Hybridization of EXIOBASE Environmental Extensions with Micro-scale LCA Data

Your Name

July 18, 2025

1 Introduction & Motivation

Hybridizing multi-regional input-output (MRIO) models, such as EXIOBASE, with process-level Life Cycle Assessment (LCA) coefficients, provides a scientifically robust methodology for integrating granular, product-level emission factors into macroeconomic accounts. The purpose of this hybridization is to better represent heterogeneity and recent data in environmental impacts, especially for the food sector, where conventional MRIO sector averages can obscure real-world differences.

2 Datasets Overview

2.1 Open Food Facts (OFF)

- **Description:** An open global database of food products. Contains product names, descriptions, ingredients, barcodes, and category data.
- **Used Shape:** $(N_{\text{products}}, N_{\text{fields}})$, e.g. $(58, 933, 12)$ after English filtering and cleaning.
- **Year:** Data downloaded in 2025; products from 2010–2025.
- **Usage:** Micro-level linkage of consumer products to LCA categories and MRIO sectors via text analysis.

2.2 Clark et al. (2022) LCA Dataset

- **Description:** Meta-analysis of LCA studies for food, extending Poore & Nemecek (2018). Contains per-product emission factors: CO_2-eq, land use, eutrophication, water use, etc.
- **Columns:** `Entity` (food type), `ghg_kg` (kg CO_2/kg), `land_use_kg` (m^2/kg), etc.
- **Shape:** (212, 18).
- **Year:** Most data: 2000–2020, with meta-study median year 2010.
- **Country coverage:** 119 countries; highest study density in Europe, US, Brazil, China, India. Represents global means.

- **Usage:** Source of process-based environmental coefficients for micro-macro integration.

2.3 EXIOBASE 3.8.2 (2022.zarr)

- **Description:** MRIO model providing global economic and environmental flows.
- **Shape:**
  - `T` (transactions): $(189, 163, 189, 163)$ – region-sector to region-sector flows.
  - `Y` (final demand): $(189, 163, 189)$ – sector to final demand by region.
  - `Q` (environmental extensions): $(19, 189, 163)$ – 19 indicators (CO_2, land, water, etc.) for region-sectors.
  - `input_region`, `input_sector`: lists of region/sector names (lengths 189 and 163).
- **Year:** 2022 snapshot.
- **Usage:** Baseline for environmental impact; receives LCA-derived coefficients.

3 Conceptual and Mathematical Framework

3.1 MRIO Model and Leontief System

Let $n = N_{\text{regions}} \cdot N_{\text{sectors}}$ (here $n = 189 \times 163 = 30,807$).

- **Transactions:** $Z \in R^{n \times n}$, from flattening T.
- **Final demand:** $y \in R^n$, from reshaping and summing Y.
- **Total output:** $x = Z\mathbf{1} + y$
- **Technical coefficients:** $A = Z \cdot \text{diag}(x)^{-1}$ (or $A_{ij} = Z_{ij}/x_j$).
- **Leontief inverse:** $L = (I - A)^{-1}$
- **Environmental extensions:** $Q \in R^{m \times n}$ (here $m = 19$), $Q_{k,j}$ is indicator k for region-sector j.
- **Intensity matrix:** $S = Q \cdot \text{diag}(x)^{-1}$, $S_{k,j} = Q_{k,j}/x_j$
- **Impact matrix:** $D = SL$ (each row k is the total impact per sector j via the full supply chain)

3.2 Hybridization: Injecting LCA Coefficients

Original: EXIOBASE Q-matrix elements $Q_{k,j}$ are based on aggregate sectoral/national statistics.

Hybridization step:

- For sectors j mapped to a product-level LCA coefficient (e.g., CO_2/kg from Clark et al.), **replace** $Q_{k,j}$ with the micro-scale value, re-scaled as necessary (e.g., from kg/kg to kg/Euro using price or output data).
- If $S_{k,j}^{\text{base}}$ is the original EXIOBASE intensity and $S_{k,j}^{\text{LCA}}$ the new micro coefficient:

$$S_{k,j}^{\text{hybrid}} = \begin{cases} S_{k,j}^{\text{LCA}} & \text{if LCA mapping found for } j \\ S_{k,j}^{\text{base}} & \text{otherwise} \end{cases}$$

- This gives a new Q^{hybrid} and S^{hybrid}, with $D^{\text{hybrid}} = S^{\text{hybrid}} L$.

3.3 Detailed Pipeline Steps (as in code)

1. **Data cleaning:** Filter OFF for English-language products and valid names/ingredients.
2. **Semantic mapping:** Use sentence transformers (MiniLM) to compute embeddings for each OFF product and Clark LCA food type. Assign best-matching LCA group.
3. **Aggregation:** For each MRIO sector, aggregate mapped products to calculate mean LCA coefficients ($\bar{c}_{\text{CO}_2}$, $\bar{l}_{\text{land}}$).
4. **Sector alignment:** Align MRIO sector descriptions to the aggregated food LCA coefficients.
5. **Hybridization:** Load EXIOBASE Q and replace sector-level coefficients with the new LCA-based values for matched sectors.
6. **Leontief solution:** Recompute Leontief inverse L, and hybridized impact matrix D.
7. **Visualization:** Compare absolute/relative changes in sectoral CO_2 intensity, map sector/region indices to names, export results.

4 Why Use Process-LCA Coefficients?

- **EXIOBASE Q** values are highly aggregated, based on national inventory and sector-level data, often lacking product or technology resolution.
- **Clark/Poore & Nemecek LCA** coefficients are product- and process-specific, integrating thousands of studies at the farm, plant, or commodity level, reflecting true heterogeneity.
- Replacing (hybridizing) MRIO Q with LCA values improves scientific accuracy for policies targeting specific products (e.g., dietary change, eco-labeling, or Scope 3 accounting).

5 Country Coverage and Comparison

- **Clark/Poore & Nemecek LCA:** 119 countries, dominated by Western (EU/US), Latin American (Brazil), and Asian (China, India) studies. *Coefficients are global means, but underlying regional bias exists.*
- **EXIOBASE:** Comprehensive global coverage with region/sector splits, but sectoral Q is often based on top-down reporting.
- **Correlation:** For sectors with LCA data dominated by Western studies, hybridized coefficients are likely most representative for US/EU regions. For others, some mismatch/uncertainty remains.

6 Mathematical Summary

Given:

$$
\begin{aligned}
Z &= \text{Flattened interindustry transactions matrix} \\
y &= \text{Flattened final demand vector} \\
x &= Z\mathbf{1} + y \\
A &= Z \cdot \text{diag}(x)^{-1} \\
L &= (I - A)^{-1} \\
Q &= \text{Environmental extensions matrix (from EXIOBASE or hybridized)} \\
S &= Q \cdot \text{diag}(x)^{-1} \\
D &= SL
\end{aligned}
$$

7 References

1. Clark, M. A., Domingo, N. G., Colgan, K., Thakrar, S. K., Tilman, D., Lynch, J., ... & Hill, J. D. (2020). Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. Science, 370(6517), 705-708.

2. Poore, J., & Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. Science, 360(6392), 987-992.

3. Stadler, K., Wood, R., Bulavskaya, T., Södersten, C. J., Simas, M., Schmidt, S., ... & Tukker, A. (2018). EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables. Journal of Industrial Ecology, 22(3), 502-515.