A PHYSICALLY EXTENDED EEIO FRAMEWORK FOR MATERIAL EFFICIENCY ASSESSMENT IN UNITED STATES MANUFACTURING SUPPLY CHAINS

PRE-PRINT

Heather Liddell* – Purdue University
Beth Kelley – Purdue University
Liz Wachs – National Renewable Energy Laboratory
Alberta Carpenter – National Renewable Energy Laboratory
Joe Cresko – U.S. Department of Energy

*Corresponding Author: liddellh@purdue.edu

Abstract

A physical assessment of material flows in an economy (e.g., material flow quantification) can support the development of sustainable decarbonization and circularity strategies by providing the tangible physical context of industrial production quantities and supply chain relationships. However, completing a physical assessment is challenging due to the scarcity of high-quality raw data and poor harmonization across industry classification systems used in data reporting. Here we describe a new physical extension for the U.S. Department of Energy's (DOE's) EEIO for Industrial Decarbonization (EEIO-IDA) model, yielding an expanded EEIO model that is both physically and environmentally extended. In the model framework, the U.S. economy is divided into goods-producing and service-producing subsectors, and mass flows are quantified for each goods-producing subsector using a combination of trade data (e.g., UN Comtrade) and physical production data (e.g., U.S. Geological Survey). Given that primary-source production data are not available for all subsectors, price-imputation and mass-balance assumptions are developed and used to complete the physical flows dataset with high-quality estimations. The resulting dataset, when integrated with the EEIO-IDA tool, enables the quantification of environmental impact intensity metrics on a mass basis (e.g., CO_2eq/kg) for each industrial subsector. This work is designed to align with existing DOE frameworks and tools, including the EEIO-IDA tool, the DOE Industrial Decarbonization Roadmap (2022), and Pathways for U.S. Industrial Transformations study (2025).

Introduction and Motivation

An economic input-output (IO) model tracks the flow of goods through an economy and can be used to analyze the accrual of value in product supply chains, from raw material extraction to intermediates to final products. This is shown conceptually in Figure 1. These models can offer insights into the overall structure of the economy, including where raw materials from extraction industries enter the economy, the extent to which industries transact with each other in domestic or international supply chains, and how these transactions between industries contribute to cumulative value addition in gross output. These features make IO models ideal for the study of circularity.[1-5]

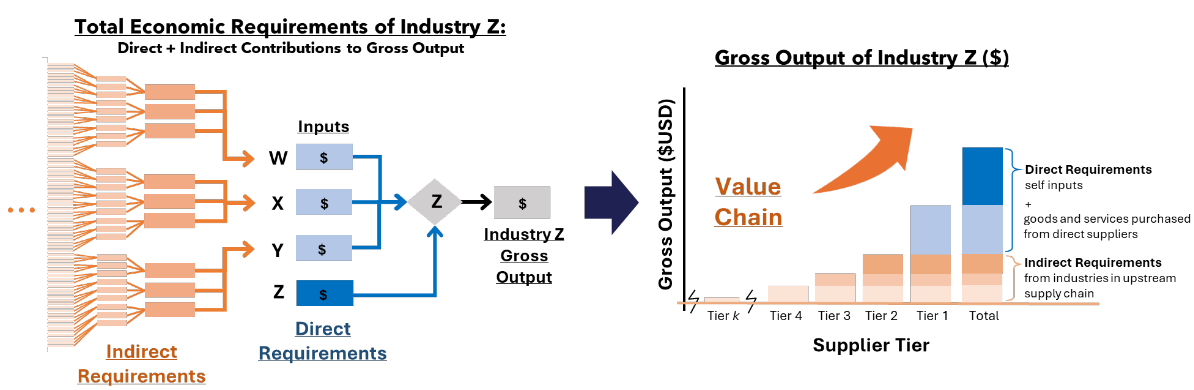

Figure 1. Schematic representation of the economic IO framework

An environmentally extended input-output (EEIO) model leverages the economic IO framework (and the comprehensive, high-quality scope of economic trade data) to quantify environmental impacts in supply chains. The economy-level view afforded by an EEIO model allows for the quantification of emissions and other environmental flows for a cross-section of goods and services representing (in aggregate) the entire economy. This includes assessment of difficult-to-measure Scope 3 emissions in a product's upstream supply chain, as illustrated in Figure 2. Furthermore, the EEIO method minimizes truncation errors that are common in process-based life cycle assessment (LCA).

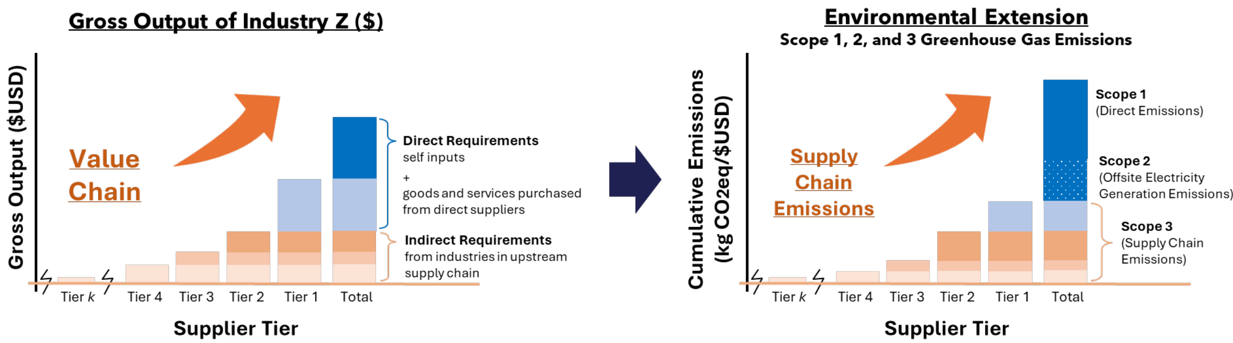

Figure 2. Environmental extension of an economic IO model to quantify greenhouse gas emissions in product supply chains

Despite the advantages of the economic foundation of the EEIO method, the monetary basis is also one of its main limitations. EEIO methods generally ignore the physical quantities of goods produced: all environmental impacts are expressed with respect to an output of goods or services in monetary units (e.g., $USD). Physical quantities do not typically factor into an EEIO assessment. As a result, EEIO models characterize environmental impact intensities in monetary terms, such as kg CO_2e/$USD. This is true of all of the major EEIO models available for the United States, including the U.S. Environmental Protection Agency's (EPA's) USEEIO model,[6,7] Carnegie Mellon's EIO-LCA model,[8] and the newer U.S. Department of Energy's (DOE's) EEIO-IDA model and scenario analysis tool.[4,9] The authors are not aware of any U.S.-focused model that allows for calculation of environmental impact intensities in mass-based units such as kg CO_2e/kg product.

Adding a physical extension to an EEIO model would enable estimation of industry-level environmental impacts on a mass basis (e.g., kg CO_2-eq/kg) rather than only on a monetary basis, which would be useful for more accurate Scope 3 product life cycle assessments, decarbonization scenario modeling at the product and industry level, and other purposes. Unfortunately, a barrier to such an extension is the limited availability of U.S. goods production data in physical units. This research develops a rational framework for estimating physical production when primary data are sparse. We use the framework to construct a new dataset for physical goods production in the United States, consisting of production data for 233 individual industries across 25 industrial subsectors. Data quality, limitations, and planned integration with DOE's EEIO-IDA model are discussed.

Review of Related Work

While EEIO models don't typically incorporate physical units, a related family of models known as physical input-output tables (PIOTs) use physical units in place of monetary units in IO tables. PIOTs express goods exchanges in terms of physical units, such as kilograms, in an input-output framework. Use of these tables could address some of the challenges associated with the current practice of using monetary units as a proxy for physical units to represent flows of physical materials—but unfortunately, PIOTs have not yet been widely adopted.[10] Presently, the availability and coverage of PIOT models for the United States is limited and incomplete. Furthermore, PIOTs are rarely combined with EEIO models. Perhaps the most prominent example of a combined PIOT/EEIO model is EXIOBASE, a global multi-regional EEIO model that is available in both monetary and hybrid monetary/physical units.[11,12]

EXIOBASE datasets have been found to be in reasonable agreement with national statistics, LCA datasets, and other IO models in multiple studies,[13,14] though (unsurprisingly for such a large model), authors have also found that the level of agreement is imperfect.[15,16,17] EXIOBASE has a resolution (number of industries) higher than that of most national statistics used to build the model, requiring disaggregation assumptions that depend on analyst judgment. Results would therefore reasonably be expected to vary across models, and such variations should be properly attributed to uncertainty in both models rather than inaccuracy in either one. More importantly, data in EXIOBASE and similar models can only be as good as the national datasets from which the model is built. For mass-based physical goods production in the United States, national data are extremely limited. This study aims to fill that gap.

Approach

Physical Extension of the IO Model

Physical extension of an economic IO model typically involves constructing a physical input-output table (PIOT), which expresses goods exchanges in terms of physical units rather than monetary units in an input-output framework. The generation of a PIOT is a time- and data-intensive process requiring the use of detailed trade data to construct an entire input-output matrix representing the material exchanges. At the level of the U.S. economy, this would always result in a "hybrid" physical/monetary model because the economy produces a combination of goods and services. The output of an industry can only be expressed in physical units for the goods-producing industries (not the services). Therefore, the unit of output for the service-producing industries remains monetary, while the goods-producing industries are represented with a mass output unit. The multi-unit, hybrid nature of such a table complicates its use and makes this method difficult and inaccessible to nonexperts.

Here we suggest a simple approach that avoids the necessity of a hybrid IO table altogether in the physical IO extension. Rather than constructing a PIOT (a complete IO table that incorporates physical units), we instead physically extend the economic IO table in a manner analogous to EEIO environmental extension. To environmentally extend an IO table in the normal way, we leverage an environmental impact vector as shown in equation (1):

$$E = (I - A)^{-1} \cdot d \cdot r, \tag{1}$$

where E is the environmental impact vector (representative unit: kg CO_2e); A is the Direct Requirements matrix (in monetary units) and $(I - A)^{-1}$ is the Leontief inverse of A; d is the final demand vector (in monetary units); and r is the environmental impact intensity (representative unit: kg CO_2e/\$). Here, we say "representative unit" because this framework can be used to examine any environmental impact, not just greenhouse gas emissions. When many impacts are assessed, all of the environmental impact vectors are combined in the environmental satellite matrix, but each is still analyzed according to equation (1).

Analogously, to physically extend the IO table, we can use the same approach, using a physical production vector P (in mass units such as kg) to compute a characteristic industry output price p (in units of \$/kg):

$$P = (I - A)^{-1} \cdot d \cdot (1/p). \tag{2}$$

Combining equations (1) and (2), we can physically and environmentally extend the economic input output model, giving us the convenient unit we seek for environmental impact intensity (kg CO_2e/kg):

$$E = (I - A)^{-1} \cdot d \cdot (1/p) \cdot r^{*}, \tag{3}$$

where r^* is a modified impact intensity vector with a new unit of kg CO_2e/kg for all goods-producing industries in the model. The original units of kg CO_2e/\$ will be maintained in r^* for the service-producing industries. In this simple approach, the minimum data requirement to physically extend an EEIO model is a production vector P of length n, containing the physical production values (in kg) of all goods-producing industries included in the model. For service-producing industries, monetary units can be maintained in P; in this case p is unity and the manipulation has no effect.

Quantifying Production in U.S. Industries

The US Bureau of Economic Analysis (BEA) releases the U.S. IO tables at the sector level (12 sectors) annually, summary level (71 sectors) annually, and detail level (over 400 sectors) every five years. In this study, we seek to quantify physical flows for the years 2018 and 2022 at both the summary level and the detail level. These years correspond to the survey years for the most recent Manufacturing Energy Consumption Survey (MECS) data release (2018) and the forthcoming MECS data release expected later this year (2022). These new datasets are intended for future integration with DOE's EEIO-IDA tool (Gause 2023) and alignment with ongoing decarbonization modeling efforts at DOE, some of which were recently published in DOE's *Industrial Decarbonization Roadmap*[18] and *Transformative Pathways for U.S. Industry: Unlocking American Innovation*[19] reports.

To satisfy the data requirements in the physical extension approach described in the previous section, primary-source production data from governmental or other high-quality sources are preferred. Unfortunately, production data in physical units from high-quality sources are extremely limited for the U.S. economy. The heatmap of Figure 3 illustrates our findings: of the 263 detail-level industries in goods-producing subsectors, we were able to locate high-quality production data with good industry coverage for only 48 (fewer than 20% of all goods-producing industries). For six additional industries, we were able to find data for selected products providing partial data coverage of the industry. Two of the industries within goods-producing subsectors provided support services and had no physical flows. That left 207 industries out of 263 (78%) with either no data found at all or very poor data coverage.

Considering the poor overall data coverage and the need for a complete dataset for physically extended EEIO modeling, alternative approaches were developed and used to estimate physical flows for cases in which primary-source data were unavailable. Physical flows were assessed using the following three methods, with the choice of principal data collection method selected depending on data availability and quality for an individual industrial subsector or goods-producing industry:

1. **Data-driven approach** (first choice, where data are available)

This method involves direct collection of physical production data from high-quality government sources such as the U.S. Geological Survey (USGS) or the Energy Information Administration (EIA), which can be used to provide estimates of mass directly. In some cases, detailed data were available for selected industries or products falling within a given subsector, but not for the full subsector. In such cases, partial data from primary sources can be combined with a complete sectoral dataset compiled using the price-driven approach (see below) to improve aggregated estimates and assess the data quality of the price-driven data for the industries with redundancy.

2. **Price-driven approach** (second choice, if data-driven method is not possible)

For industrial subsectors that do not have comprehensive physical output data available from government sources, mass is imputed by dividing the subsector's production in monetary units (available in the BEA tables[1]) by an assumed price for that subsector's output ($/kg). Prices are estimated using trade data from UN Comtrade, which reports cost, insurance, and freight (CIF) trade values (USD) and net weight (kg) by HS commodity codes,[2] and USA Trade, which provides calculated duties by HS commodity codes (the latter being necessary to convert the CIF values extracted from Comtrade to producer prices, aligning with IO table values).

Comtrade includes price information on both imports and exports, with data reported by HS commodity code. For the United States, import data are preferred to export data (when using as an estimate of domestic prices) due to the nature of the U.S. economy as a net importer. HS commodity codes for imports are translated to NAICS codes[3] and then to BEA industry groupings[4] in Python using crosswalk concordance files available from the U.S. Census Bureau.[5] The CIF values, net weight, and calculated duty data are then aggregated according to BEA industry groupings to generate a characteristic price for each subsector. To convert the CIF values to producer price values, the following steps were completed:

a. For each industry, the calculated duty value (from USA Trade) was divided by the CIF values (from UN Comtrade) to output a duty rate.
b. Three duty rate tiers were established by identifying the minimum and maximum duty rates across all industries with a non-zero duty rate and defining three equally sized intervals (low, medium, and high duty) to span the range between min/max.
c. Industries with duty rates falling within the limits of the upper interval were defined as "high duty," industries with duty rates falling within the central interval were defined as "medium duty," and industries with duty

[1] BEA tables being used for production detail (in monetary units) include "The Use of Commodities by Industries, After Redefinitions (Producer's Prices) - Summary" and "Gross Output by Industry" for the year 2018 (BEA 2024a; 2024b).
[2] Harmonized System (HS) Codes: an industry classification system administered by the World Customs Organization (WCO) that classifies traded products. Commodities are assigned six-digit codes.
[3] North American Industry Classification System (NAICS): an industry classification developed under the Office of Management and Budget (OMB) that classifies sectors, subsectors, industries, and commodities of the U.S. business economy. These are organized by 2-, 3-, 4-, and 6-digit codes, respectively.
[4] Bureau of Economic Analysis (BEA) industry groupings: classification system used by BEA to analyze industry data, including the input-output accounts. These codes are based on NAICS with some minor differences in aggregation of subsectors.
[5] U.S. Census Bureau: The U.S. Census Bureau provides crosswalk concordance files to translate HS codes to NAICS codes.

rates within the lower interval were defined as “low duty.” Industries were not necessarily evenly distributed across the three intervals.

d. An average duty value was computed for each duty tier, calculated as the mean duty rate for all industries grouped within each duty tier.
e. Industries within each duty tier were assigned the average duty rate for that tier. For the sectors with a duty rate value of zero in USA Trade, no duty was assigned.
f. The basic price for each industry was calculated from UN Comtrade CIF and net weight data, combined with the assigned duty rate, using equation (4):

$$Basic\ Price = \frac{CIF + CIF * Assigned\ Duty\ Rate}{Net\ Weight\ (kg)}. \tag{4}$$

g. Finally, the basic price is converted to producer prices using a ratio of basic to producer prices extracted from the Tau matrix[20] (i.e., the basic-to-producer-price ratio matrix) in USEEIO v2.3:

$$Producer\ Price = \frac{Basic\ Price}{Tau\ ratio}. \tag{5}$$

Physical values are then calculated by dividing the subsector-specific Gross Output in USD by the subsector producer price.

When partial data on physical flows are available from government sources (i.e., the data-driven method), these data are compared and used to assess data quality for the price-driven estimates (see Data Collection Method Comparison). In future iterations, redundant datasets may be further integrated to compose a “best available” dataset to improve data quality.

3. Input-driven approach:

The final method for assessing physical production, input-driven, is considered the least accurate but is used when neither of the previous two methods are possible, which is true (e.g.) for the Construction subsector. We are forced to use the input-driven method if a subsector produces or transforms physical goods, but no high-quality sources are available for physical production (precluding the use of the data-driven method) and the subsector does not participate in international trade (precluding the use of the price-driven method). Prices in these industries are estimated based on a mass balance as the sum of inputs, using equation (6), which is modified from an expression in the UN SUT Handbook[21] to incorporate a waste coefficient:

$$p_i = \frac{x_i}{(1-w_i)\sum_{j=1}^{n} z_{i,j}}, \tag{6}$$

where p is the characteristic price of industry i, x_i is industry i’s total output in dollars, $z_{i,j}$ is the mass of the inputs from sector j used by sector i, and w_i is a waste coefficient for industry i. The waste coefficient is a value between zero and one that accounts for material that is provided as an input but leaves the mass-balance as waste without being incorporated into the industry’s final products (e.g., construction scrap sent to landfill).

Figure 3. Availability of U.S. physical production data (in mass units such as kg) from high-quality primary sources, by subsector

Discussion

Data collection and analysis following the framework outlined in this paper is now in progress to compile a preliminary physical flows dataset for the United States. An important consideration for this work is data quality. One way we can assess the overall quality of our framework is to compare results across more than one of the proposed estimation methods (i.e., data-driven, price-driven, and input-driven) in cases where more than one data collection method is possible. Such data redundancy is present for the industries shaded green ("good data coverage") in Figure 3 because in these industries, sufficient data exists for both the data-driven and the price-driven methods to be used.

Two such subsectors are Crop Production (111CA) and Petroleum and Coal Products (324). The mass values (million metric tons) for these two subsectors are compared in Figure 4. The percent difference between mass values estimated with the two methods ranged from 17-167%. While this is a greater difference than we would prefer, the order of magnitude matches for the two completely independent methods, which represents significant progress towards a usable dataset. Work will continue to evaluate possible sources of error and opportunities for method improvement.

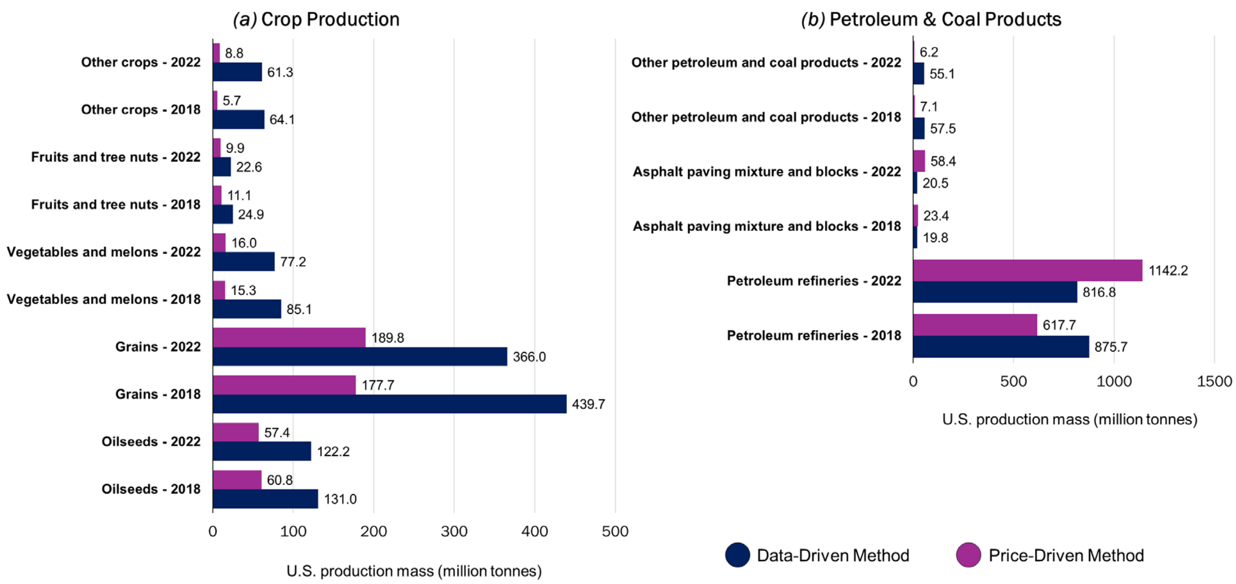

Figure 4. Comparison of physical production estimates for the data-driven and price-driven methods for two industrial subsectors: Crop Production and Petroleum and Coal Products.

A key issue for this project is that the data-driven approach is preferred (based on data quality) but data coverage is normally insufficient to apply this method to an entire subsector. To supplement this gap, the trade-based imputation (price-driven) approach is used. However, differences in estimates made using these two methods can be substantial and vary depending on the subsector. Further, different factors drive the potential inaccuracies of each of these methods. For example, inaccuracy in governmental production data (data-driven method) could be driven by incomplete reporting, whereas inaccuracy in production estimates made using the price-driven method could be driven by a wide range of factors such as price volatility, price heterogeneity between goods produced for domestic consumption vs. import or export, tariffs and other trade policies.

Conclusions & Recommendations

We have offered here a simple framework for the simultaneous physical and environmental extension of an economic input-output model that minimizes data requirements and simplifies model structure compared to a PIOT-based approach. The developed framework is now being used to develop a physical flows dataset for integration with the U.S. Department of Energy's (DOE's) EEIO for Industrial Decarbonization (EEIO-IDA) scenario modeling tool. Physical production data is needed for every goods-producing industry in the U.S. economy to enable the physical IO extension. This has proved extremely challenging due to the scarcity of high-quality raw data and poor harmonization across industry classification systems used in data reporting. Few industries have complete data coverage from primary governmental sources, meaning that an alternative price-driven (price imputation) estimation method is often the best available approach for assessing physical flows. Future work is focused on building a complete physical flows dataset for the United States for the years 2018 and 2022 and performing a comprehensive data quality assessment, highlighting areas of challenge and opportunity for mass-based EEIO.

Acknowledgments

This work is based upon funding from the Alliance for Sustainable Energy, LLC, Managing and Operating Contractor for the National Renewable Energy Laboratory for the U.S. Department of Energy.

References

1. Donati, F., Niccolson, S., de Koning, A., Daniels, B., Christis, M., Boonen, K., Geerken, T., Rodriges, J. F. D., Tukker, A., Modeling the circular economy in environmentally extended input-output: A web application, Journal of Industrial Ecology, 25, 36, 2020.
2. Yamamoto, T., Merciai, S., Mogollon, J. M., Tukker, A., The role of recycling in alleviating supply chain risk–Insights from a stock-flow perspective using a hybrid input-output database, Resources, Conservation, and Recycling, 185, 106474, 2022.
3. Vunnava, V. S. G., Singh, S., Integrated mechanistic engineering models and macroeconomic input–output approach to model physical economy for evaluating the impact of transition to a circular economy, Energy & Environmental Science 14, 5017, 2021.
4. Gause, S., Liddell, H. P. H., Dollinger, C., Steen, J., Cresko, J., Environmentally extended input-output (EEIO) modeling for industrial decarbonization opportunity analysis: a circular economy case study, in: Technology Innovation for the Circular Economy: Recycling, Remanufacturing, Design, System Analysis and Logistics, N. Nasr (Ed.), pp. 739-753, John Wiley & Sons, 2024.
5. Liddell, H. P. H., Ray, B. M., Cresko, J. W., A retrospective time-series analysis of circular economy and industrial decarbonization metrics in the United States, 1998-2022. Under review at the Journal of Advanced Manufacturing and Processing, 2025.
6. Yang, Y., Ingwersen, W. W., Hawkins, T. R., Srocka, M., and Meyer, D. E., USEEIO: A New and Transparent United Stated Environmentally-Extended Input-Output Model, Journal of Cleaner Production, 158, 308, 2017.
7. Ingwersen, W. W., Li, M., Young, B., Vendries, J., and Birney, C., USEEIO v2.0, The US Environmentally-Extended Input-Output Model v2.0, Scientific Data, 9, 194, 2022.
8. Hendrickson, C. T., Lave, L. B., and Matthews, H. S., Environmental life cycle assessment of goods and services: an input-output approach, Routledge, New York, 2006.
9. Gause, S., Liddell, H. P., Dollinger, C., Yüzügüllü, E., Steen, J., Morrissey, K., Cresko, J., Environmentally Extended Input-Output for Industrial Decarbonization Analysis (EEIO-IDA) Tool. Software tool, beta version 1.0, 2023. Available at: https://www.energy.gov/eere/iedo/articles/environmentally-extended-input-output-industrial-decarbonization-analysis-eeio
10. Wachs, L., Singh, S., A modular bottom-up approach for constructing physical input-output tables (PIOTs) based on process engineering models, Journal of Economic Structures, 7, 26, 2018.
11. Stadler, K., Wood, R., et al., EXIOBASE 3: Developing a time series of detailed environmentally extended multi-regional input-output tables, Journal of Industrial Ecology 22, 502, 2018.
12. Wood, R., Stadler, K., et al., Global Sustainability Accounting—Developing EXIOBASE for multi-regional footprint analysis, Sustainability, 7, 138, 2015.
13. Schulte, S., Jakobs, A., Pauliuk, S., Estimating the uncertainty of the greenhouse gas emission accounts in global multi-regional input-output analysis, Earth System Science Data, 16, 2669, 2024.
14. Moran, D., Wood, R., Convergence between the Eora, WIOD, EXIOBASE, and OPENEU's consumption-based carbon accounts, Economic Systems Research, 26, 245, 2014.

15. Giljum, S., Wieland, H., Lutter, S., Eisenmenger, N., Schandl, H., Owens, A., The impacts of data deviations between MRIO models on material footprints, Journal of Industrial Ecology, 23, 946, 2019.
16. Steubing, B., de Konig, A., Merciai, S., Tukker, A., How do carbon footprints from LCA and EEIOA databases compare? Journal of Industrial Ecology, 26, 1406, 2022.
17. Owen, A., Wood, R., Barrett, J., Evans, A., Explaining value chain differences in MRIO databases through structural path decomposition, Economic Systems Research, 28, 243, 2016.
18. Cresko, J., Rightor, E., et al., U.S. Department of Energy Industrial Decarbonization Roadmap, DOE Technical Report No. DOE/EE-2635, U.S. Dept. of Energy, 2022.
19. Cresko, J., Dollinger, C., et al., Transformative Pathways for U.S. Industry: Unlocking American Innovation, DOE Technical Report No. DOE/EE-2963, U.S. Dept. of Energy, 2025.
20. Ingwersen, W., USEEIO Models with Import Emission Factors for Greenhouse Gases for 2017-2022 from EXIOBASE coupled model, U.S. Environmental Protection Agency (EPA), 2025..
21. UN Statistics Division, Handbook on Supply and Use Tables and Input-Output tables with Extensions and Applications, United Nations National Accounts, 2018.

About the Authors

Heather Liddell is an Assistant Professor at Purdue University with a joint appointment in Mechanical Engineering and Environmental & Ecological Engineering. Her research focuses on technology and systems strategies to improve resource utilization and reduce adverse environmental impacts of manufacturing. Liddell holds BS and MS degrees in mechanical engineering and a PhD in materials science from the University of Rochester.

Beth Kelley is a Graduate Research Assistant at Purdue University. Her research focuses on quantifying mass flows in the U.S. manufacturing sector and developing a database of mass-based emission factors (kg CO_2e/kg) for U.S. goods production. She completed her BS in May 2023 and expects to graduate with her MS in May 2025 from the Environmental & Ecological Engineering program at Purdue University.

Liz Wachs is a Researcher at the National Renewable Energy Laboratory, focusing on industrial decision making for sustainability. Liz holds a BA in ancient history and classical civilization from the University of Texas at Austin, an MA in Desert Studies from Ben Gurion University of the Negev, a BS in chemical engineering and PhD in agricultural and biological engineering from Purdue University.

Alberta Carpenter is a Distinguished Member of Research Staff and a Senior Researcher at the National Renewable Energy Laboratory and leads NREL's strategic analysis work focusing on sustainable manufacturing to include industrial decarbonization, circular economy, environmental and justice impacts.

Joe Cresko is the Chief Engineer for DOE's Industrial Efficiency and Decarbonization Office, where he leads the Strategic Analysis Team. This team helps to assess the life cycle and cross-sectoral impacts of new manufacturing advances. Cresko holds a BS in chemical engineering from Bucknell University and an MS in engineering science and mechanics from Penn State University.