Conformity of LCA Data to Benford's Law: A Country-Level Perspective

Bogdan Šinik
bogdan.sinik@famnit.upr.si
UP FAMNIT
Koper, Slovenia

Aleksandar Tošić
aleksandar.tosic@upr.si
UP FAMNIT, InnoRenew CoE
Koper, Slovenia

Abstract

Life cycle assessment (LCA) data is essential for evaluating the environmental impacts of products, services, and processes throughout their life cycle. Its accuracy depends largely on the underlying life cycle inventory. Currently, few tools exist to verify the integrity of such data. In earlier work, we introduced a novel approach to this issue by applying Benford's law as a preliminary data quality indicator. Benford's law describes the distribution of leading digits in naturally occurring numerical datasets. We first demonstrated that overall LCA datasets conform to Benford's law and that substantial differences exist among continents when analyzed in more detail.

In this study, we examined conformity at the country level to assess whether certain countries could represent their continents. All countries showed conformity when data were aggregated. However, performance varied: for example, Brazil aligned with the positive trend of its continent, while Switzerland performed notably worse. Similar patterns were observed when data were aggregated by compartments (air, water, soil, natural resources) and by selected elementary flow groups. Brazil again showed the strongest conformity, followed by the EU.

Our results confirm that country-level LCA data generally conform to Benford's law and that some countries can serve as continental representatives. At the same time, caution is needed when using data from countries with weaker conformity, as these datasets may require additional verification.

Keywords

Anomaly Detection, Benford's law, Data Integrity, Life-cycle assessment

1 Introduction

Life-Cycle Assessment (LCA) has established itself as a principal methodology for quantifying the environmental impacts of products and processes throughout their entire life cycle [12]. Given its extensive application across industrial sectors and in policy formulation, the verification of data integrity represents a critical requirement [5, 13]. The process of data acquisition is particularly decisive, as it directly conditions the accuracy of LCA outcomes and thereby determines the robustness and credibility of subsequent analyses [4]. Data collection is the crucial step in assuring accuracy of LCA results and it influences all subsequent steps of the analyses [4]. As there is no universal tool that is used for this, we aimed to employ Benford's law, widely known statistical tool for testing integrity of the underlying data and fraud detection [11]. Benford's law, also known as first-digit law, describes distribution of the first digits of numbers in naturally generated datasets. It is widely used in many fields, from finances and banking to fields such as astronomy [1], election result analysis [10], and scientific publishing networks [20]. As the general conformity of LCA data to Benford's law has been demonstrated [17], and prior research has also identified differences in conformity across continents and environmental compartments [18], we extend this analysis by testing several countries to assess whether conformity can also be expected at the national level. Furthermore, we investigate whether individual countries can serve as suitable representatives of their respective continents.

Information Society 2025, Ljubljana, Slovenia

2 Literature review

In our previous research, we examined differences in LCA data across continents with respect to heavy metals, greenhouse gases, toxic substances and carbon-based parameters. The results revealed significant intercontinental variation, particularly between Europe and Africa [18]. Not only did we demonstrate differences across continents, but we also validated these findings by establishing a correlation between Benford's conformity levels and environmental performance scores for the respective continents. Building on these results, we aimed to investigate whether similar patterns could also be observed at the country level.

Beyond our own application of Benford's law to environmental datasets, numerous studies have employed the same approach in related contexts. In the environmental field, Benford's law has been used to test ecotoxicological data for anomalies [9], verify the conformity of large biodiversity datasets [19], and assess the validity of reported emission reduction claims [6]. These applications demonstrate the versatility of Benford's law as a diagnostic tool for evaluating the integrity of environmental data.

In parallel, several studies have analyzed differences in LCA results among countries. For instance, one study compared LCA outcomes for dwellings in Spain, a developed country, and Colombia, a country under development. The results showed substantial differences driven by electricity mix, climate, and consumption habits: Spain's fossil fuel–intensive energy system and higher household demand produced larger impacts, whereas Colombia's reliance on hydropower and distinct usage patterns lowered overall burdens [15]. This evidence underscores the importance of examining LCA data at the national level, rather than relying solely on continental or global averages.

Finally, research has also pointed to the issue of uneven data quality across countries. A study from 2019 highlighted that LCA data coverage and completeness are generally more reliable for Europe and major economies, whereas many developing countries suffer from significant uncertainties, often requiring extrapolations or proxies [7]. Such discrepancies emphasize the need for careful validation when comparing national-level LCA data, and they provide further justification for assessing the integrity of country-level datasets.

3 Methodology

The first-digit law describes the characteristic frequency distribution of leading digits in numerical datasets. It is also referred to as Newcomb–Benford's law or simply Benford's law. According to Benford's law [16], the leading digits of many naturally occurring datasets follow a fixed probability distribution, provided that the data meet several conditions:

- Values are generated through mathematical combinations of numbers from multiple distributions.
- The dataset spans a wide numerical range (e.g., including values in the hundreds, thousands, tens of thousands, etc.).
- The dataset is sufficiently large, with at least 50–100 observations as a general guideline [14].
- The data are right-skewed, with the mean exceeding the median and a long right tail rather than a symmetric distribution.
- The dataset lacks a strict upper or lower bound, aside from a minimum of zero.

The equation for the distribution of the first digits of observed data is given in Equation 1.

$$P(d) = log_{10}(d+1) - log_{10}(d) = log_{10}(1+\frac{1}{d}) \quad (1)$$

One important aspect that needs to be decided when using Benford's law is minimal sample size. Although Benford's Law can be applied to relatively small samples ($50 \leq N \leq 100$) [14], recent studies recommend employing Monte Carlo simulations for more robust evaluation of sample size adequacy [8, 2]. One study suggests that conformity tests are most reliable when $N \geq 200$[2], while for $N < 1000$, researchers should allow for larger deviations before inferring non-conformity, consistent with Nigrini's recommendations [8].

In our previous continent-level study [18], we applied Monte Carlo sampling of the ecoinvent database to determine an appropriate sample size. We showed that conformity rates rise with sample size up to about $N = 500$, after which the results stabilize with more than 95% of samples conforming. Based on this, we established $N = 500$ as the threshold for analysis, consistent with recent recommendations [8], and we adopt the same threshold in the present country-level research.

The dataset employed in this study was ecoinvent version 3.10 (cut-off cumulative LCI), which is available under a paid license. We conducted the analysis in R using the benford.analysis package [3]. Zero values were automatically excluded, and we adapted the Benford function to handle negative numbers. From the 2654 columns in the ecoinvent dataset, 2648 numeric columns were used, representing 1190 distinct chemical substances. These were classified into five categories based on place of disposal: Air, Water, Soil, Natural Resources, and Inventory Indicator.

For this study, we focused on four groups of elementary flows from ecoinvent: carbon-based flows, heavy metals, toxic substances, and greenhouse gases. These groups were analyzed only for countries with sufficient observations (≥ 500), which included China, India, Switzerland, the USA, Canada, and Brazil. In addition, all EU member states were combined and reported under a single 'EU' category.

4 Results

The results for overall conformity across the countries align closely with those for the continents [18]. All examined countries demonstrated close conformity when their data were aggregated. All MAD values were less than 0.001, and all p-values were either 0 or extremely close to it (see Table 1).

Upon individually evaluating each elementary flow, the discrepancies are clearly visible. Brazil is clearly the superior performer, with less than 30% of elementary flows classified as non-conforming and over 40% exhibiting acceptable conformity. It is the only country that have multiple close conforming columns. It is succeeded by the EU, however it already exhibits a majority of non-conforming columns. India and China have less than 2% of conforming columns in total (Table 2). It is noteworthy that Switzerland exhibits poor results despite ecoinvent being based in the country.

After evaluating the selected categories of elementary flows, a consistent pattern emerged in the performance of individual countries. Notably, Brazil was the only nation to exhibit 'close conformity' for any elementary flows within these specific subgroups. Throughout the analysis, Brazil consistently demonstrated the highest level of conformity, whereas China and India consistently ranked the lowest, with their elementary flows being almost entirely non-conforming (see Table 3).

In the analysis of carbon-based flows, Brazil is the most prominent performer, with approximately 22% of its elementary flows classified as non-conforming. The performance of Canada matched that of the EU, showing one fewer non-conforming column. In contrast, China, India, and Switzerland did not have conforming columns at all. Notably, four elementary flows were non-conforming across all tested countries: fossil carbon-dioxide in the lower stratosphere and upper troposphere (X=1), non-fossil carbon dioxide in the air (X=9), carbon dioxide in the soil (X=12), and fossil carbon monoxide in the lower stratosphere and upper troposphere (X=17) (Figure 1).

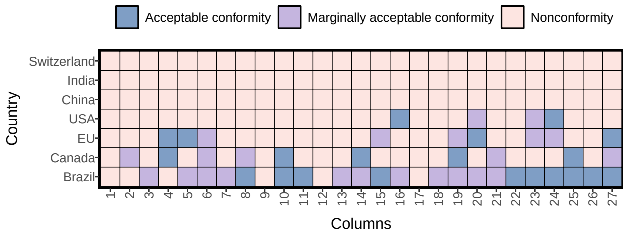

Figure 1: Conformity to Benford's law for carbon-based elementary flows across selected countries in the ecoinvent database.

A similar situation is present in the results for toxic substances. Brazil retains the top position, closely followed by the EU in this instance. At the bottom of the list, China and India again show 100% non-conforming flows, whereas Switzerland now exhibits more than 17% conforming columns. Only one flow, Lead-210 in urban air close to the ground (X=14), did not conform for any country. It is also interesting to note that Lead II in surface water (X=13) only marginally conforms for Switzerland (Figure 2).

With respect to the elementary flows of greenhouse gases, a consistent trend was observed. China and India demonstrated the lowest conformity. However, India presented one conforming elementary flow, specifically trifluoromethane in urban air close to the ground (X=48). This particular flow exhibited acceptable conformity for both India and Brazil, which were the sole countries to conform for this elementary flow. Conversely, trifluoromethane

Table 1: Statistical tests on Benford's conformity for all observations in the selected LCI database by country.

Country	ChiSq	ChiSq(P)	MantissaArc	MantissaArc(P)	MADConformity	MAD	Observations
Switzerland	266.6629	0.0000	0.0000	0.0000	Close conformity	0.0006	4,828,955
European Union	71.5715	0.0000	0.0000	0.0000	Close conformity	0.0004	4,700,559
Brazil	7.6982	0.4635	0.0000	0.1694	Close conformity	0.0002	1,658,861
China	69.9057	0.0000	0.0000	0.5208	Close conformity	0.0005	1,936,686
Canada	23.3071	0.0030	0.0000	0.2541	Close conformity	0.0003	1,719,686
India	88.8233	0.0000	0.0000	0.0371	Close conformity	0.0006	1,200,466
USA	33.5820	0.0000	0.0000	0.1689	Close conformity	0.0004	1,240,267

Table 2: MAD Conformity counts and percentages by country, sorted in increasing order by number of nonconforming values, where CC = Close conformity, AC = Acceptable conformity, MAC = Marginally acceptable conformity, and NC = Non-conformity

Country	CC (%)	AC (%)	MAC (%)	NC (%)	Dataset
Brazil	41 (1.87%)	900 (41.1%)	620 (28.31%)	629 (28.72%)	ecoinvent
EU	9 (0.41%)	400 (18.26%)	464 (21.19%)	1317 (60.14%)	ecoinvent
Canada	1 (0.05%)	125 (5.71%)	287 (13.11%)	1777 (81.14%)	ecoinvent
USA	1 (0.05%)	114 (5.21%)	262 (11.96%)	1813 (82.79%)	ecoinvent
Switzerland	0 (0.0%)	38 (1.74%)	117 (5.34%)	2035 (92.92%)	ecoinvent
China	0 (0.0%)	7 (0.32%)	27 (1.23%)	2156 (98.45%)	ecoinvent
India	0 (0.0%)	5 (0.23%)	25 (1.14%)	2159 (98.63%)	ecoinvent

Table 3: MAD Conformity counts per country, where CC = Close conformity, AC = Acceptable conformity, MAC = Marginally acceptable conformity, and NC = Non-conformity

Country	CC (%)	AC (%)	MAC (%)	NC (%)	Data
Brazil	0 (0.0%)	18 (37.5%)	13 (27.08%)	17 (35.42%)	Green house gases
European Union	0 (0.0%)	8 (16.67%)	8 (16.67%)	32 (66.67%)	Green house gases
Canada	0 (0.0%)	3 (6.25%)	5 (10.42%)	40 (83.33%)	Green house gases
Switzerland	0 (0.0%)	0 (0.0%)	2 (4.17%)	46 (95.83%)	Green house gases
USA	0 (0.0%)	1 (2.08%)	1 (2.08%)	46 (95.83%)	Green house gases
India	0 (0.0%)	1 (2.08%)	0 (0.0%)	47 (97.92%)	Green house gases
China	0 (0.0%)	0 (0.0%)	0 (0.0%)	48 (100.0%)	Green house gases
Brazil	4 (0.02%)	56 (34.15%)	47 (28.65%)	57 (34.75%)	Heavy metals
European Union	0 (0.0%)	34 (20.73%)	39 (23.78%)	91 (55.49%)	Heavy metals
Canada	0 (0.0%)	7 (4.27%)	23 (14.02%)	134 (81.71%)	Heavy metals
Switzerland	0 (0.0%)	6 (3.66%)	19 (11.59%)	139 (84.76%)	Heavy metals
USA	0 (0.0%)	3 (1.83%)	10 (6.1%)	151 (92.07%)	Heavy metals
India	0 (0.0%)	0 (0.0%)	3 (1.84%)	160 (98.16%)	Heavy metals
China	0 (0.0%)	2 (1.22%)	1 (0.61%)	161 (98.17%)	Heavy metals
Brazil	0 (0.0%)	10 (37.04%)	11 (40.74%)	6 (22.22%)	Carbon
Canada	0 (0.0%)	5 (18.52%)	5 (18.52%)	17 (62.96%)	Carbon
European Union	0 (0.0%)	4 (14.81%)	5 (18.52%)	18 (66.67%)	Carbon
USA	0 (0.0%)	2 (7.41%)	2 (7.41%)	23 (85.19%)	Carbon
China	0 (0.0%)	0 (0.0%)	0 (0.0%)	27 (100.0%)	Carbon
India	0 (0.0%)	0 (0.0%)	0 (0.0%)	27 (100.0%)	Carbon
Switzerland	0 (0.0%)	0 (0.0%)	0 (0.0%)	27 (100.0%)	Carbon
Brazil	0 (0.0%)	11 (47.83%)	5 (21.74%)	7 (30.43%)	Toxic substances
European Union	0 (0.0%)	8 (34.78%)	6 (26.09%)	9 (39.13%)	Toxic substances
Canada	0 (0.0%)	1 (4.35%)	6 (26.09%)	16 (69.57%)	Toxic substances
Switzerland	0 (0.0%)	2 (8.7%)	2 (8.7%)	19 (82.61%)	Toxic substances
USA	0 (0.0%)	0 (0.0%)	1 (4.35%)	22 (95.65%)	Toxic substances
China	0 (0.0%)	0 (0.0%)	0 (0.0%)	23 (100.0%)	Toxic substances
India	0 (0.0%)	0 (0.0%)	0 (0.0%)	23 (100.0%)	Toxic substances

in non-urban the air or from high stacks (X=47) conformed exclusively in the European Union, although marginally. Furthermore, ten elementary flows were identified as non-conforming across all analyzed countries. (see Figure 3). Regarding heavy metal elementary flows, the observed trend remained consistent. Brazil continued to exhibit the lowest proportion of non-conforming columns, approximately 35%. Notably, Brazil demonstrated close conformity for four elementary flows: Chromium III in surface water (X=38), long-term Mercury II in ground water (X=130), Nickel II in soil (X=143), and Zinc II in surface water (X=160). For the first time, China exhibited conforming elementary flows.

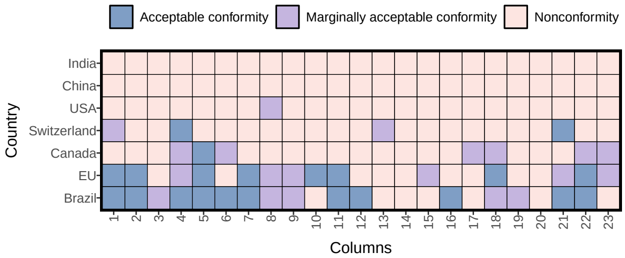

Figure 2: Conformity to Benford's law for toxic substances-based elementary flows across selected countries in the ecoinvent database.

Specifically, conformity was observed for long-term Iron ion in air in low population density (X=71), Lead II in ground water (X=94), and Manganese II in water (X=117). China was the sole country to conform for Lead II in the ground water. India also demonstrated conformity for three elementary flows. Similar to China, India conformed for long-term Iron ion in air in low population density, though only marginally. The other two conforming flows were Mercury II in lower stratosphere and upper troposphere (X=122) and Nickel II in lower stratosphere and upper troposphere (X=136). It is important to note that for Nickel II in agricultural soil (X=140), insufficient data prevented calculation due to high incidence of missing values. Furthermore, a total of 27 elementary flows were non-conforming across all analyzed countries. (Figure 4).

5 Conclusion and Discussion

When testing the conformity of countries with sufficient data ($N \geq 500$ observations), the results clearly revealed which nations were driving continental patterns. China and India consistently exhibited non-conformity across nearly all elementary flows, aligning with the broader Asian trend. The European Union demonstrated moderate performance, while Switzerland frequently ranked among the poorest performers alongside China and India. Notably, none of these selected countries achieved conformity levels comparable to the overall European results, suggesting that Europe's superior performance may be attributed to non-EU European countries not individually analyzed in this study.

Particularly interesting was Switzerland's poor performance despite ecoinvent being a Swiss-based database. This unexpected finding challenges assumptions about the relationship between database origin and data quality for specific geographical regions.

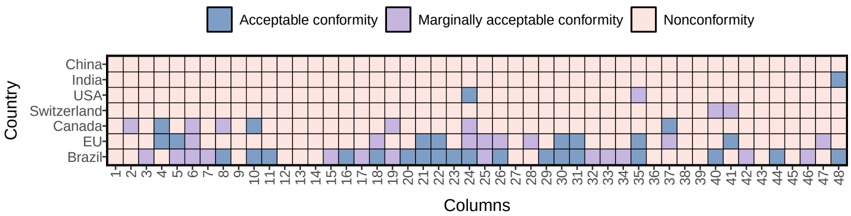

Figure 3: Conformity to Benford's law for greenhouse gases-based elementary flows across selected countries in the ecoinvent database.

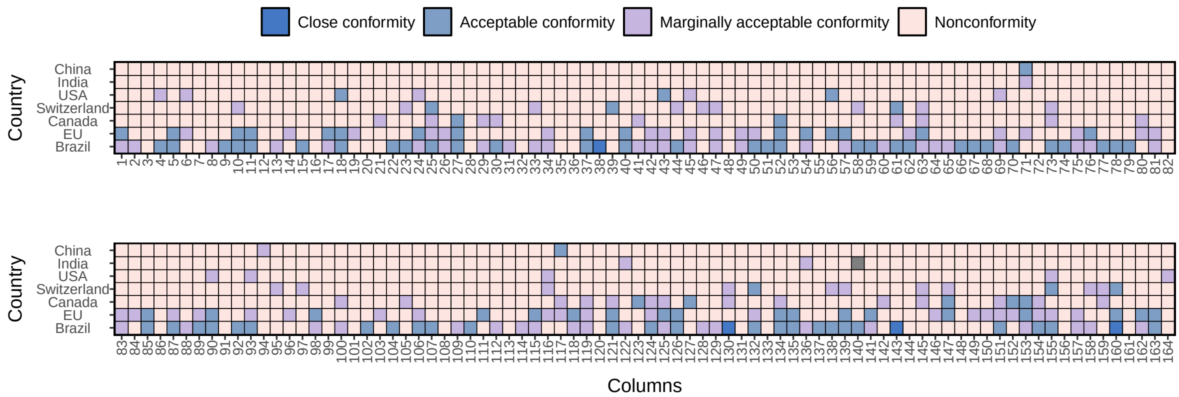

Figure 4: Conformity to Benford's law for heavy metal-based elementary flows across selected countries in the ecoinvent database.

Conversely, Brazil emerged as the clear standout performer, contributing substantially to South America's strong overall conformity results. Canada and the United States demonstrated similar performance levels, maintaining consistent middle-tier positioning relative to other countries, which aligned well with North America's intermediate continental ranking.

These country-specific results provide valuable insights into the regional variations observed at the continental level and highlight the importance of examining data quality at multiple geographical scales when assessing LCA database integrity.

References

[1] Theodoros Alexopoulos and Stefanos Leontsinis. 2014. Benford's law in astronomy. *Journal of Astrophysics and Astronomy*, 35, 639–648.

[2] Andrea Cerasa. 2022. Testing for benford's law in very small samples: simulation study and a new test proposal. *PLoS One*, 17, 7, e0271969.

[3] Carlos Cinelli and Maintainer Carlos Cinelli. 2022. Package 'benford. analysis'. *Benford analysis for data validation and forensic analytics. Version 0.1*, 5.

[4] Andreas Cirotha, Claudia Di Noia, Salwa Syed Burhana, and Michael Srockaa. 2019. Lca database creation: current challenges and the way forward. *Indonesian J. Life Cycle Assess. Sustain*, 3, 41–51.

[5] Steven A Cohen. 2013. Life cycle assessment and the us policy-making context. *Emerging technologies: socio-behavioral life cycle approaches*, 217–44.

[6] Matthew A Cole, David J Maddison, and Liyun Zhang. 2020. Testing the emission reduction claims of cdm projects using the benford's law. *Climatic Change*, 160, 407–426.

[7] Eleonora Crenna, Michela Secchi, Lorenzo Benini, and Serenella Sala. 2019. Global environmental impacts: data sources and methodological choices for calculating normalization factors for lca. *The International Journal of Life Cycle Assessment*, 24, 10, 1851–1877.

[8] Archibald de Araújo Silva and Maria Aparecida Gouvêa. 2023. Study on the effect of sample size on type i error, in the first, second and first-two digits excessmad tests. *International Journal of Accounting Information Systems*, 48, 100599. DOI: https://doi.org/10.1016/j.accinf.2022.100599.

[9] Pepijn de Vries and Albertinka J. Murk. 2013. Compliance of lc50 and noec data with benford's law: an indication of reliability? *Ecotoxicology and Environmental Safety*, 98, 171–178. DOI: https://doi.org/10.1016/j.ecoenv.2013.09.002.

[10] Joseph Deckert, Mikhail Myagkov, and Peter C Ordeshook. 2011. Benford's law and the detection of election fraud. *Political Analysis*, 19, 3, 245–268.

[11] R. M. Fewster. 2009. A simple explanation of benford's law. *The American Statistician*, 63, 1, 26–32. eprint: https://doi.org/10.1198/tast.2009.0005. DOI: 10.1198/tast.2009.0005.

[12] Jeroen B Guinée, Reinout Heijungs, Gjalt Huppes, Alessandra Zamagni, Paolo Masoni, Roberto Buonamici, Tomas Ekvall, and Tomas Rydberg. 2011. Life cycle assessment: past, present, and future. (2011).

[13] Miao Guo and RJ Murphy. 2012. Lca data quality: sensitivity and uncertainty analysis. *Science of the total environment*, 435, 230–243.

[14] David A Kenny. 2015. Measuring model fit. (2015). http://davidakenny.net/cm/fit.htm.

[15] Oscar Ortiz-Rodríguez, Francesc Castells, and Guido Sonnemann. 2010. Life cycle assessment of two dwellings: one in spain, a developed country, and one in colombia, a country under development. *Science of The Total Environment*, 408, 12, 2435–2443. DOI: https://doi.org/10.1016/j.scitotenv.2010.02.021.

[16] Tommie W Singleton. 2011. IT Audit Basics: Understanding and Applying Benford's Law. *Isaca Journal*, 3, 6.

[17] Bogdan Šinik and Aleksandar Tošic. [n. d.] Lca data conforms to benford's law.

[18] Bogdan Šinik and Aleksandar Tošić. 2025. Testing life-cycle assessment data quality with benford's law reveals geographic variation. *Ecological Informatics*, 103227.

[19] Judit K. Szabo, Lucas Rodriguez Forti, and Corey T. Callaghan. 2023. Large biodiversity datasets conform to benford's law: implications for assessing sampling heterogeneity. *Biological Conservation*, 280, 109982. DOI: https://doi.org/10.1016/j.biocon.2023.109982.

[20] Aleksandar Tošić and Jernej Vičič. 2021. Use of benford's law on academic publishing networks. *Journal of Informetrics*, 15, 3, 101163. DOI: https://doi.org/10.1016/j.joi.2021.101163.