boilers	Latin Hypercube sampling, PCA	Three scenarios for two power station reactivation cases, including financial, social, and environmental aspects

2. Materials and methods

The first steps of the literature research were an unstructured search with Google Scholar to evaluate the availability of publications and to define the range of search terms for the structured research, as well as searching via WTI-AG (2022) (no results) and ScienceDirect (Elsevier B.V., 2023) in 2021. An update search for new publications was performed on March 30 th 2023 as well. The inclusion criteria for the review were the language (English and German) and the type of publication (review articles and research articles in scientific journals). Subsequently, the identified literature was systematically categorised based on the specific topics of Life Cycle Assessment (LCA) under investigation, resulting in distinct groups such as agriculture and food production. Within each category, statistical methods employed, the corresponding utilisation cases, the extent of data availability, and the software or programming languages utilised were meticulously recorded and analysed to discern prevailing patterns and trends in the application of multivariate statistics in LCA research.

3. Results

The results summarize 17 scientific articles that employed multivariate statistical methods such as Principal Component Analysis (PCA), Multiple Linear Regression (MLR), Canonical-Correlation Analysis (CCA), Factor Analysis (FA), Exploratory Factor Analysis (EFA), Linear Discriminant Analysis (LDA), Hierarchical Cluster Analysis (HCA), Hierarchical Clustering on Principal Components (HCPC) and other cluster analysis (CA) to analyse the environmental impact of various products and systems. Bivariate statistical tools, like Pearson Correlation (PC), Spearman Rank Correlations (SRC) and univariate tools, including Analysis of Variance (ANOVA), Linear Regression Analysis (LRA), as well as others like Monte Carlo Analysis (MC) were also used. A summary of the scope of the article, as well as the statistical methods used as well as the sample size, is visualised in Table 1.

3.1. Overview of multivariate statistical methods and software used

Most publications were in the area of agriculture. Fifteen of the 17 studies used PCA, and eight other methods were also used, visible in Table 2.

Table 2: Summary of publications by categories and statistical methods used by topic

Topic	Nr. of studies	PCA	CA	HCA	MLR	(E)FA	HCPA	CCA	LDA
Agriculture and Farming	7	5	2	1	2	1	1	1	1
Materials and Elements	2	2	0	0	0	0	0	0	0
Supply Chain	2	2	0	0	0	0	0	0	0
Wastewater Treatment	2	2	0	1	0	1	0	0	0
Multiple and other Categories	4	4	0	0	0	0	0	0	0
Sum	17	15	2	2	2	1	1	1	1

For the statistical part, the software used were Gurobi, GAMS, MATLAB, STATISTICA, Analytica, SAS, Chemometric Agile Tool, S+ and SPSS or it was just mentioned which programming language (and packages) were used like the FactoMineR package. The LCA was modelled in Excel and Simapro. No automation-like interface between the software for data transfer and no trend for the choice of statistic software could be observed.

4. Discussion

The use cases for multivariate statistics could be identified as a grouping of products/systems, the reduction of parameters, the evaluation of parameters, and the support for decision-making. One approach discussed by González-Quintero et al. (2021) and Michos et al. (2012) involves using multivariate statistical techniques like PCA, HCA, and HCPC to group products or systems based on defined parameters. This provides an objective view of similarities between them, leading to new categorisations like organic, integrated, and conventional for other sectors. However, not all products can be easily differentiated through this method, and adaptation of input parameters may be necessary for better accuracy. Future studies are needed to validate these results. Another approach involves reducing parameters by employing multiple regression analysis and CCA (Mu et al., 2017), PCA (Genovese et al., 2017), as well as multidimensional scaling (Gutiérrez et al., 2010). This eliminates redundant information about the system and improves the cost-efficiency of evaluation. However, if multiple parameters are added up without considering redundancies, it can lead to misleading results. It is important to note that the redundant parameters may still contain valuable data that was not selected for analysis initially.
Several studies used multivariate statistics to evaluate parameters and gain insight into their significance compared to each other. Relationships between input values can be revealed, giving information about their interactions and modification effects. Linkages between different domains of objectives (environmental, economic, technical, and legal) can also be created, leading to improved control strategies. However, the locality should

be considered as an additional influencing factor in the decision of input parameters, and validation of the used publication on a larger scale is required to remove this problem (e.g. Basson and Petrie, 2007).
Multivariate statistics can also support decision-makers by providing new categories for evaluation and creating a chemical element sustainability index (Smith et al., 2021) or a new evaluation method (Pozo et al., 2012). This gives stakeholders easier access to necessary information and tools to estimate new systems. Mathematical or statistical solutions allow a less subjective view of the evaluated system, but the risk of appearing perfect and complete should be questioned. Withholding information from decision-makers can consciously or subconsciously influence their actions.

5. Implications of Findings

The findings of this literature review present a valuable foundation for future investigations to expand the scope of this research in various sectors and domains. Products evaluated can vary over different Technology Readiness Levels (TRL). However, most studies focused on higher TRL. Lower TRL appears to be a promising avenue for further exploration due to its growing importance and the lack of current studies in this area. However, like the literature sources examined (Table 1), these also often have limitations in terms of the amount of data. In multivariate analysis, a sufficiently large sample size is desirable to ensure reliable and robust results. A small sample size may affect the accuracy and generalisability of the results.
The current trend in LCA research reveals a dearth of comprehensive testing and comparison of various statistical models, with a particular emphasis on the exclusive application of PCA. Furthermore, there is a high necessity to explore specialised approaches such as Repeated Cross-Validation (Filzmoser et al., 2009), tailored for situations where data is scarce, and other techniques of machine learning, depending on the size of the data sets. As such, it is imperative to undertake comprehensive evaluations of these alternative methods compared to the commonly employed multivariate statistical models, with the aim of providing LCA practitioners with an effective toolkit suited to different use cases.
Moreover, to enhance the accessibility of multivariate statistics to a wider audience of LCA practitioners, the implementation of integration for existing LCA software or interfaces between statistical and LCA software is proposed as a necessary step forward, which could, for example, be done by using the LCA software Brightway (Mutel, 2014), which is written in Python. Also, guidelines on how to include multivariate statistics for the different use cases (grouping of products/systems, reduction or evaluation of parameters and support for decision makers) could lead to standardisation and, therefore, to higher quality and comparability of results in the future. However, as a first step, it should be clarified in which use cases multivariate statistic is beneficial for LCAs, and also an evaluation to compare it to other methods of Artificial Intelligence should be made in different data and use cases.

6. Conclusion

Overall, this review demonstrates the usefulness of statistical multivariate analysis techniques for addressing the challenges of LCA, such as reducing the complexity of data and the identification of non-redundant impact categories and their relationships. The review of multivariate statistics in LCA highlights the benefits and applications of using statistical multivariate analysis techniques in addressing the challenges of LCA. The findings of the literature review revealed several key points:

- Use Cases: Multivariate statistical methods are utilised in LCA for various purposes, including grouping of products/systems, parameter reduction, parameter evaluation, and support for decision making. These approaches provide objective views of similarities between products/systems, reduce redundant information, evaluate parameter significance, and create new evaluation categories or indices. Other possible use cases which should be explored are
    - Uncertainty Analysis: Multivariate statistics can be leveraged to assess uncertainties by analysing the variability of input parameters and their impact on the overall results. This would enhance the robustness of LCA findings.
    - Time Series and Spatial Analysis: Time-dependent LCA data and geographical variations can be explored using multivariate techniques to identify trends, patterns, regional/temporal differences and correlations. This approach can offer insights into the dynamic nature of environmental impacts and guide strategies for continuous improvement.
    - Integration with Artificial Intelligence: Integrating multivariate statistics with machine learning algorithms, such as neural networks and genetic algorithms, can lead to enhanced predictive capabilities in LCA. This integration could facilitate scenario modelling and aid in the identification of optimal solutions for sustainable product design and policy-making.

  - Comparative LCA: Multivariate methods can support the comparison of LCA results across different products or systems, even when they have varying numbers of parameters or dimensions. This will enable fair and comprehensive assessments.
- The implications of the findings emphasise the need for further research in various sectors and domains, particularly exploring lower TRL and addressing the limitations of small sample sizes in multivariate analysis. Comprehensive testing and comparison of statistical models beyond PCA is necessary. Integration of multivariate statistics into existing LCA software and the development of guidelines for different use cases can enhance accessibility and standardisation.

In conclusion, statistical multivariate analysis techniques provide valuable insights, reduce data complexity, identify non-redundant impact categories, and support decision-making processes in LCA. By incorporating these methods, LCA practitioners can improve the quality, comparability, and robustness of LCA results, ultimately facilitating more effective environmental performance assessments and informed decision-making.

References

Algunaibet I. M., Guillén-Gosálbez G., 2019, Life cycle burden-shifting in energy systems designed to minimize greenhouse gas emissions: Novel analytical method and application to the United States. Journal of Cleaner Production, 229, 886–901.

Algunaibet I. M., Pozo, C., Galán-Martín Á., Guillén-Gosálbez G., 2019, Quantifying the cost of leaving the Paris Agreement via the integration of life cycle assessment, energy systems modeling and monetization. Applied Energy, 242, 588–601.

Almeida A. R., Alvarenga R. A. F., Sartor L. B., Simioni F. J., 2019, Mitigating environmental impacts using Life Cycle Assessment in Brazilian companies: A stakeholders' perspective. Journal of Environmental Management, 236, 291–300.

Basson L., Petrie J. G., 2007, An integrated approach for the consideration of uncertainty in decision making supported by Life Cycle Assessment. Environmental Modelling & Software, 22(2), 167–176.

Bava L., Sandrucci A., Zucali M., Guerci M., Tamburini A., 2014, How can farming intensification affect the environmental impact of milk production? Journal of Dairy Science, 97(7), 4579–4593.

Bersimis S., Georgakellos D., 2013, A probabilistic framework for the evaluation of products' environmental performance using life cycle approach and Principal Component Analysis. Journal of Cleaner Production, 42, 103–115.

Chen X., Samson E., Tocqueville A., Aubin J., 2015, Environmental assessment of trout farming in France by life cycle assessment: using bootstrapped principal component analysis to better define system classification. Journal of Cleaner Production, 87(1), 87–95.

Del Duca V., Chirone R., Coppola A., Scala F., Salatino P., 2022, Application of multivariate statistical analysis for pyrolysis process optimization. Chemical Engineering Transactions, 96, 277–282.

Elsevier B.V., 2023, ScienceDirect.com | Science, health and medical journals, full text articles and books. <www.sciencedirect.com>, accessed 02.04.2023.

Filzmoser P., Liebmann B., Varmuza K., 2009, Repeated double cross validation. Journal of Chemometrics, 23(4), 160–171.

Flores-Alsina X., Gallego A., Feijoo G., Rodriguez-Roda I., 2010, Multiple-objective evaluation of wastewater treatment plant control alternatives. Journal of Environmental Management, 91(5), 1193–1201.

Foronda R. M., Fracassio V. M., Santos R. B., Santos B. F., 2023, Statistical Analysis in Database of Offshore Naturally Flowing Wells with Abnormal Events. Chemical Engineering Transactions, 99, 601–606.

Fraterrigo Garofalo S., Cavallini N., Demichelis F., Savorani F., Mancini G., Fino D., Tommasi T., 2023, From tuna viscera to added-value products: A circular approach for fish-waste recovery by green enzymatic hydrolysis. Food and Bioproducts Processing, 137, 155–167.

Genovese A., Morris J., Piccolo C., Koh S. C. L., 2017, Assessing redundancies in environmental performance measures for supply chains. Journal of Cleaner Production, 167, 1290–1302.

González-Quintero R., Kristensen T., Sánchez-Pinzón M. S., Bolívar-Vergara D. M., Chirinda N., Arango J., Pantevez H., Barahona-Rosales R., Knudsen M. T., 2021, Carbon footprint, non-renewable energy and land use of dual-purpose cattle systems in Colombia using a life cycle assessment approach. Livestock Science, 244, 104330.

Grados D., Schrevens E., 2019, Multidimensional analysis of environmental impacts from potato agricultural production in the Peruvian Central Andes. Science of The Total Environment, 663, 927–934.

Guo D., Li B., Yu W., Han J.-C., Zhou Y., Ye Z., Wu X., Young B., Huang Y., 2023, Revisiting China's domestic greenhouse gas emission from wastewater treatment: A quantitative process life-cycle assessment. Science of The Total Environment, 876, 162597.

Gurobi Optimization, n.d., Gurobi Optimizer <ww.gurobi.com/solutions/gurobi-optimizer> accessed 02.04.2023

Gutiérrez E., Lozano S., Moreira M. T., Feijoo G., 2010, Assessing relationships among life-cycle environmental impacts with dimension reduction techniques. Journal of Environmental Management, 91(4), 1002–1011.

Husson F., Josse J., Lê S., Rennes d'Agrocampus, Mazet J., 2008, FactoMineR: Exploratory Multivariate Data Analysis with R. <factominer.free.fr>, accessed 02.04.2023.

ISO, 2006, ISO 14040:2006 - Environmental management — Life cycle assessment — Principles and framework. <www.iso.org/standard/37456.html> , accessed 02.04.2023.

Leardi R., Melzi C., Polotti G., 2023, CAT, Chemometric Agile Tool. <www.gruppochemiometria.it/index.php/software/19-download-the-rbased-chemometric-software>, accessed 02.04.2023.

Li B., Boiarkina I., Young B., Yu W., 2015, Substance flow analysis of phosphorus within New Zealand and comparison with other countries. Science of The Total Environment, 527–528, 483–492.

Melián E. P., Santiago D. E., León E., Reboso J. V., Herrera-Melián J. A., 2023, Treatment of laundry wastewater by different processes: Optimization and life cycle assessment. Journal of Environmental Chemical Engineering, 11(2), 109302.

Michos M. C., Mamolos A. P., Menexes G. C., Tsatsarelis C. A., Tsirakoglou V. M., Kalburtji K. L., 2012, Energy inputs, outputs and greenhouse gas emissions in organic, integrated and conventional peach orchards, Ecological Indicators, 13(1), 22–28.

Mu W., van Middelaar C. E., Bloemhof J. M., Engel B., de Boer I. J. M., 2017, Benchmarking the environmental performance of specialized milk production systems: selection of a set of indicators. Ecological Indicators, 72, 91–98.

Mutel C., 2014, Brightway Software Framework <documentation.brightway.dev/en/legacy/source/lca/static_lca.html>, accessed 02.04.2023.

Pozo C., Ruíz-Femenia R., Caballero J., Guillén-Gosálbez G., Jiménez L., 2012, On the use of Principal Component Analysis for reducing the number of environmental objectives in multi-objective optimization: Application to the design of chemical supply chains. Chemical Engineering Science, 69(1), 146–158.

Rowley H. V., Geschke A., Lenzen M., 2015, A practical approach for estimating weights of interacting criteria from profile sets. Fuzzy Sets and Systems, 272(1), 70–88.

Rydh C.J., Sun M., 2005, Life cycle inventory data for materials grouped according to environmental and material properties. Journal of Cleaner Production, 13(13–14), 1258–1268.

Sala S., Amadei A.M., Beylot A., Ardente F., 2021, The evolution of life cycle assessment in European policies over three decades. International Journal of Life Cycle Assessment, 26(12), 2295–2314.

Sartorius A.G., 2023, SIMCA® - Multivariate Data Analysis Software. <www.sartorius.com/en/products/process-analytical-technology/data-analytics-software/mvda-software/simca>, accessed 02.04.2023.

Smith L., Ibn-Mohammed T., Reaney I.M., Koh S.C.L., 2021, A Chemical Element Sustainability Index. Resources, Conservation and Recycling, 166, 105317.

Ügdüler S., De Somer T., Van Geem K.M., De Wilde J., Roosen M., Deprez B., De Meester S., 2022, Analysis of the kinetics, energy balance and carbon footprint of the delamination of multilayer flexible packaging films via carboxylic acids. Resources, Conservation and Recycling, 181, 106256.

Varmuza K., Filzmoser P., 2009, Introduction to multivariate statistical analysis in chemometrics, CRC Press, Boca Raton, FL, United States, ISBN 978-1420059472.

Venables W. N., Ripley B. D., 2002, Modern Applied Statistics with S. Springer Science+Business Media, New York, United States.

WTI AG, 2022, WTI AG. <www.wti-ag.com/home>, accessed 02.04.2023.

Zhai Y., Shen X., Quan T., Ma X., Zhang R., Ji C., Zhang T., Hong J., 2019, Impact-oriented water footprint assessment of wheat production in China. Science of The Total Environment, 689, 90–98.