

FINAL REPORT

Strategy and Roadmap for Advancing Life Cycle Assessment Methodologies

Wesley W. Ingwersen
Jane Bare
U.S. Environmental Protection Agency

Kelly A. Scanlon
IPC
Formerly, Office of the Assistant Secretary of Defense for Energy, Installations, and Environment

Michael S. Bruckner
Noblis

June 2023

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 23-06-2023		2. REPORT TYPE SERDP Final Report		3. DATES COVERED (From - To) 12-01-2018 - 31-12-2022	
4. TITLE AND SUBTITLE Strategy and Roadmap for Advancing Life Cycle Assessment Methodologies: A Proposal for a Special Study				5a. CONTRACT NUMBER IA RW-096-95925701-1	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Wesley Ingwersen (1), Jane Bare (1), Kelly Scanlon (2), Michael Bruckner (3)				5d. PROJECT NUMBER WP-2757	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) 1. U.S. Environmental Protection Agency, Washington, DC 2. Department of Defense. Arlington, VA 3. Noblis. Reston, VA				8. PERFORMING ORGANIZATION REPORT WP-2757	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Department of Defense Strategic Environmental Research and Development Program 4800 Mark Center Drive, Suite 16F16 Alexandria, VA 22350-3605				10. SPONSOR/MONITOR'S ACRONYM(S) SERDP	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) WP-2757	
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited. This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. This is a work of the U.S. Government and is not subject to copyright protection in the United States.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The objective of this study was to make life cycle assessment (LCA) advancements that could be integrated to make the Department of Defense's (DoD's) Defense Input Output (DIO) model and associated methods more robust. In Task 1, three software prototypes were developed to automate product system model construction. In Task 2, new life cycle impact assessment characterization factors were created for eutrophication effects and a new software tool to harmonize impact assessment with life cycle data, the LCIA Formatter, was created. In Task 3, a new version of the DIO model was built that uses the USEEIO build architecture and integrates select elements from the former DIO and current USEEIO models.					
15. SUBJECT TERMS DIO, USEEIO, life cycle impact assessment, LCIA Formatter, input-output, eutrophication factors, product system model assembly					
16. SECURITY CLASSIFICATION OF: UU			17. LIMITATION OF ABSTRACT UNCLASS	18. NUMBER OF PAGES 92	19a. NAME OF RESPONSIBLE PERSON Wesley Ingwersen
a. REPORT UU	b. ABSTRACT SAR	c. THIS PAGE SAR			19b. TELEPHONE NUMBER (include area code) 870-493-6885

SERDP FINAL REPORT

Project: WP-2757

TABLE OF CONTENTS

	Page
ABSTRACT	IX
EXECUTIVE SUMMARY	ES1
1 TASK 1 - EVALUATING LCA DATA STRUCTURES AND PLATFORMS TO ENABLE AN LCA DATA PIPELINE FOR DOD	1
1.1 INTRODUCTION	1
1.2 METHODS	2
1.3 RESULTS	4
1.3.1 Presentation of Prototypes	5
1.3.2 Prototype Testing	9
1.3.3 <i>autoprox</i> – An Additional Prototype for Direct openLCA Software Integration.....	14
1.4 DISCUSSION	15
1.4.1 Potential and Limitations	17
1.5 CONCLUSION.....	17
1.6 CODE AVAILABILITY	17
1.7 REFERENCES	18
2 TASK 2.1 - IDENTIFY AND IMPROVE AN EXISTING IMPACT ASSESSMENT METHOD WITH RELEVANCE TO DOD	21
2.1 INTRODUCTION	21
2.2 METHODS	22
2.2.1 Nutrient Framework	22
2.2.2 Fate Calculations.....	24
2.2.3 Aggregation.....	25
2.2.4 Overall Calculations.....	28
2.3 RESULTS	30
2.3.1 Freshwater Fate Factors and the Great Lakes	33
2.3.2 Large Marine Ecosystems (LMEs)	34
2.3.3 Application Case: N and P Inventories	35
2.4 DISCUSSION	38
2.5 CONCLUSIONS.....	38
2.6 REFERENCES	39
3 TASK 2.2 – ADDRESS THE DISCONNECT BETWEEN DATA DEVELOPED IN IMPACT ASSESSMENT METHODS AND DATA FOR LIFE CYCLE INVENTORIES 42	42
3.1 SUMMARY	42
3.2 STATEMENT OF NEED	42

TABLE OF CONTENTS (Continued)

	Page
3.3	STRUCTURE 43
3.4	AVAILABLE METHODS 44
3.4.1	TRACI2.1..... 44
3.4.2	ReCiPe2016 44
3.4.3	ImpactWorld+ 44
3.4.4	FEDEFL Inventory Methods 45
3.4.5	Valuation..... 45
3.5	APPLICATIONS 45
3.6	REFERENCES 45
4	TASK 3.1 – STRATEGIC ROADMAP: USEEIO AND DIO ALIGNMENT 48
4.1	INTRODUCTION 48
4.1.1	Input-Output (IO) Modeling 48
4.1.2	DoD Needs..... 49
4.2	CURRENT STATE 51
4.2.1	Defense Input-Output (DIO) Model 51
4.2.2	U.S. Environmentally-Extended Input-Output Model (USEEIO) 52
4.2.3	Comparison Table..... 53
4.2.4	Barriers and Issues 54
4.3	STRATEGY..... 54
4.3.1	Goals 54
4.4	ROADMAP..... 55
4.4.1	Short-term 55
4.4.2	Medium-term 55
4.4.3	Long-term 56
4.5	REFERENCES 57
5	TASK 3.2 USEEIO-DIO MODEL 58
5.1	OBJECTIVE 58
5.2	BACKGROUND 58
5.3	METHODS 60
5.3.1	Process Extensions..... 61
5.3.2	Indicators and Scoring Factors..... 63
5.3.3	Air Emissions Model 68
5.4	RESULTS AND DISCUSSION 68
5.4.1	Data Records..... 68
5.4.2	Technical Validation..... 69
5.4.3	Code Availability..... 69
5.5	REFERENCES 69
6	CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH 71

LIST OF FIGURES

	Page
Figure 1. Schematic Drawing of One Landing Gear Product System	3
Figure 2. Semantic Network of Product Relations in <i>pslink</i>	6
Figure 3. Calculating the Traversal Factor in <i>pslink</i>	7
Figure 4. Exchanges and Dependency Relationships in a Product System Model.....	8
Figure 5. PSM of Landing Gear Connected to USEEIO in openLCA	11
Figure 6. LCIA Results Computed From the Linked Landing Gear Model in openLCA	11
Figure 7. A Portion of a Model's Internal Nodes (foreground) Documented in a Spreadsheet.....	12
Figure 8. A Portion of a Model's Leaf Nodes (background) Documented in a Spreadsheet ...	13
Figure 9. Matching the CDD MRF Dataset Using <i>perdu</i>	14
Figure 10. The Four Steps for Developing a PSM From a Product Dataset Identified in the Study, and the Contributions of Each Prototype to Each Step	15
Figure 11. Calculation of Characterization Factors for Eutrophication Potential	22
Figure 12. Schematic of the Characterization Factor Aggregation Process	25
Figure 13. Spatial Intersect of Native-Scale Characterization Factors	27
Figure 14. Freshwater Eutrophication Grid for Illinois	32
Figure 15. Maps of Characterization Factors for Freshwater Eutrophication to Freshwater (Top Row) and Freshwater Eutrophication to Soil (Bottom Row)	33
Figure 16. Maps of Characterization Factors for Marine Eutrophication Emissions to Freshwater, LME, and Soil (Top Row) and Air (NH ₃ and NO _x ; Middle and Bottom Rows).....	35
Figure 17. Case Application for P at the State Level: Agricultural Fertilizer (Top Row), Urban Fertilizer (Middle), and Human Waste (Bottom)	36
Figure 18. Application Case for N at the State Level: Agricultural Fertilizer (Top Row), Urban Fertilizer (Middle), and Human N Waste (Bottom).	37
Figure 19. Components of EEIO Modeling.....	49
Figure 20. Data Components of DIO.....	52
Figure 21. Simplified Structure of USEEIO Model Assembly, Validation, and Calculation	53

LIST OF TABLES

	Page
Table 1. Software Requirements	4
Table 2. Summary Characteristics of Three Initial Prototypes	9
Table 3. <i>pslink</i> Match Statistics for F-18 LG ANON Dataset	10
Table 4. Summary of Characterization Factors Calculated in this Work	29
Table 5. Eutrophication Impact Categories, Likelihoods, and Geographic Aggregation Targets, With Associated Characterization Factors Summarized Across the Interdecile Range (IDR)	30
Table 6. Statistical Variation Across the CONUS Aggregated Characterization Factors, Using Agricultural, Nonagricultural, and General Likelihoods for State-Level and Country- Level Geographic Targets, as Appropriate	32
Table 7. Impact Methods Processed by lciafmt (Young et al. 2021d, 2021c, 2021b, 2021a) ..	44
Table 8. Inputs Into DIO Model A and B Matrices	61
Table 9. Data Sources for the Energy Use and Transportation Extensions to the <i>A</i> Matrix	62
Table 10. Midpoint Impact Categories	64
Table 11. Endpoint Impact Categories	64
Table 12. Valuation Categories, Values, and Methods	67
Table 13. Data Records	68

ACRONYMS AND ABBREVIATIONS

AAF	Adaptive Acquisition Framework
AFFF	Aqueous Film-Forming Foam
AoA	analysis of alternatives
BLS	Bureau of Labor Statistics
BOM	Bill of Materials
CBA	cost benefit analysis
CEA	cost-effectiveness analysis
CF	characterization factor
CONUS	continental United States
CREP	Conservation Reserve Enhancement Program
CRP	Conservation Reserve Program
DALY	disability-adjusted life year
DoD	Department of Defense
DoDI	Department of Defense Instruction
DIO	Defense Input-Output
EEIO	environmentally-extended input-output
eGRID	Emissions & Generation Resource Integrated Database
EO	Executive Order
EPA	U.S. Environmental Protection Agency
ESOH	Environment, Safety, and Occupational Health
ESTCP	Environmental Security Technology Certification Program
FEDEFL	Federal LCA Commons Elementary Flow List
FLCAC	Federal Life Cycle Assessment Commons
FLOWSA	Flow Sector Attribution
FWP	Farmable Wetlands Program
FY	Fiscal Year
GDP	Gross Domestic Product
GEP	Gross Economic Product
GHP	Gross Household Production
GPS	global positioning system
GREET	Greenhouse gases, Regulated Emissions, and Energy use in Transportation
ICD	International Classification of Diseases
IEA	International Energy Agency
IO	input-output
ISO	International Organization for Standardization
KPP	key performance parameter

LCA	Life Cycle Assessment
LCC	life cycle costing
LCI	Life Cycle Inventory
LCIA	Life Cycle Impact Assessment
m ²	square meter(s)
MCDM	multi-criteria decision-making
MILCON	Military Construction
MJ	megajoule(s)
NAVSEA	Naval Sea Systems Command
NDAA	National Defense Authorization Act
NEI	National Emissions Inventory
O&M	operations and maintenance
OH&S	occupational health and safety
PDF-m2-yr	potentially disappeared fraction of species on a meter squared during a year
PESHE	Programmatic Environment, Safety, and Occupational Health Evaluation
PFAS	per- and polyfluoroalkyl substances
PSM	Product System Model
RDT&E	research, development, test, and evaluation
SAG	Sustainability Analysis Guidance
SC-GHG	social cost of greenhouse gases
SCC	social cost of carbon
SERDP	Strategic Environmental Research and Development Program
StEWI	Standardized Emission and Waste Inventories
t-km	tonne-kilometre
TOC	total ownership cost
TRACI	Tool for Reduction and Assessment of Chemicals and Other Impacts
TRI	Toxics Release Inventory
UAC	urban areas and urbanized clusters
USD	U.S. dollar(s)
USEEIO	U.S. Environmentally-Extended Input-Output model
USN	U.S. Navy
VSL	value of a statistical life
WHO	World Health Organization
yr	year

KEYWORDS

life cycle impact assessment, eutrophication factors, LCIA Formatter, USEEIO, Defense Input-Output, LCA data pipeline, sustainability assessment

ACKNOWLEDGEMENTS

In addition to the principal investigators, numerous scientists and engineers contributed technical work to one or more tasks in this project. They include: Andrew Henderson, Ben Young, Andrew Beck, Ben Morelli, and Sarah Cashman from Eastern Research Group; Michael Srocka from GreenDelta; Brandon Kuczenski and Chris Mutel from GDIT; Heather Golden from U.S. EPA; Shane Mead, Clark Bee, Heather DiCaprio, and Emma Williams from Noblis; Shannon Lloyd from Concordia University. Chrys Starr (NAVAIR) provided data used in Task 1. Bill Michaud (GDIT), Sarah Cashman (Eastern Research Group), and Braxton Lewis (Noblis) provided contract management support. Christina Gannett (Noblis) helped edit the report. The U.S. EPA Sustainable and Healthy Communities National Research Program and the Office of the Assistant Secretary of Defense Energy, Installations, and Environment provided additional research management for this project, as well as funding for related projects without which this project would not have been possible. The U.S. EPA National Risk Management Research Laboratory and Center for Environmental Solutions and Emergency Response provided administrative and quality assurance support.

ABSTRACT

INTRODUCTION AND OBJECTIVES

As part of defense acquisitions, Life Cycle Assessment (LCA) is a tool to assist in designing more sustainable systems—those that use fewer resources over the life cycle, have fewer impacts on human health and the environment, and thus have a lower total ownership cost. The objective of this study was to make LCA advancements that could be integrated to make the Department of Defense’s (DoD’s) Defense Input Output (DIO) model and associated methods more robust.

TECHNICAL APPROACH

Advancements in LCA data, impact assessment, and input-output LCA were completed as part of three tasks. In **Task 1**, the research team worked with the United States Navy and IPC to get bills of materials (BOM) for the F-18 landing gear and electronic printed circuit board, respectively, to develop a prototype software to construct product system models (PSMs). In **Task 2**, an existing eutrophication life cycle impact assessment (LCIA) method was selected. The research team updated the method with increased resolution of U.S. regional specific fate and transport factors. Also, an open-source software was created to address the data disconnect between impact assessment characterization factors (CFs) and data used in life cycle inventory and to create standardized sets of CFs. In **Task 3**, DoD’s DIO and the U.S. Environmental Protection Agency’s (EPA’s) Environmentally-Extended Input-Output (USEEIO) models were evaluated to determine where they shared elements or differed, and to select desirable elements from each model and associated framework to develop DIO v2.0. The research team developed a strategic plan for aligning both models and developed DIO v2.0 combining elements of USEEIO and DIO. The new model will be documented in a peer-reviewed journal.

RESULTS

In **Task 1**, the research team developed three complementary software prototypes, and each took a distinct approach to creating a complete PSM through linkages to background datasets. Each prototype addresses different steps to automate the translation of the BOMs into unit processes; however, manual work is still required in each case. In **Task 2**, the research team generated a new set of freshwater and marine eutrophication factors specific to counties and states in the contiguous United States. The team also created the LCIA Formatter—a Python package that produces standardized datasets of CFs optionally using the Federal LCA Commons (FLCAC) Elementary Flow List (FEDEFL) as a standard flow list. In **Task 3**, the research team aligned the DoD DIO with the EPA USEEIO model and increased the resolution of air emissions in USEEIO. Also, a strategic plan for USEEIO and DIO was developed that included recommendations for short-, medium-, and long-term steps to improve EEIO models for use by DoD.

BENEFITS

Task 1 led to a finding that greater investment of time into refining or improving the prototypes, or into further development of methods to tackle the steps delineated by the study, is likely to yield very useful tools that automate data collection processes important to conducting LCA studies. **Task 2** demonstrated how to create more spatially-explicit CFs for impacts of concern that are

spatially-sensitive (like eutrophication) for areas in which the DoD operates around the globe. The LCIA methods generated by the LCIA Formatter for use with the FEDEFL can be hosted publicly on the FLCAC website for use by LCA practitioners and researchers and can support LCAs performed by many parties, including FLCAC member agencies such as EPA, Department of Energy, Department of Agriculture, and the DoD. As demonstrated in this task, data from across Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program (ESTCP) Program Areas could improve existing environmental models and impact assessment methods. **Task 3 led to the conclusion that USEEIO and DIO model alignment will lower DoD's operations and maintenance costs to maintain a large background dataset that includes all indirect activities in the economy and corresponding data on extractive, industrial, agricultural, transport, and disposal processes that captures their material and energy requirements and environmental emissions and wastes. As EPA updates USEEIO, the DIO model will be updated with the corresponding data, giving DoD access to the latest authoritative background LCA modeling information that allows DoD to concentrate resources on development of defense specific process datasets currently lacking to complete LCA studies. The output of this research supports DoD's efforts to meet executive order and statutory sustainability requirements as part of defense acquisitions.**

EXECUTIVE SUMMARY

INTRODUCTION

The Strategic Environmental Research and Development Program (SERDP) Weapons Systems and Platforms Program Area supports the development and demonstration of innovative environmental methods and technologies. These technologies enable the Department of Defense (DoD) to reduce or eliminate hazardous materials and associated impacts across the life cycle of the system: research, development, test, and evaluation (RDT&E); production; fielding or deployment; sustainment; and disposal. Promoting cost-effective, sustainable solutions requires life cycle-based assessments that evaluate potential environmental impacts and costs. Currently, DoD’s “Sustainability Analysis Guidance: Integrating Sustainability into Acquisition Using Life Cycle Assessment” (SAG) provides the defense community with a methodology, standardizing the approach to conducting the assessment and reporting findings that can be used to refine concepts and promote safer, more sustainable solutions.

For DoD, Life Cycle Assessment (LCA) is a tool to assist in designing more sustainable¹ systems — those that use fewer resources over the life cycle, have fewer impacts on human health and the environment, and thus have a lower total ownership cost (TOC; DoD, 2022a). LCA is considered part of the “sustainability analysis” systems engineering technique and incorporated within the Defense Acquisition System (DoD, 2022a) and used to “minimize negative impacts on resource availability, climate change, human health, and ecosystem quality” (DoD, 2020a). As part of sustainability analysis, LCA supports decision making throughout acquisition activities, including analysis of alternatives, trade-space analysis, business case analysis, preliminary design, and supportability analysis (DoD, 2022a).

The SAG (2022a) describes a “hybrid” approach to LCA where users work with a common background LCA model, the Defense Input-Output (DIO) model. DIO v1.0 was created by DoD and is an environmentally-extended input-output (EEIO) model type that can estimate impacts of U.S. goods and services in aggregate product categories (~380). DIO also contains additional extensions for international and U.S.-region energy and transportation process. A custom set of impact assessment methods is used by DIO to translate chemical release and resource use across the life cycle into potential environmental impacts.

OBJECTIVES

This special study was designed to make advancements to LCA that can potentially be integrated to make the DIO and associated methods more robust for DoD LCA. As part of this special study, research in three areas was undertaken on the following topics:

(1) LCA Data: Automate data collection and assembly. Collecting and managing data relevant to the life cycle of a complex system requires strategic thinking to ensure structured, harmonized, and useable data and models. Assembling full life cycle models (i.e., product systems) quickly and

¹ “Sustainability” differs from “sustainment” in that it relates to the use of resources, and the associated impacts and costs over the system’s life cycle. In contrast, sustainment is concerned with the end user’s ability to operate and maintain a system once it is in inventory and deployed. Sustainability requires consideration of all phases of a system’s life cycle, including system design, development, production, operations, sustainment, and disposal (DoD, 2020a).

through automation will create a data pipeline for DoD that will streamline model building for complex products.

(2) Impact Assessment: Advance accuracy of a selected impact category method and make methods more interoperable with life cycle inventory data. Life cycle impact assessment (LCIA) methods produce characterization factors (CFs) that link data on emissions and resource use over the life cycle of a product to environmental, human health, climate change, and resource depletion threats. Impact assessment methods, particularly those of relevance to DoD, need to be advanced as well as better linked to life cycle inventory (LCI) data.

(3) Input-output LCA Models: Integrate the disparate EPA and DoD EEIO models. The U.S. Environmental Protection Agency (EPA) and DoD have each constructed EEIO models for the U.S. economy that can be used as a standard and complete background life cycle model. A model built on the best aspects of each existing model with new enhancements can provide a robust background LCA model to use for high-level analysis or in the background of detailed product or technology LCA studies as part of Defense Acquisitions.

TASK 1 - EVALUATING LCA DATA STRUCTURES AND PLATFORMS TO ENABLE AN LCA DATA PIPELINE FOR DOD

Technical Approach

The software used to construct and manage product system models (PSMs) for LCA is receiving increasing attention. In this task, three software developers with experience in LCA were invited to collaborate on the problem of generating an operable PSM from a foreground specification. The team worked with the U.S. Navy (USN) to get a bill of materials (BOM) for the landing gear of the F-18 to use for the main foreground specification; a second BOM was obtained on an electronic printed circuit board from IPC, a global trade association representing electronics manufacturers, to use as a case study. The software developers were given the foreground specifications as a set of spreadsheets and were asked to develop and test prototype software under a limited time period. The resulting prototypes were combined and compared.

Results

The developers automated the translation of the BOMs into a set of unit processes that became the foreground model. Each submission took a distinct approach to the problem of creating a complete PSM through linkages to background datasets. The first approach used semantic similarity relations to identify best-fit background datasets. The second approach focused on producing a flexible description of the model structure that removed redundancy and permitted aggregation. The third approach provided an interactive web application for matching product components to standardized classification systems to facilitate characterization and linking.

Discussion

An analysis of these approaches identified four common steps in the broader problem of automating PSM assembly: (1) specifying the model structure; (2) specifying flow properties; (3) linking flows to background datasets; and (4) serializing the linked model in a format that could be used by LCA software. The three prototypes are complementary in that they address different

steps and demonstrate alternative approaches. None of the approaches realized the goal of full automation, especially in the flow property description step.

Implications for Future Research and Benefits to DoD

Greater investment of time into refining or improving these prototypes, or into further development of methods to tackle one of the four steps delineated by this study, is likely to yield very useful tools that automate data collection processes important to conducting LCA studies. These prototypes are not yet integrated into the LCA software tools already used in the community by practitioners. However, they share assets that would facilitate their incorporation. These assets include: (1) each prototype is already packaged in a way that facilitates automated install or embedding in other tools; (2) the prototypes are written in a language (Python) that can already be run by tools such as openLCA or Brightway via their existing interfaces; (3) the tools export PSMs in standard exchange formats, like JSON-LD and ILCD, that are readable by LCA software tools; (4) the MIT open-source licenses permit reuse, embedding, and modification without licensing fees or copyright infringement; and (5) management of the source code for the prototypes is in a version control system that supports collaboration and issue tracking (github).

The incorporation of one or more of the prototypes into an LCA tool would automate PSM assembly from a BOM accessible to a larger body of practitioners. A prototype could be customized to link BOM-based DoD installation or equipment data with the DIO model. As the DoD SAG (2022a) recommends hybrid modeling, such a tool could help practitioners automatically link their primary data to the DIO model. The team recommends continued support to translate prototype(s) software to automated modeling that will enable assessment of many complex defense-related products to ensure compliance with the SAG (2022) and the goals of the DoD's (2018) Digital Engineering Strategy.

TASK 2 – IMPROVING LIFE CYCLE IMPACT ASSESSMENT METHODS AND INTEROPERABILITY

Technical Approach

This task was divided into two subtasks. The first subtask (Task 2.1) was to identify and improve an existing impact assessment method with relevance to DoD. Eutrophication, or the effect of excess nutrients in freshwater and marine ecosystems, was selected as an area where U.S. regional specific factors did not exist, and where models for fate and transport were not included in the method. Existing models and methods were reviewed along with available fate and transport models for nutrients to identify the best model to use to update eutrophication factors.

The second subtask (Task 2.2) addressed the disconnect between data developed in impact assessment methods and data for LCIs. The team evaluated ways to make impact assessment methods more readily-accessible and decided on creating lightweight, open-source software to perform this function.

Results

The team developed a new set of freshwater and marine eutrophication CFs specific to counties and states in the contiguous United States. The marine factors characterize nitrogen releases to

freshwater and marine environments, and ammonia and nitrogen oxide releases to air. The freshwater factors characterize the impacts of phosphorus releases to freshwater or soil in agricultural and non-agricultural land uses. The factors were evaluated by applying them to nitrogen and phosphorus inventory data with full contiguous-U.S. coverage.

To address the data disconnect between LCIA CFs and data used in LCI, the team developed the LCIA Formatter tool. The LCIA Formatter is a Python package that produces standardized datasets of CFs optionally using the Federal LCA Commons (FLCAC) Elementary Flow List (FEDEFL) as a standard flow list. The LCIA Formatter produced standard sets of impact assessment methods including TRACI v2.1, Impact World +, and ReCIpe 2016.

Discussion

The impacts of the nutrient losses as estimated with the new CFs illustrate the spatial variation and patterns in potential eutrophication impact in the United States. The importance of variation in geographic location is demonstrated using nutrient data likelihood categories of agricultural (rural fertilizer), non-agricultural (urban fertilizer), and general (human waste processing). Proximity to large bodies of water, as well as individual hydraulic residence times, were shown to affect the comparative values of CFs across the United States.

The LCIA methods generated by the LCIA formatter for use with the FEDEFL can be hosted publicly on the FLCAC for use by LCA practitioners and researchers. These methods support LCAs performed by many parties, including FLCAC member agencies such as EPA, Department of Energy, Department of Agriculture, and DoD. The system was built to be flexible enough to support creating outputs for LCIA spatially-explicit CFs as those become more common.

Implications for Future Research and Benefits

Impact assessment methods need to be advanced, and they need to be better linked to LCI data. For example, eutrophication is one of many impact categories with relevance to DoD training and operations in marine environments, where enrichment may promote corrosion through excessive biofilm growth on DoD ships and infrastructure. In the current task, freshwater and marine eutrophication CFs are spatially differentiated to improve the scientific robustness of the impact category. Information on the sources of nutrients is available for some types of impacts, e.g., agricultural production, but uncertainty in data sources and data gaps prevent useful assessments and solutions. Recommendations that would require further research include: (1) refining the spatial and temporal resolutions of the fate and transport models; (2) further developing the residence time estimates for freshwater and marine systems; (3) integrating soil data, such as soil type, to better estimate application and run-off of land-applied nutrients; and (4) using hydrodynamic models to expand the marine coverage.

In addition to the eutrophication impact category, the team recommends continued support to better incorporate DoD-generated data and information on emissions, fate, and transport to create more spatially-explicit CFs for impacts of concern that are spatially-sensitive (like eutrophication) for areas in which the DoD operates around the globe. Related, SERDP and ESTCP investments could continue to advance research to incorporate primary ecological toxicity data to increase scope of LCIA methods to incorporate impacts of potential chemicals of concern, such as the 2022

Per- and Polyfluoroalkyl Substances (PFAS)-Free Foam Sustainability Analysis, which incorporated results from across the SERDP and ESTCP Environmental Restoration Program Area to increase the resolution of the USEtox database and create new product-specific CFs. Such investment would allow DoD to meet future requirements like the Fiscal Year (FY) 2021 National Defense Authorization Act (NDAA) requirement to promote and advance the research and development of PFAS-free alternatives to Aqueous Film-Forming Foam (AFFF) and prioritize green and sustainable chemicals that do not pose a threat to public health or the environment. DoD data may be extended to advance impact assessment methods and associated CFs for other relevant impact categories. Data on energy and water use (e.g., from installation resilience studies), on air and noise emissions (e.g., from waste reduction studies), on formulation and processing of materials (e.g., from sustainable energetics studies), and on environmental health and human health (e.g., from PFAS studies) would have utility and can be used by the FLCAC partners, including DoD, to improve existing environmental models and impact assessment methods.

TASK 3 - IMPROVING DOD AND EPA INPUT-OUTPUT MODELS

Technical Approach

The EPA developed a U.S. economy-wide EEIO (USEEIO) life cycle model that provides comparable capability to the DoD DIO v1.0 model. USEEIO is arguably the most comprehensive, transparent, and reproducible national EEIO model to date (Yang et al, 2017). Task 3 is comprised of four steps: (1) evaluate the models side-by-side to determine where they shared elements or differed; (2) identify desirable elements from each model and associated framework from which to develop a DIO v2.0; (3) develop a strategic plan for aligning the USEEIO and DIO models; and (4) develop DIO v2.0 that combines elements of USEEIO and DIO and to document the new model in a peer-reviewed journal.

Steps 1–3 were performed as part of the first subtask (Task 3.1) “Strategic Roadmap: USEEIO and DIO Alignment.” Step 4 was performed as part of the second and final subtask (Task 3.2) “USEEIO-DIO Model.”

Results

A strategic plan for USEEIO and DIO was developed that included a comparison of DIO and USEEIO elements. The strategic plan calls for alignment of the core EEIO economic and environmental components, use of the USEEIO modeling framework for model building after expanding it to enable integrated hybrid models, expansion of DoD-specific components for DoD uses, and numerous other recommendations. The recommendations were divided into short-, medium-, and long-term steps.

The team built DIO v2.0. This integrated hybrid LCA model includes both the EEIO core along with the energy and transportation ‘process extensions’ that were a part of DIO v1.0. The EEIO portion of the model incorporates economic transactions between sectors and environmental flow components of the EPA USEEIO v2.0 model. The model for criteria and hazardous air emissions by industry sector used in USEEIO v2.0 was improved to include increased resolution of air emissions, expanding on a less-refined approach used in DIO v1.0. The environmental flows were made consistent with the FLCAC elementary flow list. The same customized set of impact methods

used by DIO v1.0 were selected but built using updated data with the LCIA Formatter tool. These impact methods include valuation factors that put impacts in dollar terms that include the social cost of carbon. The model was assembled using the USEEIO modeling framework in which USEEIO models are maintained and regularly updated to allow the DIO to be easily updated over time and to be able to incorporate future advancements in the USEEIO model. Documentation for the DIO v2.0 model was written with the intent of publishing in the public domain.

Discussion

The DIO v2.0 model implements all the short-term steps identified in the strategic plan and many of the medium- and long-term steps. The model core of DIO and USEEIO was unified, with existing DoD specific enhancements still available in DIO. There remains a need to accomplish additional alignment and enhancements steps, including adding occupational health indicators and to align plans for new DoD applications or end-user software to incorporate DIO v2.0 updates.

Application of DIO v2.0 and results of this research will allow DoD to better meet new sustainability requirements as part of defense acquisitions. This potentially affects decision making over a large share of Federal procurement. FY 2023 acquisition funding requested by DoD totalled \$276.0 billion, which includes funding of \$145.9 billion for Procurement and \$130.1 billion for RDT&E (DoD, 2022b). The scale of this potential impact is further put into perspective when considering that over half of the entire discretionary Federal budget is spent by DoD (USDT, 2022). This aligns with Executive Order (EO) 14057, *Catalyzing Clean Energy Industries and Jobs Through Federal Sustainability*, Section 102 requirement for the Federal Government “to use its scale and procurement power to achieve... net-zero emissions from Federal procurement.”

As this research supports incorporation of the social cost of greenhouse gases (SC-GHG) in acquisitions, this work directly supports the EO 14030, *Climate-Related Financial Risk*, requirement to consider SC-GHG in procurement decisions. Also, this research directly supports the FY 2021 NDAA Section 263 requirement for Agencies “to incorporate sustainable chemistry into existing research, development, demonstration [Programs].”

Implications for Future Research and Benefits

By aligning the DoD DIO with EPA’s USEEIO model and building on EPA’s model, this will lower DoD’s operations and maintenance (O&M) costs to maintain a large background dataset. This background dataset includes all indirect activities in the economy and corresponding data on extractive, industrial, agricultural, transport, and disposal processes that captures their material and energy requirements and environmental emissions and wastes. EPA will continue to update USEEIO, which gives DoD access to the latest authoritative background LCA modeling information and allows DoD to concentrate resources on development of defense-specific process datasets currently lacking to complete LCA studies. Examples of such datasets DoD could focus on include development of sustainable aviation fuel, propulsion systems, support infrastructure for decarbonization, materials needed for Military Construction (MILCON) activities, surface engineering, and weapons systems sustainment. The team recommends that DoD establish a repository to contain DIO, as its authoritative LCA modeling support tool, and continue to regularly engage with EPA. Likewise, DoD participation in the FLCAC may spur additional synergies across Federal agencies.

This task also allowed for an update to the air emissions inventory created for the DIO and USEEIO models. This advancement will serve as a national resource and increases resolution of impact assessment inferences that can be made by using both DIO and USEEIO. Future collaborations by DoD and EPA may yield additional advancements available for application across the U.S. economy.

As DoD's Sustainability Analysis Guide (SAG) evolves, it should reflect the updated nature of the DIO model. To increase adoption of this research in support of acquisition and sustainability requirements, SERDP and ESTCP should provide additional guidance to their principal investigators for how to deploy this research to estimate impacts, costs, and benefits of their own research.

The team recommends continued support for short-, mid-, and long-term steps described in the Strategic Roadmap (Chapter 4). The ideas expressed provide feedstock for several collaborative research efforts that will build on an already existing strong relationship between SERDP and ESTCP government partners for the benefit of all agencies.

REFERENCES

- United States Department of Defense (DoD). (2018). Digital Engineering Strategy. Office of the Deputy Assistant Secretary of Defense for Systems Engineering. Washington, D.C.
- United States Department of Defense. (2020a). Sustainability Analysis Guidance: Integrating Sustainability into Acquisition Using Life Cycle Assessment. Version 7.0.
- United States Department of Defense (DoD). (2022a). Systems Engineering Guidebook. Office of the Deputy Director for Engineering, Office of the Under Secretary for Research and Engineering.
- United States Department of Defense (DoD). (2022b). Program Acquisition Cost By Weapon System. United States Department of Defense Fiscal Year 2023 Budget Request. Office of the Under Secretary of Defense (Comptroller)/Chief Financial Officer.
- United States Department of the Treasury (USDT). (2022). How much has the U.S. government spent this year? *FiscalData.Treasury.gov*

1 TASK 1 - EVALUATING LCA DATA STRUCTURES AND PLATFORMS TO ENABLE AN LCA DATA PIPELINE FOR DOD

This chapter contains the content of the following published journal article that was written as part of this project, with minor modifications:

Kuczenski B, Mutel C, Srocka M, et al (2021) Prototypes for automating product system model assembly. The International Journal of Life Cycle Assessment. <https://doi.org/10.1007/s11367-021-01870-9>

1.1 INTRODUCTION

Modeling the potential direct and indirect environmental impacts of chemicals, materials, and the products is the aim of LCA studies. Including all indirect activities in the economy requires significant data on extractive, industrial, agricultural, transport, and disposal processes that captures their material and energy requirements and their environmental emissions and wastes. Practitioners must rely on existing datasets to fulfill these “background” data requirements. Furthermore, the data must all be interoperable so that they can be used together and be based on coherent modeling assumptions. Existing life cycle inventory data are often not interoperable (Ingwersen 2015; Suh et al. 2016) but achieving this has been acknowledged as essential for broader LCA data sharing and usage of data from multiple sources (Ingwersen et al. 2015; Canals et al. 2016).

Identification, acquisition, and integration of data continues to pose a significant challenge to those who create models for LCA studies, particularly for studies that do not rely on a single life cycle inventory database for background data. This task involves acquisition of data from various sources, structuring the data for use in an LCA as a network of processes, and linking the processes together to build a product system model (PSM). Data discovery applications and data portals (e.g., Nexus (GreenDelta 2019), the Federal LCA Commons (FLCAC) (USDA 2019), the Global LCA Data Access Initiative (UNEP/SETAC Life Cycle Initiative, 2019)) that provide structured life cycle inventory data from multiple sources are in their early stages. Data mining approaches are also in early development to create more data for life cycle assessment from public sources (Cashman et al. 2016; Citroth and Srocka 2017). Building variations of PSMs using a single LCI database has also been automated in the Wurst software for ecoinvent (Mutel 2017). While there are great challenges to finding and or developing new LCI data for life cycle assessments, the greater challenge in the field right now might be linking LCI data together from different sources and assembling them in a structured PSM that can be reviewed, revised, and reused by others (Kuczenski et al. 2018). Because PSMs may consist of hundreds or thousands of processes combined from various sources, it is necessary to be able to describe those models precisely and in a transparent fashion.

Often, the description of a PSM can be described in two parts: the description of the foreground as a network of unit processes, nominally a tree whose root is the functional unit of the model; and the linking of terminal nodes or “leaves” in that network to process models found in background databases (Kuczenski 2019). While this description may not hold for some activities, particularly in chemical production (which may involve loops in the foreground system), it is suitable to describe the foreground of most product LCAs. The linking of LCI processes from various data

sources could be accomplished using various methods. One method that has been proposed is through the creation of “bridge” processes, which prevent the need to alter the existing datasets that are linked (Ingwersen et al. 2018).

The purpose of this study was to develop prototype software capable of generating reusable product system models from a standardized product specification and automatically linking these models with existing LCI data from disparate sources to create PSMs. The study adopted a methodology that draws from the “hackathon” model (Briscoe and Mulligan 2014), in which participants are invited to develop code to solve a well-specified technical problem. The contributions are available for review and adaptation by the public.

1.2 METHODS

A challenge was initiated to create prototype software that can assemble a full PSM from disparate data sources. Three experienced LCA software developers were invited to participate, with a fourth participating as a convener. Each developer was allotted up to 80 hours of independent work to perform the task. A set of requirements was given to each developer to follow (Table 1). Neither a graphical user interface, nor elementary and product flow harmonization, were required.

The developers were given data for an assembled product and asked to use their software to develop PSMs consisting of a foreground system for the selected assembly along with selected supporting background systems. The selected assembled product was the landing gear for a Boeing F/A-18E/F “Super Hornet” aircraft (F-18 LG). The landing gear consisted of independent nose (1) and rear (2) systems in the form of assemblies, where each assembly consisted of one or more subassemblies and/or individual components (

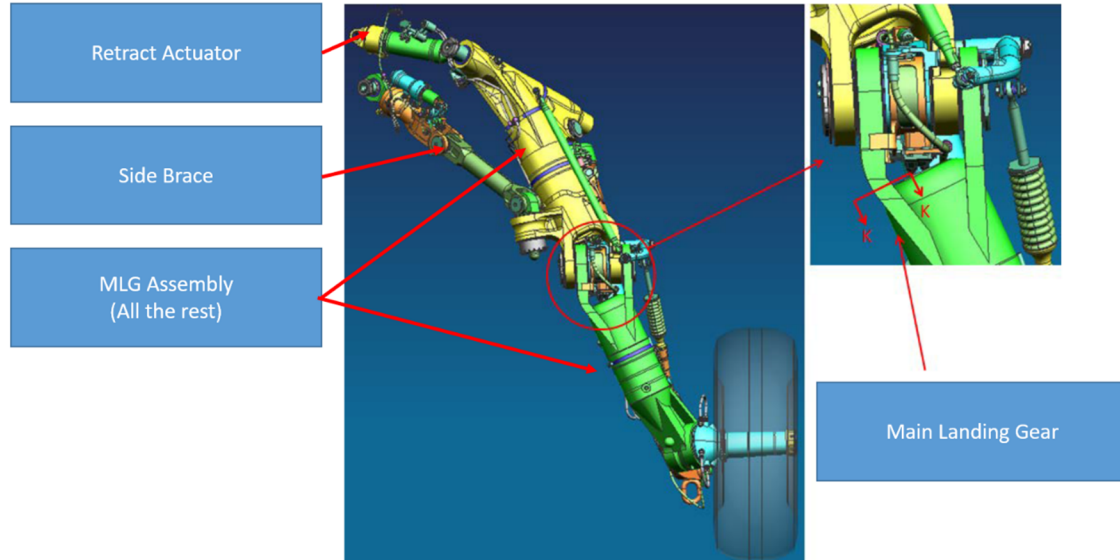


Figure 1). The data were provided in the form of a standard Bill of Materials (BOM) and were confidential.

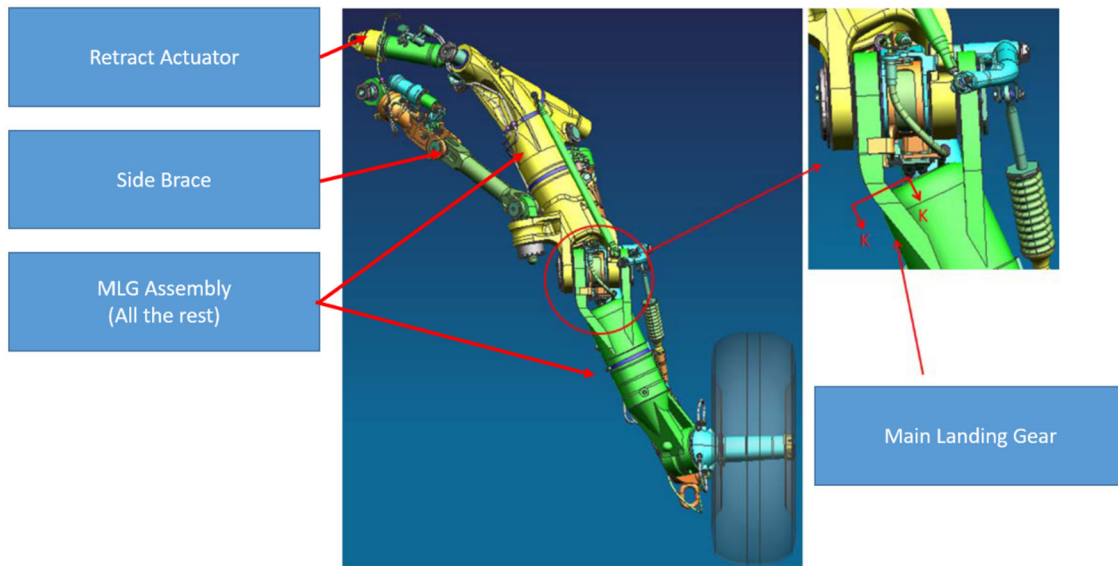


Figure 1. Schematic Drawing of One Landing Gear Product System

Table 1. Software Requirements

No.	Requirement
1	The software may make use of any existing software or source code, as long as the software/code license permits its reuse without any licensing fees.
2	LCI data used may come from open or proprietary sources. However, any proprietary source data must be aggregated in a manner suitable for data sharing.
3	The software must include automated data discovery functions to identify LCI to satisfy process requirements.
4	The software must be functional with LCI datasets from the Federal LCA Data Commons.
5	LCI data product flow names used in the PSM should be modified in content to the most limited extent possible. ‘Bridge processes’ provide an alternative method of linking unit processes without altering the product flow names (Ingwersen et al. 2018).
6	All dependencies must be documented and be publicly available or otherwise made available to the project team.
7	The software should read and generate LCI data in a described LCA data exchange format, preferably in JSON-LD according to the openLCA JSON-LD schema, but minimally in JSON-LD, ILCD, or ES2 formats.
8	Development in the Python 3 language is preferred, but not required.
9	The software must provide metadata on the PSM content and structure. Model contents may include description of sub-PSMs within the modeled PSM. See Kuczenski et al. (2018), section 5.
10	The software must work on current Windows and Macintosh operating systems and preferably on others, including Linux.

The team performed two tests of the extensibility of the prototypes, testing their performance in creating PSM models from two additional datasets. The first dataset was a BOM spreadsheet for the assembly of a printed circuit board (PCB). The spreadsheet included twelve components, of which seven were assigned part numbers and five were given only text descriptions. The second dataset was a life cycle inventory unit process for the sorting of construction and demolition debris in a material recovery facility (CDD MRF) taken from the FLCAC (USDA 2019). This dataset was provided in JSON-LD openLCA format and had approximately 65 exchanges, including both intermediate and elementary flows.

1.3 RESULTS

The input BOMs could all be interpreted as trees of assemblies, sub-assemblies, and components that could be converted into a foreground PSM. The components, or leaf nodes of the trees, represent inputs to the foreground from the background. Converting a BOM into a complete PSM was shown to require three modeling steps: constructing the trees; determination of quantitative properties of the leaf nodes; and linking the leaf nodes to background databases. These were accompanied by a fourth step of describing the models in a format that can be read into an LCA program, referred to as “serialization.” A complete solution to the study problem would complete all four steps. In this section, each prototype is presented along with a brief description. This is followed by an assessment of the prototypes in their success in creating PSMs from the F-18 LG dataset and with alternative datasets.

1.3.1 Presentation of Prototypes

Three prototypes *pslink*, *antelope*, and *perdu* were initially constructed by the three developers. All prototypes are written in Python and require a Python interpreter to use or launch. The prototypes are designed to run independently, though they required different levels of configuration and interaction.

1.3.1.1 *pslink*

pslink imports and stores BOM data as trees and uses them to create a foreground PSM. The user must then collect or mine data on the material composition and dimensions of the component inputs to the foreground model / leaves of the PSM tree and stores these attributes as key-value pairs in a text file for each component. The *pslink* quantitative solver tests each attribute file against a collection of symbolic formulas that compute the volume of various solids. These formulas can be dynamically registered, allowing a user to specify formulas specific to the input data. If a formula is found that can be computed using the supplied attribute information, the result is taken to be the volume of the part. In case multiple formulas could be satisfied by the supplied data, only the first formula encountered is used. *pslink* also requires that the user store density data on the materials identified in the composition data. Keywords from the component attribute data are matched against the stored densities. If a match is found, an exchange can be created, converting the previously calculated volume to a mass for a given material in a component. If multiple matches are found, the mass is partitioned equally among them.

The linking step in *pslink* starts from a simplified semantic graph that a user creates for the materials present in the attribute data. The users' input is a simple text file where each line declares the material name and then one or more related names, and their relationship to the material name, where relationship types are "same as, =", "broader than, ^" or "derived from, <". An example of records related to steel are:

```
"steel," "ferrous metal"^  
"steel product", "steel"<  
"stainless steel", "steel"^  
"stainless steel alloy", "alloy"^, "stainless steel"^  
"alloy 660", "stainless steel alloy 660"=, "stainless steel alloy"^
```

A network of semantic relations is created from this file, like shown in Figure 2, and stored as a graph.

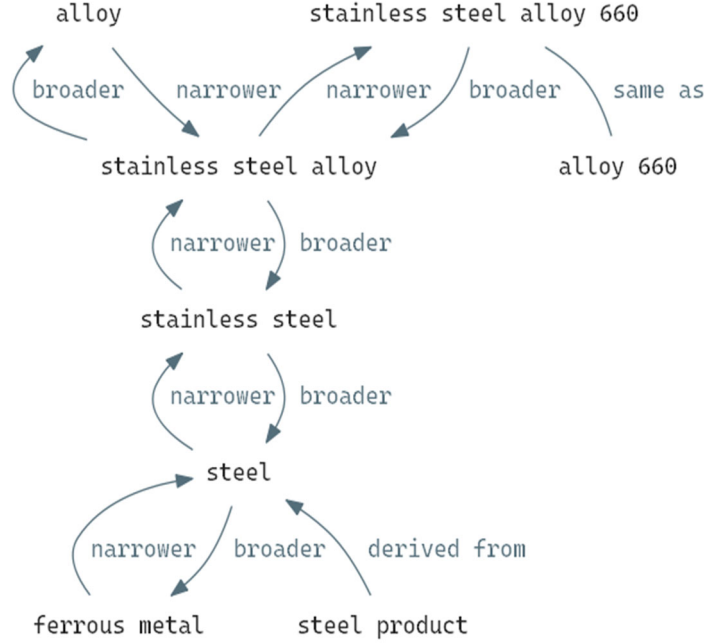


Figure 2. Semantic Network of Product Relations in *pslink*

To derive the product flow information from the background database of interest, the user is required to extract a list of processes and product flows names and compile them in a comma-separated text file. The products of the background database are bound to the products of the graph via a lexical analysis. The function of the lexical matching can be configured and different functions are implemented in *pslink* that consider specific syntactic elements of product names in LCI databases. For each binding, the lexical matching factor, f_l , is stored which has a value in the range of [0,1].

For each product in the foreground system, entry points into the semantic network are again searched via a lexical analysis. Starting from these entry points the graph is traversed along the relations to the nodes with bindings to products of the background database. Each relation is assigned a relation factor, f_r , that is specific for the relation type (e.g., “same as”: 1.0, “broader”/“narrower”: 0.75, “derived from”: 0.5). A traversal factor f_t is calculated for each path from an entry point to a binding of a background product by multiplying the lexical and relation factors across the segments, s , of the path p :

$$f_t = \prod_{s \in p} f_l f_r \quad \text{Eq. 1}$$

Figure 3 illustrates this procedure. The calculation of the traversal factors can be documented with the integrated *explain* function, which presents a list of the relations used to determine each factor.

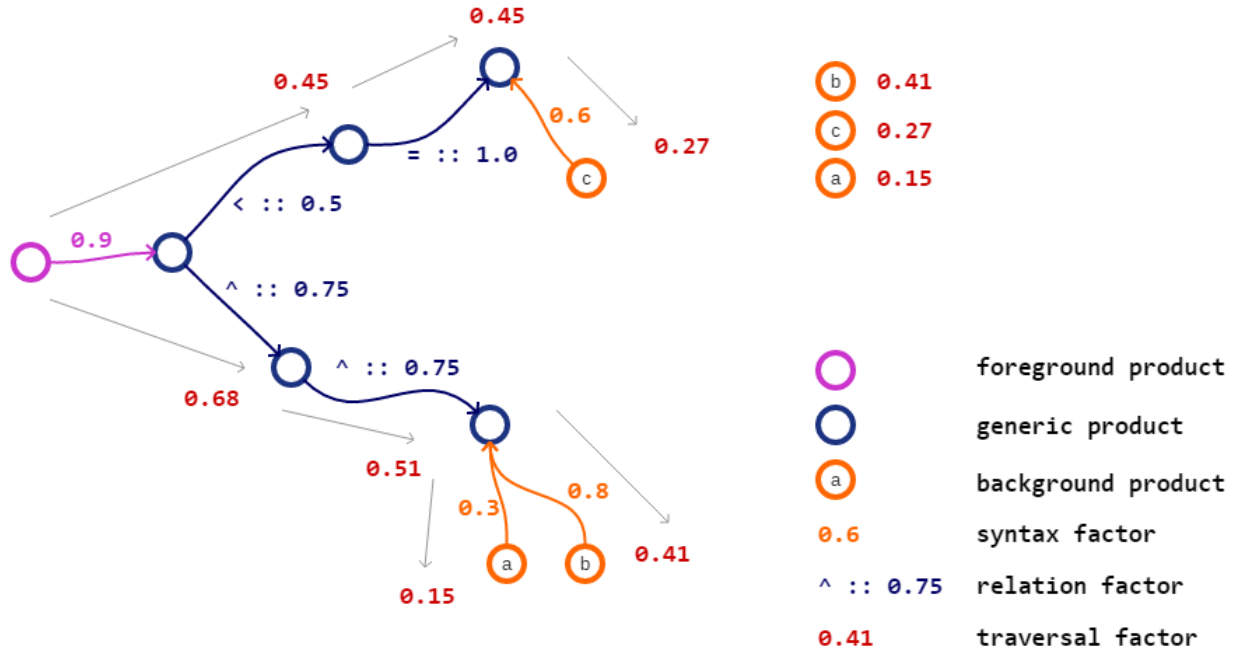


Figure 3. Calculating the Traversal Factor in *pslink*

The products of the background databases with the highest traversal factors are finally assigned to a product of the foreground system in the linking step. The collection of flows and linked unit processes are then serialized as JSON-LD following the openLCA schema (Srocka et al. 2020), allowing them to be imported into OpenLCA software for practitioner use. If the model is imported into an existing database that includes the linked background processes, then those processes would be linked to the designated datasets upon import.

1.3.1.2 *antelope*

antelope is a Python tool based on the [Antelope LCA](#) framework (Kuczenski and Beraha 2015). *antelope* also converts BOMs into product system models with a tree-like structure, with the reference flow at the root of the tree. *antelope* includes an algorithm for detecting and removing duplicate subassemblies using tree isomorphism (Valiente 2002). By treating duplicated assemblies as distinct product models, the duplicates are removed and the models are re-used in separate places in the product system.

When a PSM is described as a tree, the interior nodes and leaf nodes serve different purposes (Figure 4). Branches connecting interior nodes reflect direct relationships between foreground elements and define the structure of the PSM. Leaf nodes, on the other hand, represent flows which enter or leave the foreground system, and which must be linked to compute the model. If they are linked with a background activity, they become intermediate exchanges that represent a dependency of the foreground on that activity. Note that the direction of dependency is unrelated to the physical direction of the flow (i.e., a foreground process could depend on both material suppliers and waste disposal services). If they are linked with an environmental context, they become elementary exchanges (none of those were included in the current study).

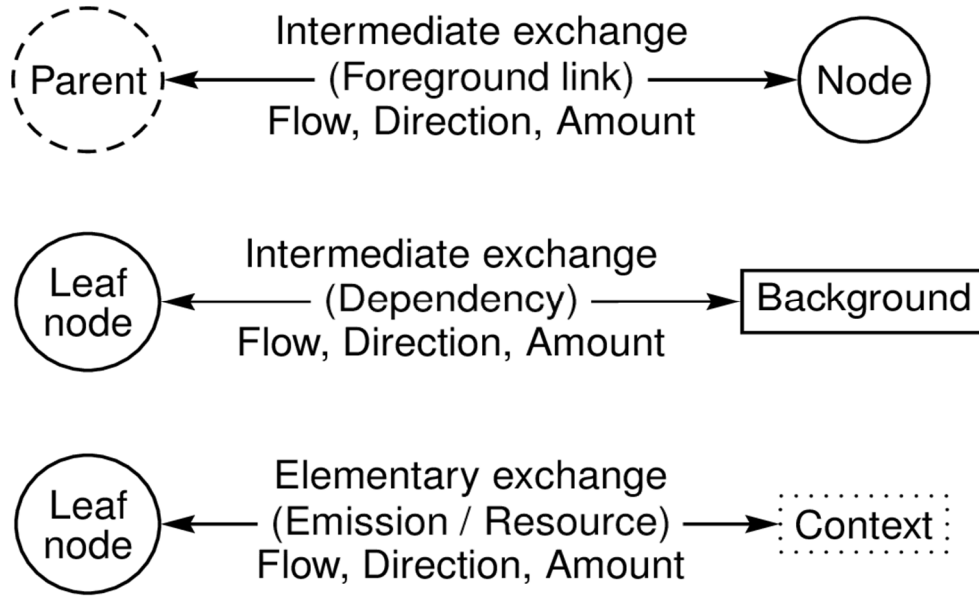


Figure 4. Exchanges and Dependency Relationships in a Product System Model

Circles represent foreground nodes, solid rectangle represents activities in a background database, and dotted rectangle represents an environmental context. The direction of the exchange can be either as an input to or output from the foreground.

Each link in the model must be assigned a weight, which is the exchange value of the link, or the quantity of the child flow that is required by the parent. During traversal of the tree, each node's weight is calculated by multiplying its parent's node weight by its exchange value. The record of this traversal can be used to reconstruct the model. In the *antelope* traversal implementation, any node can be specified as private, in which case the nodes interior to that node's subtree will be concealed and only the aggregate result reported.

Once the product models are created and linked to background datasets, the product system can be described as an LCA disclosure, a data structure that simplifies the expression of the model structure (Joyce 2019; Kuczynski 2019). Disclosures are constructed from a traversal of the product tree under a particular scenario. The *antelope* user can specify certain sub-assemblies or the whole assembly to be aggregated during the traversal, so that private or proprietary information in the PSM can be concealed from view while still contributing to the disclosure. The disclosures are saved as excel files for easier accessibility.

1.3.1.3 *perdu*

perdu is a Flask (Pallets Project, 2020) web application for matching product data against standard industry and product classifiers. *perdu* embeds classification systems, including the Global Product Classification (GPC), the North American Industry Classification System (NAICS), and U.S. Bureau of Economic Analysis Detail Input-Output schema as implemented in the USEEIOv1 model (Yang et al. 2017). *perdu* extracts names and descriptions from these classification systems and uses that data to build indices for searches. Various search algorithms provided by the Whoosh library (Matt Chaput, 2012) are implemented to search these indices using terms from the product data.

The web application was designed around a simple workflow (Figure 9). A user can upload a spreadsheet or table (Excel or csv file) with “name” and “description” fields or an openLCA JSON-LD archive with processes and flows to be matched. *perdu* will extract the names and descriptions from the tabular data or the processes and flows from the JSON-LD and store them as a term list and present them to the user for interactive matching. With the product data loaded and names presented to the user, the user selects a classification system to use, and an initial list of match suggestions is provided. The search terms can be refined in cases where the given labels are difficult for automatic algorithms to interpret. Multiple matches per query are possible. The user then selects a proposed match for each term and describes the relationship type; matches are either “exact,” “approximate,” “narrower than,” or “broader than.” The interface does not include any mechanism for quantitative characterization (i.e., the mass or dollar value of each node). The review process occurs in a browser window, and after it is complete the user can export a static file that contains the saved matching information. Following interactive search, review, and linking, *perdu* can produce output files of matches in a CSV, JSON, or Turtle (TTL) formats. The web application also includes a stand-alone search window.

1.3.1.4 Summary

Table 2 presents a summary of the prototypes.

Table 2. Summary Characteristics of Three Initial Prototypes

Prototype	Primary Dependencies	Data Discovery	Supervised or Unsupervised Matching?	Data Sources Linked	Product System Statistics?	Output formats
<i>pslink</i>	olca-py	Semantic similarity	Unsupervised	USEEIOv1.1, LCA Commons, Ecoinventv2.2	Y	JSON-LD
<i>antelope</i>	Antelope LCA	Tree isomorphism	Supervised	USEEIOv1.1, USLCI, ELCD	Y	JSON, Excel
<i>perdu</i>	Whoosh, RDFLib	Metadata search	Supervised	USEEIOv1.1, GPC, NAICS	N	CSV, TTL, JSON-LD

1.3.2 Prototype Testing

1.3.2.1 Landing Gear Model

The principal test case, the F-18 LG dataset, consisted of eight assemblies supplied in two BOMs in the form of Excel spreadsheets. The BOM was anonymized using a software routine included in the *antelope* contribution and is henceforth referred to as F-18 LG ANON. The resulting foreground models were observed to have a total of 1,528 distinct elements, where each element corresponds to a row in a BOM. These models included 1,045 distinct flows, of which 810 were leaf nodes, some of which were reused more than once.

In order to prepare *pslink* for use with the F-18 LG ANON foreground model, a mechanical parts database (PartTarget Inc. 2019) was mined to extract material composition and dimensions. As

different components can have very different attributes that describe their dimensions, it was necessary to accommodate a wide range of attribute names. Density data for materials identified in the parts were manually input from various online sources (amesweb.info 2019; MatWeb, LLC 2020; ThoughtCo Inc. 2020). A semantic graph of associated material relationships was manually created. *pslink*, enabled with these input data, constructed PSMs from the F-18 LG ANON dataset and selected background datasets. The linkages between the F-18 LG ANON dataset and background databases are summarized in Table 3.

Table 3. *pslink* Match Statistics for F-18 LG ANON Dataset

FS	FS unique materials	BS	BS unique product flows	Materials with at least 1 possible match	Possible material matches	Leaves with links established
F-18 LG ANON	27	EIv2.2	1924	26	66	77
F-18 LG ANON	27	USEEIOv1.1	388	26	30	77
F-18 LG ANON	27	LCA Commons	752	25	32	70

FS: Foreground data source; BS: Background data source

The data mined had identified 27 unique materials in the F-18 LG ANON dataset that became the basis of *pslink*'s matches to background database products. In all three background sources, corresponding product flows were found for nearly all materials (25 or 26 out of 27). Each product flow could then be linked to one or more providing processes; the database with the most unique product flows (Ecoinventv2.2) able to provide the most potential matches to these materials (66). But the number of foreground model leaf nodes for which linkages were created was the same between the USEEIOv1.1 and Ecoinventv2.2 sources (77) and only slightly less for the LCA Commons source (70), which could be explained by the one less material provided by the LCA Commons in comparison with the other two databases.

The created models were exported from *pslink* in JSON-LD and successfully imported into openLCA (Figure 5) where it was verified that life cycle inventory results could be calculated to generate nonzero scores (Figure 6). However, the simple tree-traversal algorithm led to doubled exchanges in 19 duplicate processes.

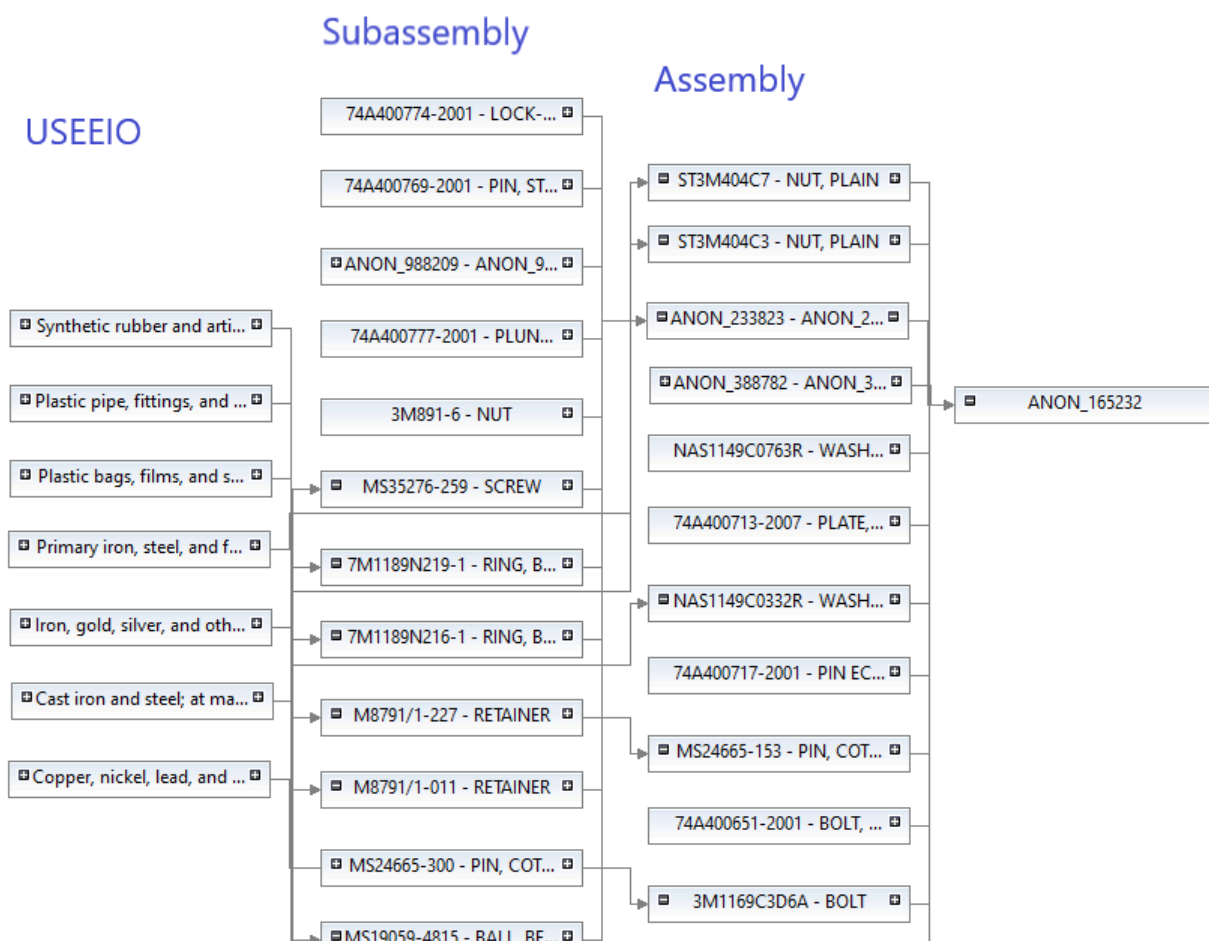


Figure 5. PSM of Landing Gear Connected to USEEIO in openLCA

Contribution	Process	Amount	Unit
✓ 100.00%	P ANON_165232	4.04240	kg
✓ 64.45%	P ANON_233823 - ANON_233823 Sub-assembly	2.60531	kg
✓ 61.69%	P [REDACTED]	2.49373	kg
> 61.69%	P Stainless steel; Manufacture; Production mix, at plant; 316 2B, 80% recycled - GLO	2.49373	kg

Figure 6. LCIA Results Computed From the Linked Landing Gear Model in openLCA. Text of component intentionally obscured to protect privacy of data.

antelope captured the tree structure of the F-18 LG ANON dataset, including detection and removal of duplicated subassemblies. Within the structure of the model, nineteen duplicate subassemblies were detected, making up 126 elements or about 8 percent of the total. The models were exported as disclosures in spreadsheet format to facilitate review. Visualizations of portions of the foreground and background for the study test case are shown in Figure 7 and Figure 8. This prototype did not use data discovery to determine flow properties of parts, and instead used manual designation of each part's "part type," "material," and "size" as a basis to link to valid datasets in the USLCI, USEEIO, and ELCD databases. Because this procedure was done by hand it was very time-intensive and potentially error-prone but could be applied consistently to all 810 unique

components. A total of 25 USLCI processes, 19 USEEIO sectors, and 17 ELCD processes were identified as linking targets. The curated characterizations were included in the *antelope* contribution in spreadsheet form. The price and mass characterizations were random.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2	0. 74A400500-1017 ASSEMBLY Item(s) (Input)																							
3	1. 74A400500-1017 ASSEMBLY Item(s) (Input)	1																						
4	2. TRUNNION ASSY Item(s) (Input)		1																					
5	3. RING, TIEDOWN Item(s) (Input)			2																				
6	4. RING, TIEDOWN Item(s) (Input)				1																			
7	5. TRUNNION ASSY Item(s) (Input)			1																				
8	6. TRUNNION Item(s) (Input)					1																		
9	7. TRUNNION Item(s) (Input)						1																	
10	8. SCREW, CAP Item(s) (Input)		2																					
11	9. SCREW, CAP Item(s) (Input)									1														
12	10. SUPPORT ASSY Item(s) (Input)		1																					
13	11. SUPPORT Item(s) (Input)										1													
14	12. SUPPORT ASSY Item(s) (Input)		1																					
15	13. SUPPORT Item(s) (Input)													1										
16	14. LEVER/AXLE ASSY Item(s) (Input)		1																					
17	15. BRKT ASSY, PROX Item(s) (Input)															1								
18	16. BRACKET Item(s) (Input)																1							
19	17. PIN ASSY Item(s) (Input)															1								
20	18. PIN Item(s) (Input)																	1						
21	19. BOLT, ASSY, RETNR Item(s) (Input)															1								
22	20. BOLT, ASSY, RETNR Item(s) (Input)																			1				
23	21. BRKT ASSY, SPR Item(s) (Input)															1								
24	22. CRANK ASSY Item(s) (Input)															1								
25	23. CRANK, AXLE Item(s) (Input)																						1	
26	24. CRANK, AXLE Item(s) (Input)																							1
27	25. BEARING, PLAIN Item(s) (Input)																							
28	26. BEARING, PLAIN Item(s) (Input)																							
29	27. LEVER ASSY Item(s) (Input)															1								
30	28. LEVER Item(s) (Input)																							
31	29. LEVER Item(s) (Input)																							
32	30. BEARING Item(s) (Input)																							
33	31. BEARING Item(s) (Input)																							
34	32. LINK ASSY, UPR Item(s) (Input)															1								
35	33. YOKE ASSY Item(s) (Input)															1								
36	34. LINK ASSY, LWR Item(s) (Input)															1								
37	35. LINK Item(s) (Input)																							
38	36. ROD ASSY, SPR Item(s) (Input)															1								
39	37. ROD ASSY, SPR Item(s) (Input)																							

Figure 7. A Portion of a Model's Internal Nodes (foreground) Documented in a Spreadsheet

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
steel hot rolled coil, blast furnace route, production mix, at plant, thickness 2 to 7 mm, width 600 to 2100 mm [RER] kg (Output)		1.57	0.58													
steel rebar, blast furnace and electric arc furnace route, production mix, at plant [GLO] kg (Output)		0.21	0.22							0.03		0.05				1.05
Steel sections [GLO] kg (Output)		0.44														0.37
Steel hot rolled coil [GLO] kg (Output)			2													0.23
aluminum extrusion profile, primary production, production mix, at plant, aluminum semi-finished extrusion product, including primary production, transformation and recycling [RER] kg (Output)						2.88										0.81
copper tube, technology mix, consumption mix, at plant, diameter 15 mm, 1 mm thickness [EU-15] kg (Output)						0.24										0.15
aluminum sheet, primary production, production mix, at plant, aluminum semi-finished sheet product, including primary production, transformation and recycling [RER] kg (Output)		1.49														
nylon 66 GF 30 compound (PA 66 GF 30), production mix, at plant [RER] kg (Output)																0.62
acrylonitrile-butadiene-styrene granulate (ABS), production mix, at plant [RER] kg (Output)																0.21
polyethylene high density granulate (PE-HD), production mix, at plant [RER] kg (Output)		0.01														
Other aircraft parts; at manufacturer [United States] USD (Output)		666														
Screws, nuts, and bolts; at manufacturer [United States] USD (Output)		65														
Other plastic products; at manufacturer [United States] USD (Output)		999														
Aircraft engines and parts; at manufacturer [United States] USD (Output)		999														
Relay and industrial controls; at manufacturer [United States] USD (Output)		999														
Ball and roller bearings; at manufacturer [United States] USD (Output)		999														
Metal pipe fittings, ball and roller bearings, industrial patterns, metal bath fixtures, other misc. fabricated metals; at manufacturer [United States] USD (Output)		999														
Synthetic rubber and artificial and synthetic fibers; at manufacturer [United States] USD (Output)		107														
Aluminum products; at manufacturer [United States] USD (Output)		443														
Secondary steel products; at manufacturer [United States] USD (Output)		60														
Valve and fittings (except for plumbing); at manufacturer [United States] USD (Output)		721														
Secondary copper products; at manufacturer [United States] USD (Output)		999														
Fluid meters and counting devices; at manufacturer [United States] USD (Output)		999														
Springs and wires; at manufacturer [United States] USD (Output)		999														
Other primary nonferrous metals; at manufacturer [United States] USD (Output)		999														

Figure 8. A Portion of a Model’s Leaf Nodes (background) Documented in a Spreadsheet

perdu did not perform automatic model generation, but instead focused on the classification of leaf nodes in the model. The prototype included search indexing frameworks for three different classification schemes: [NAICS](#), [GSI](#), and the U.S. system of national accounts, as implemented for LCA in the [USEEIO](#) database (Yang et al. 2017). Other classifications can be added by following the patterns laid out in the prototype.

1.3.2.2 Handling of Alternative Input Datasets

Because of the similarity of first alternative dataset, the PCB BOM, to the main test case, all three tools were successful in processing the PCB BOM. There were no duplicate flows. *antelope* and *pslink* prototypes correctly interpreted the spreadsheet and constructed a foreground model; however, the *pslink* prototype omitted the flows that had no part numbers. Because of lack of any text descriptions accompanying the part numbers in the BOM, the flow property description step failed for both *pslink* and *antelope*, and no linking was performed. With the *perdu* prototype, on the other hand, it was possible to identify matching classifications from the available schemes.

The attempt to use the second alternative dataset, the CDD MRF unit process, was less successful. Although openLCA and the *antelope* core software can both access OpenLCA datasets, *perdu* was the only prototype able to accommodate input data in this format directly (see Figure 9). The CDD MRF test revealed another issue that was relevant to the project. Ten exchanges in the model included links to “bridge processes” for various activities found in the USEEIO database, using the “defaultProvider” property defined in the openLCA schema (Srocka et al. 2020). However, the bridge processes themselves were not included along with the CDD MRF dataset, and so there was no way for any software to “build” the links or to identify the linking targets. This outcome revealed that further development is needed to successfully make use of bridge process specifications.

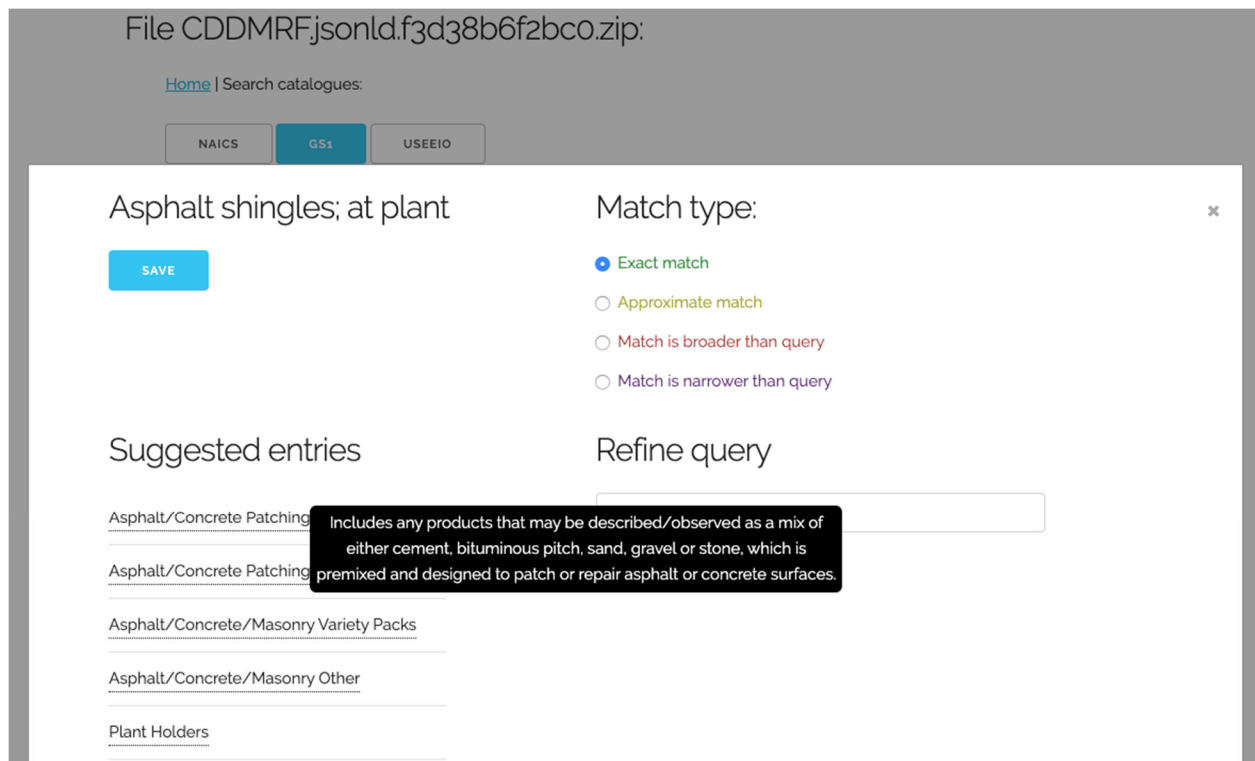


Figure 9. Matching the CDD MRF Dataset Using *perdu*

1.3.3 *autoprox* – An Additional Prototype for Direct openLCA Software Integration

The challenges associated with the second alternative dataset - namely an existing unit process - inspired creation of an additional prototype, *autoprox*, that can link a unit process with background datasets, and make these links in the form of bridge processes (Ingwersen et al. 2018). *autoprox* is written in Kotlin and depends on the openLCA core API and a Java virtual machine. It reads directly from an openLCA database rather than from input files like the Excel files storing the BOMs. The user supplies two inputs: the id of the unit process in the openLCA database for which linking to datasets to provide product flow inputs is desired, and one of three alternative *matchers*, which are implementations of common matching algorithms. The best matches are then used to automatically create bridge processes within the given openLCA database between the given process and other available processes in the database.

The three different *matchers* each use principles of lexical and/or semantic matching between flow names (Pawar and Mago 2018), but use different forms of information to evaluate the match. The first matcher, *BigramsDiceMatcher*, uses “bigrams,” which are sets of two consecutive letters, syllables, or words in a text string. It works well for matching flows that have distinctive words or phrases, like “asphalt shingles,” but less well for flows that are made up of common or generic words. The second matcher, *InfoContentMatcher*, uses the “information content” of a word, defined as the length of the word times the inverse of its frequency in the collection of candidate processes. Longer words and less-frequent words have higher information content. The information content of flow words is then compared to the information content of provider flows to find the best matches. The third matcher, *WordNetPathMatcher* also uses the information

content of words, but combines this information with a semantic similarity score derived from a large natural language corpus. The reader is encouraged to read more about these matchers on the *autoprox* site (link provided in the Code Availability section).

Like the other solutions, *autoprox* is limited to identifying possible matches from the background processes that are available, which in this case come from the given openLCA database. The techniques implemented in the *autoprox* prototype could provide inspiration for further research in semantic and lexical matching for LCA applications.

1.4 DISCUSSION

While each prototype has its strengths and deficiencies, none could transform a product specification into a complete PSM with all product flows provided by background datasets without manual intervention. The team found that the solution to automatically generating product system models from standardized product data has several steps, and that some of those steps are more amenable to automation than others. Each prototype demonstrated progress toward this end, yet all still require user involvement for one or more steps.

Figure 10 shows the overall problem and identifies the key contributions of each prototype.

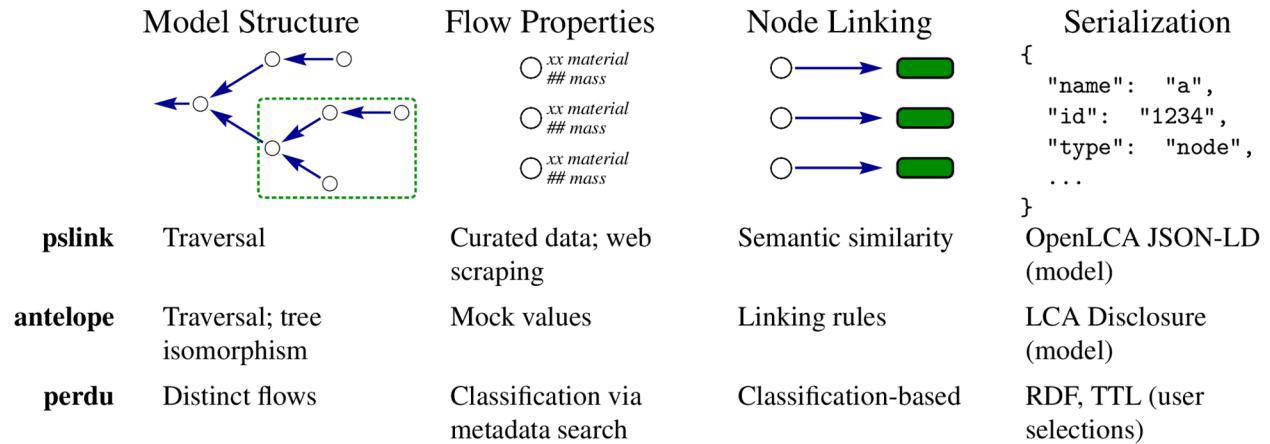


Figure 10. The Four Steps for Developing a PSM From a Product Dataset Identified in the Study, and the Contributions of Each Prototype to Each Step

The first step was specifying the model structure. The *pslink* and *antelope* prototypes were capable of transforming BOM input data into a product system model to serve as a foreground system. Notably, this was not the first demonstration of BOM to LCI automation (Sundaravaradan et al. 2011; Tao et al. 2014; McIntosh and Koffler 2014; Mishra and Singh 2019), but these are the first open source implementations of such conversions. Detecting and removing duplicate subtrees may be required to avoid double counting, with *antelope* the only prototype designed to detect and remove duplicate subtrees. *antelope* is also capable of optionally anonymizing assemblies and subassemblies in a BOM, which adds value for protecting primary data confidentiality. Modeling of the second test dataset, a unit process, had different requirements, which were met through the creation of an additional prototype, *autoprox*.

The second step – flow property description – was found to be the most challenging of the four. No prototype was able to perform this step without user input. Each component from a BOM that was determined to be a leaf node was identified only as a “piece,” but background datasets provide flows in terms of physical extent, most frequently mass, or by price. Therefore, before linking could be performed, each component required qualitative and quantitative description. This step has not received much attention in the literature for process LCA, although it is also present in hybrid LCA, where physical flows must be expressed in terms of their monetary value in order to link to economic input-output models (Suh and Huppes 2005). Each prototype took a unique approach to flow description. *pslink* takes a material-centric approach, where it uses component material composition and dimensions along with density data and registered volume formulas to calculate the mass of each material in a component. For *pslink* to perform this step, the user is required to collect these physical data and prepare them in defined formats, some of which can be automated. In the *antelope* case, the qualitative description of each leaf node is performed by the user by manually mapping a flow to a reference background dataset product flow, and then the mappings were used to generate mock quantitative data according to heuristics. *perdu* works on the assumption that the flow property description problem is a nomenclature problem, recognizing the lack of a common nomenclature for products/intermediate flows across life cycle inventory sources. Classifications systems are a common way of product identification used in product category rules and environmental product declarations (Ingwersen and Stevenson 2012; Minkov et al. 2015), and *perdu* classifies flows according to one of these formal systems. In an application, it is possible that a modeler will already have direct access to information about the parts contained in their product system, and this information must be used in a tool-dependent manner to describe these properties.

In the node linking step, the use of a semantic linking demonstrated by *pslink* and *autoprox* shows great promise to select “best-fit” background process to provide the required foreground input flow. The use of large lexical databases and machine learning algorithms has been identified as a promising approach for reconciling different background LCI databases (Davis et al. 2010; Kuczenski et al. 2016). The *pslink* prototype implementation depended on a locally-defined semantic network, and the *autoprox* prototype extended this approach to use a variety of matching algorithms, including more robust semantic databases. These techniques can be used for foreground systems when sufficient lexical content is available. The novel linkage metrics from *pslink* - the lexical, relation and traversal factors - demonstrate a way to assess the validity of machine-based links. However, the approaches may generate false matches, suggesting that the results would still need to be manually reviewed. They also can require substantial computational resources. Each tool was limited to searching one user-specified background data source specified by the user at a time.

Serialization of the model to a format readable by LCA software is the simplest problem to solve, with the *pslink* contribution demonstrating straightforward implementation of JSON-LD export and linking to target background datasets present in openLCA. The *perdu* prototype does not construct LCA models. *perdu* prepares lists of database matches for further use in a linked data approach, which would be possible to integrate with emerging data resources such as BONSAI (BONSAI 2020). The LCA disclosure framework used in *antelope* can define the relationships, but further development is required to support serialization. Only *pslink* met the study requirements to generate direct inputs to openLCA but a linked dataset is dependent upon the presence of the background dataset in the software to forge the link. In contrast, the *antelope*

prototype uses an LCA disclosure format that is not limited to a single source of background data. This is intended to be accomplished by providing “stable semantic references” to the background datasets used. In practice, this is accomplished by naming each background dataset with both an “origin” that identifies the specific data source, and an “external reference” that unambiguously specifies a particular dataset. In principle, an LCA software system could obtain the exact dataset specified as long as it can properly interpret the named origin. While it would be straightforward to express the LCA disclosure format using the openLCA schema, this functionality is not yet implemented. It would still be subject to the same constraints as the *pslink* serialization: the referenced datasets would need to be present in the database in order to be linked.

1.4.1 Potential and Limitations

These prototypes are not yet integrated into the LCA software tools already used in the community by practitioners. However, shared assets exist that would facilitate their incorporation. These assets include: (1) each prototype is already packaged in a way that facilitates automated install or embedding in other tools; (2) the prototypes are written in a language (Python) that can already be run by tools such as openLCA or Brightway via their existing interfaces; (3) the tools exports matches and PSM in standard exchange formats already readable by LCA software tools like JSON-LD and ILCD; (4) the open-source licenses permit reuse, embedding, and modification without licensing; and (5) management of the source code for the prototypes in a version control system that supports collaboration and issue tracking (github).

The integration of one or more of the prototypes into an LCA tool would make it more accessible to a larger body of practitioners.

1.5 CONCLUSION

The PSM automation project demonstrated success in both principle and practice. In principle, this method of challenging experts to solve specific problems by independently and creatively writing code within a set of technical and logistical constraints can be effective in the LCA domain. In practice, the present study generated three operable prototypes that together describe the solution space of the study problem. The process of converting a foreground into a product system model requires four different steps: description of the model structure, characterization of flow properties, linking flows to background databases, and serializing the model into a format that can be interpreted by LCA software. The three prototypes, along with the additional *autoprox* tool, showed that experienced LCA software developers could write software to address, in some cases, all of the steps, in a relatively short period of time. Thus, with greater investment of time into refining or improving these prototypes, or into further development of methods to tackle one of the four steps delineated by this study, is likely to yield very useful product system assembly tools. It is our hope that the community continue to build on the work that is included in this project.

1.6 CODE AVAILABILITY

The project and prototypes are available as open-source software on the collaborative coding platform GitHub. The resources provided to the developers, including test datasets can be found in the [Product System Assembly Resources](#) repository. Each prototype was published independently by its author, and the repositories are publicly available:

1. *antelope*: https://github.com/bkuczenski/antelope_epa
2. *pslink*: <https://github.com/msrocka/pslink>
3. *perdu*: <https://github.com/cmutel/perdu>
4. *autoprox*: <https://github.com/msrocka/autoprox>

1.7 REFERENCES

- amesweb.info (2019) Amesweb - Advanced Mechanical Engineering Solutions. amesweb.info
- BONSAI (2020) BONSAI. bonsai.uno
- Briscoe G, Mulligan C (2014) Digital innovation: The hackathon phenomenon. <http://www.creativeworkslondon.org.uk/wp-content/uploads/2013/11/Digital-Innovation-The-Hackathon-Phenomenon1.pdf>
- Canals LM i, Boureima F, Brago T, et al (2016) Towards an Interoperable Network of LCA Databases
- Cashman SA, Meyer DE, Edelen A, et al (2016) Mining Available Data from the United States Environmental Protection Agency to Support Rapid Life Cycle Inventory Modeling of Chemical Manufacturing. ENVIRON SCI TECHNOL 50:9013–9025. <https://doi.org/10.1021/acs.est.6b02160>
- Ciroth A, Srocka M (2017) Introducing the LCA data machine. <http://brussels.setac.org/>
- Davis C, Nikolic I, Dijkema GPJ (2010) Industrial Ecology 2.0. J IND ECOL 14:707–726. <https://doi.org/10.1111/j.1530-9290.2010.00281.x>
- GreenDelta (2019) openLCA Nexus. <https://nexus.openlca.org/>
- Ingwersen W, Lesage P, Mutel C, et al (2015) LCA Data Interoperability: New Solutions to Old Challenges. Special Session. Vancouver, Canada
- Ingwersen WW (2015) Test of U.S. federal life cycle inventory data interoperability. J CLEAN PROD 101:118–121. <https://doi.org/http://dx.doi.org/10.1016/j.jclepro.2015.03.090>
- Ingwersen WW, Kahn E, Cooper J (2018) Bridge Processes: A Solution for LCI Datasets Independent of Background Databases. INT J LIFE CYCLE ASS 23:2266–2270. <https://doi.org/10.1007/s11367-018-1448-6>
- Ingwersen WW, Stevenson MJ (2012) Can we compare the environmental performance of this product to that one? An update on the development of product category rules and future challenges toward alignment. J CLEAN PROD 24:102–108. <https://doi.org/10.1016/j.jclepro.2011.10.040>
- Joyce PJ (2019) Lca_disclosures. https://github.com/pjamesjoyce/lca_disclosures/tree/typed_flows
- Kuczenski B (2019) Disclosure of Product System Models in Life Cycle Assessment: Achieving Transparency and Privacy. J IND ECOL 23:574–586. <https://doi.org/10.1111/jiec.12810>
- Kuczenski B, Beraha S (2015) [Antelope a web service for publishing life cycle models and results](#). In: Proceedings of the International Symposium on Sustainable Systems and Technology

- Kuczenski B, Davis CB, Rivela B, Janowicz K (2016) Semantic catalogs for life cycle assessment data. J CLEAN PROD 137:1109–1117. <https://doi.org/10.1016/j.jclepro.2016.07.216>
- Kuczenski B, Marvuglia A, Astudillo MF, et al (2018) LCA Capability Roadmap: Product System Model Description and Revision. INT J LIFE CYCLE ASS 23:1685–1692. <https://doi.org/10.1007/s11367-018-1446-8>
- Matt Chaput (2012) Whoosh. <https://whoosh.readthedocs.io/en/latest/index.html>
- MatWeb, LLC (2020) Matweb. <http://matweb.com>
- McIntosh BS, Koffler C (2014) Using Bill of Materials imports and dynamic EPD templates to minimize the costs of Type III Environmental Product Declarations. https://www.researchgate.net/profile/Christoph_Koffler/publication/280880839_Using_Bill_of_Materials_imports_and_dynamic_EPd_templates_to_minimize_the_costs_of_Type_III_Environmental_Product_Declarations/links/55ca0e0808aebc967dfbd7d5.pdf
- Minkov N, Schneider L, Lehmann A, Finkbeiner M (2015) Type III Environmental Declaration Programmes and harmonization of product category rules: Status quo and practical challenges. J CLEAN PROD 94:235–246. <https://doi.org/10.1016/j.jclepro.2015.02.012>
- Mishra S, Singh SP (2019) Carbon management framework for sustainable manufacturing using life cycle assessment, IoT and carbon sequestration. BENCHMARKING ahead-of-print: <https://doi.org/10.1108/BIJ-01-2019-0044>
- Mutel C (2017) Wurst. <https://wurst.readthedocs.io/>
- Pallets Project (2020) Flask. <https://palletsprojects.com/p/flask/>
- PartTarget Inc. (2019) Part Target. <http://parttarget.com/>
- Pawar A, Mago V (2018) [Calculating the similarity between words and sentences using a lexical database and corpus statistics](#). arXiv:180205667 [cs]
- Srocka M, Ingwersen, Wesley, Ciroth, Andreas (2020) openLCA-schema. <https://github.com/GreenDelta/olca-schema>
- Suh S, Huppes G (2005) Methods for Life Cycle Inventory of a product. J CLEAN PROD 13:687–697. <https://doi.org/10.1016/j.jclepro.2003.04.001>
- Suh S, Leighton M, Tomar S, Chen C (2016) Interoperability between ecoinvent ver. 3 and U.S. LCI database: A case study. INT J LIFE CYCLE ASS 21:1290–1298. <https://doi.org/10.1007/s11367-013-0592-2>
- Sundaravaradan N, Marwah M, Shah A, Ramakrishnan N (2011) [Data mining approaches for life cycle assessment](#). In: Proceedings of the 2011 IEEE International Symposium on Sustainable Systems and Technology. pp 1–6
- Tao F, Zuo Y, Xu LD, et al (2014) Internet of Things and BOM-Based Life Cycle Assessment of Energy-Saving and Emission-Reduction of Products. IEEE Transactions on Industrial Informatics 10:1252–1261. <https://doi.org/10.1109/TII.2014.2306771>
- ThoughtCo Inc. (2020) ThoughtCo. <https://www.thoughtco.com/>
- UNEP/SETAC Life Cycle Initiative, (2019) GLAD: Global LCA Data Access network. <https://www.globalcadataaccess.org/>

USDA (2019) Federal LCA Commons Data Portal. <https://www.lcacommons.gov/>

Valiente G (2002) [Tree Isomorphism](#). In: Algorithms on Trees and Graphs. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 151–251

Yang Y, Ingwersen WW, Hawkins TR, et al (2017) USEEIO: A New and Transparent United States Environmentally-Extended Input-Output Model. J CLEAN PROD 158:308–318.

2 TASK 2.1 - IDENTIFY AND IMPROVE AN EXISTING IMPACT ASSESSMENT METHOD WITH RELEVANCE TO DOD

This chapter contains the content of the following published journal article that was written as part of this project:

Henderson AD, Niblick B, Golden HE, Bare JC (2021) Modeling spatially resolved characterization factors for eutrophication potential in life cycle assessment. Int J Life Cycle Assess. <https://doi.org/10.1007/s11367-021-01956-4>

2.1 INTRODUCTION

Eutrophication is a global environmental issue that spans the freshwater-marine continuum (Paerl 2009). Excess nutrients, especially nitrogen (N) and phosphorus (P), enter the biosphere through point and non-point sources of pollution and often contribute to harmful algal blooms (HABs). These nutrients eventually reach surface waters and result in degraded water quality conditions, impacting aquatic ecosystems and human health as well as local economies.

Life cycle assessment (LCA) provides a standardized quantitative framework from which to evaluate eutrophication potential of freshwater and marine systems (measured in kg N_{eq} and kg P_{eq}, respectively) of a product or process across its lifespan. Life cycle inventory (LCI) models provide the data needed for LCA, while life cycle impact assessment (LCIA) methods translate the data into potential impacts. Various existing LCIA methods incorporate eutrophication models, and a few models exist separately, however each method and model inform a different aspect of this environmental issue (Morelli et al. 2018). Furthermore, many LCIA models lack the spatial precision needed to operate on a refined global scale, often assigning a single characterization factor (CF) to areas representing many different conditions. The data and models required to support this high level of spatial resolution have been developed inconsistently across the globe, leaving major gaps in the research.

Characterization factors for eutrophication potential in the Tool for Reduction and Assessment of Chemical and other environmental Impacts (TRACI) estimate the eutrophication potential of a release of chemicals containing N or P to air or water, per kg of chemical released (Bare 2011, Bare et al. 2002, Norris 2003). Characterization of eutrophication potential in LCA would benefit substantially from improved spatial resolution of the environmental processes, including fate, transport, and potential environmental impacts to air, land, and water (Potting 1994). This paper builds from the eutrophication model review presented in Morelli et al. (2018) and its set of recommendations for improvement to the eutrophication potential category in LCA.

The current work presents a first effort to spatially differentiate eutrophication CFs to improve the scientific robustness of fate and transport mechanism characterization. Efforts are focused on the contiguous United States, using TRACI as the LCIA method (Bare 2012), and thereby update TRACI's eutrophication potential impact category to provide a state-of-the-art global site-specific model for research and public use.

2.2 METHODS

2.2.1 Nutrient Framework

In keeping with current LCIA practice, this project uses the simplified assumption that N and P cause eutrophication impacts in systems where they are the limiting nutrient, i.e., freshwater systems are P-limited and marine systems are N-limited. This simplification should be updated in future work, as co-limitation can occur in marine and freshwater systems (Morelli et al., 2018; Payen and Ledgard, 2017; Schindler, 2006). The calculations and models below have been developed for both systems. Figure 11 summarizes each step leading up to the result.

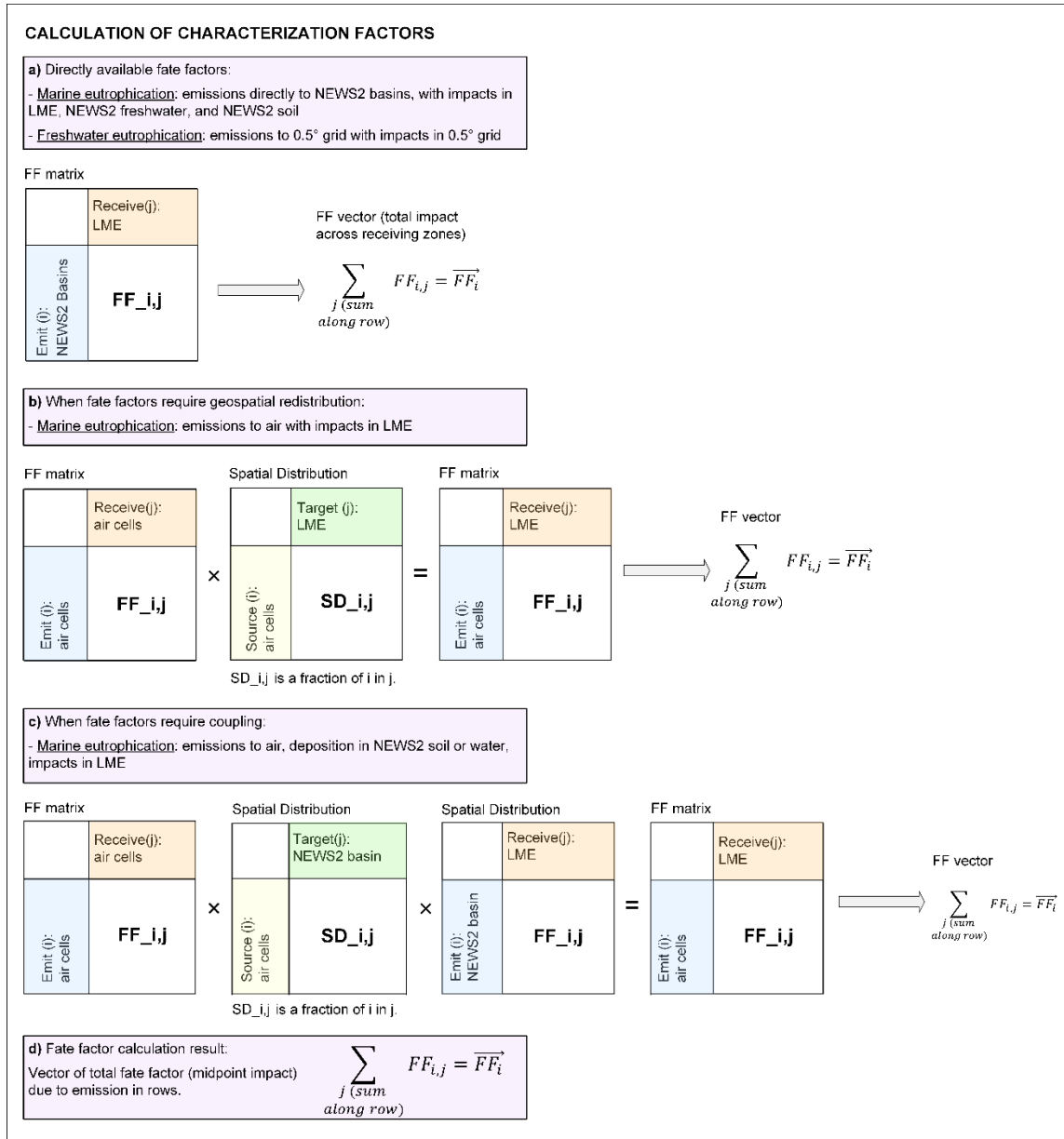


Figure 11. Calculation of Characterization Factors for Eutrophication Potential

Each step of the process is described in the purple boxes. For each of the matrices, where applicable: the source vector is shown in yellow, the emissions vector is shown in blue, the receiving vector is shown in orange, and the target vector is shown in green.

2.2.1.1 Freshwater Eutrophication

The freshwater eutrophication fate model is provided by Helmes et al. (2012), and is driven by phosphorus limitations. This model is based on a $0.5^\circ \times 0.5^\circ$ (approximately 50 km $\times$ 50 km at mid-latitudes) hydrologic model and returns characterization factors (CFs) for phosphorus emissions (Fekete et al., 2002). While higher resolution datasets are becoming available (Lehner and Linke, 2015), these have not yet been linked to phosphorus fate in freshwater systems. The current fate model accounts for phosphorus removal processes (i.e., advection, retention, and water use) through water withdrawal, uptake by biomass, and adsorption to suspended solids in waterways. It does not account for particulate P in the transfer fraction from agricultural soils to surface freshwater. Nor does it account for sewage treatment removal of P, which can lead to an overestimation of P transport for flows passing through urban centers.

2.2.1.2 Marine Eutrophication

Recent modeling efforts by Cosme et al. (2017b) provide the first LCIA model to address marine eutrophication. This model provides spatially explicit fate factors describing N emissions to soil, freshwater, and marine systems. The CFs are derived from the NEWS2 model (Mayorga et al., 2010), which provides inland soil and freshwater fate for dissolved inorganic N at a watershed scale. In the Cosme et al. work, the residence time of N in the receiving marine systems is modelled as a function of advection and denitrification.

The NEWS2 model, and thus its output, operates at the basin scale, even though some inputs are added at the sub-basin scale and some sub-basin processes are modeled explicitly. While natural and anthropogenic sources of N are distinguished at the sub-basin level, this modeling refinement is only for source contribution analysis in the final model output. In the LCIA application of Cosme et al. (2017b), the NEWS2 soil export fraction is a combination of natural and agricultural soil.

The NEWS2 model does not distinguish between various emission fates within a watershed. For large watersheds, such as the Lower Mississippi River Basin in the US, there could be significant differences in fate of emissions at the headwaters compared to the outlet of a watershed. Cosme et al. (2017a) and Morelli et al. (2018) identify other mechanistic models for N fate (e.g., AQUATOX, IMAGE-GNM), but both authors note the high data demands and computational intensity of these models across large spatial domains, which make global applications difficult.

In order to translate air emissions to marine impacts, the team used the source-receptor model of Roy et al. (2012), which models deposition of NH_x and NO_x to land resulting from air emissions at $2^\circ \times 2.5^\circ$ grid cell resolutions. Roy et al. used GEOS-Chem (Bey et al., 2001), a global 3-D model for atmospheric chemistry and emission transport, to create these source-receptor matrices based on meteorological data for 2005, which represents average annual temperature and precipitation data from 1961 to 1990.

Given this background, the marine eutrophication model proposed for incorporation into TRACI is the combination of the GEOS-Chem air transport model and the NEWS2 hydrological model, as implemented in Bjorn et al. (Bjørn et al., 2020) and further discussed in Payen et al. (in preparation). Coupling these two models to the NEWS2 basins and marine systems of Cosme et

al. makes it possible to trace an emission of N to air, following it to soil, freshwater, and finally to a marine system.

2.2.2 Fate Calculations

In the LCIA context (Rosenbaum et al., 2007), the fate factor (FF) describes the fate of an emission in a given compartment and location, tracing the emitted substance to its receiving compartments (in the case of multi-media transfer) and locations. This is commonly called the midpoint characterization factor (CF). In the case of a single compartment, FF can be expressed as a matrix, as in Equation 2 **Error! Reference source not found.**:

$$FF_{e-r}, \text{ where } ff_{e-r,ij} = \frac{\text{mass substance in compartment } r, \text{ location } j}{(\text{mass emitted to compartment } e, \text{ location } i / \text{time})} \quad \text{Eq. 2}$$

Note that FF is often represented in units of time (e.g., kg/day). In the case of a unit emission, FF can describe the steady-state mass in the receiving compartment. In a spatially explicit framework, a single emission location, *i*, may lead to a steady-state mass in one, many, or all other compartments, *j*.

FF_{e-r} can be used directly if the compartments and locations modeled match the needs of the analysis. In the case of this work, the following fate factors were used directly:

- freshwater eutrophication due to phosphorus emissions to freshwater (Helmes et al., 2012), with a 0.5° × 0.5° grid cell resolution
- marine eutrophication in large marine ecosystems (LMEs) due to nitrogen emissions to LMEs, inland freshwater, or inland soil (Cosme et al., 2017a), with an emission resolution of NEWS2 hydrological basins (Mayorga et al., 2010), and an impact resolution of LMEs (Sherman and Alexander, 1986).

However, other fate factors require manipulation prior to use in a eutrophication context:

- freshwater eutrophication due to emissions to soil are calculated based on 10% factor soil-applied nutrients similar to the ReCiPe approach.
- Marine eutrophication due to nitrogen emissions to air, followed by deposition on LMEs.
- Marine eutrophication due to air emissions of NH_x and NO_x (Roy et al., 2012), with a 2° × 2.5° grid cell resolution, followed deposition to freshwater or land, and subsequent transfer to LME,

Coupling this air quality model to the NEWS2 basins and marine eutrophication models makes it possible to trace an emission of N substances to air, following it to soil, freshwater, and finally a marine system. This coupling is briefly described below, and it has been documented more thoroughly in Bjørn et al. (2020).

A marine eutrophication fate factor, FF_{NEWS2_comp-LME}, relates emissions to a given NEWS2 basin compartment (freshwater, soil, or LME) and location, *i*, to an increased mass of nitrogen in a corresponding LME, *j*. An air fate factor, FF_{air-ground}, relates emissions to air in location *i* to a deposition of that substance in location *j*. The team used an intermediate matrix, a spatial transform (ST), to relate the geographic fractions of the cells modeled in the air model to the geographic (and

compartment) fraction of the NEWS2 basin. In the matrix $ST_{\text{air-NEWS2}}$, each element of the matrix describes the fraction of the corresponding air cell (rows) to the corresponding NEWS2 basin's appropriate compartment (freshwater or soil). Therefore, the matrices may be multiplied as shown in Equation 3.

$$FF_{\text{air-LME}} = FF_{\text{air-ground},ij} \times \frac{\text{moles } N}{\text{moles subst.}} \times ST_{\text{air-NEWS2_comp}} \times FF_{\text{NEWS2_comp-LME}} \quad \text{Eq. 3}$$

Where the resulting matrix, $FF_{\text{air-LME}}$, describes the steady-state mass of nitrogen in the LMEs due to an emission to air. Note that there is a stoichiometric conversion of the NH_x or NO_x in the air model to the nitrogen required by the marine model.

The output of the approach described above includes matrices of the fate data, linking all emission locations and compartments to all receiving locations and compartments. To make this data usable in an LCIA model, it is necessary to aggregate these locations to political boundaries (Section 2.2.3).

2.2.3 Aggregation

The goal of geospatial aggregation using a geographic information system (GIS) is to provide representative CF values for a set of target polygons. In this work, the target polygons are (1) the US, (2) its states, and (3) its counties. The native CF spatial system is defined by the underlying fate models; for air emissions, the $2^\circ \times 2.5^\circ$ grid of Roy et al (2012); for freshwater emissions of P, the $0.5^\circ \times 0.5^\circ$ grid of Helmes et al. (2012); for freshwater and soil emission of N, the NEWS2 basins of Cosme et al. (2017); and for emissions of N to the marine system, the LMEs of Cosme et al. (2017). The team used likelihood data as described below to calculate a representative aggregated value of the target polygon.

The output of the aggregation calculation is a set of CFs for the target polygons.

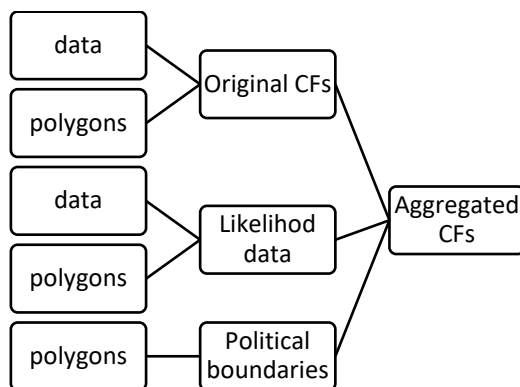


Figure 12. Schematic of the Characterization Factor Aggregation Process

The process of aggregating CFs into political units (e.g., states, counties) requires likelihood data that represent a generalized type of land use activity on each land parcel (here, a grid cell). For each impact category and substance, there are (potentially) three sets of likelihood data required. One set is for agricultural activity, which includes applications of fertilizer (or directly modeled

emissions from agriculture) as the proxy, or a representative literature value for agricultural activity inside the target polygon. The second set is for nonagricultural activity, which includes industrial emissions, transport emissions, and general population (e.g., human waste) emissions. Finally, an unknown or “general” likelihood dataset is needed for those cases in which the source of an emission is unknown or is being used for the background of an LCI model. The general likelihood dataset should be the sum of the agricultural and nonagricultural likelihood values. See the Electronic Supplementary Material (ESM 1) for a summary of phosphorus and nitrogen data used for the likelihood datasets.

When aggregating characterization factors, a few points need to be kept in mind. First, CFs, target, and likelihood polygons need not be co-located, nor do they need to have the same spatial extent. For example, the CF grid does not need to overlap with the political grid, which does not need to overlap with the likelihood grid. Furthermore, the CF grid may extend beyond the borders of the target polygon. Third, likelihood data may be considered extensive or intensive, where extensive values depend on the area considered and intensive values are independent of the area. For example, population (number of people) or fertilizer application (kg N applied) would be extensive values whereas population density (people/km²) or fertilizer application rate (kg N/ha) would be intensive values. Note that CF values are intensive values. If two equal-area, equally weighted CF polygons are contained in a single target polygon, the resulting target polygon will be the arithmetic average (not the sum) of the two CFs. Finally, there are a few scenarios in which CF and likelihood values are assumed to be zero. When target polygons are missing CF data, the CF is assumed to equal zero. This situation occurs if a target contains areas outside those covered by the CF model. For example, the 0.5° × 0.5° freshwater grid does not extend to some coastal border areas of states (i.e., where it is difficult to model hydrologic flow), such as the southern tip of Maryland. In the case of Maryland, since other parts of the state have CF and likelihood data, there will still be a state-level CF value for the state. Another scenario for assumed-zero values is when an area has no likelihood polygons; in this case, the likelihood value is assumed to be zero. This could occur, for example, if a state has areas without fertilizer application data; in this scenario, it would be assumed that no fertilizer is applied in the polygon.

The type of likelihood data also affects the aggregated CF value. For example, consider a target (state) polygon containing several native CF polygons. Suppose that the rural area of the state has high agricultural activity and happens to have high CF values. Suppose that the populous area of the state, with low agricultural activity, tends to have low CF values. An arithmetic average of CFs in the state would yield an overall state CF with a value between the low and high values. However, if agricultural activity is introduced for the likelihood dataset, then an aggregated CF for the entire state would have a higher CF than the arithmetic average because it would be weighted by the higher-value CFs in the agricultural areas of the state.

Figure 13 illustrates the GIS process required to calculate the aggregated CF using the native CF (CF), target (T), and likelihood (L) datasets. Areas *S*, *T*, and *L* can exhibit a variety of relationships with respect to spatial extent and scale. For example, while Roy air cells have global area coverage and tend to be similar in size to counties in the Eastern United States, Roy freshwater grid cells are restricted to continents and tend to be smaller than U.S. counties. It is also possible to have native CFs (i.e., CF data) extend beyond the target polygons, as would be the case with an air emission CF grid in this study.

To perform the CF calculations in a geodatabase, a spatial intersect of CF, T, and L polygons is needed. This results in a table of new polygons, or set N , created by the intersect.

Next, aggregated values are calculated for each target zone t , according to Equation 4 for intensive L data (already normalized to area) and Equation 5 for extensive L data (which requires dividing L data by total L area):

$$CF_{target} = \sum_{n \in t} CF_n * \frac{A_n * L_n}{\sum_{j \in n} A_j * L_j} \quad \text{Eq. 4}$$

$$CF_{target} = \sum_{n \in t} CF_n * \frac{A_n * L_n * \left(\frac{A_n}{A_{Ln}}\right)}{\sum_{j \in n} A_j * L_j * \left(\frac{A_j}{A_{Lj}}\right)} \quad \text{Eq. 5}$$

where:

n are intersection polygons;

CF_n and L_n are the source and likelihood data corresponding to polygon n created by the spatial intersect; and

A_x is the area of polygon x .

Each polygon N_i inherits information from the parent polygons of which it is a member. In the example in Figure 13, the red polygon $N_{CF=9112, t=Washington \text{ County VA}, l=50052}$ (here called N_{ex}) represents an area in target T Washington County, VA where the CF of CF_{9112} and likelihood data L_{50052} apply. N_{ex} has an area smaller than any of its parent polygons, and thus fractional areas can be calculated as needed. For example, if L holds total fertilizer application (in total kg, an extensive variable), then the fertilizer applied in N_{ex} is equal to $A_{N_{ex}} / A_{L50052}$, where A represents the area of each. If L holds fertilizer rate (area density in kg/m²), then the fertilizer rate in the polygon N_{ex} is equal to the fertilizer rate in L_{50052} . A given polygon, N_i , can be larger than some of the parent polygons, but cannot be larger than all three parent polygons, since it was created from a spatial intersect.

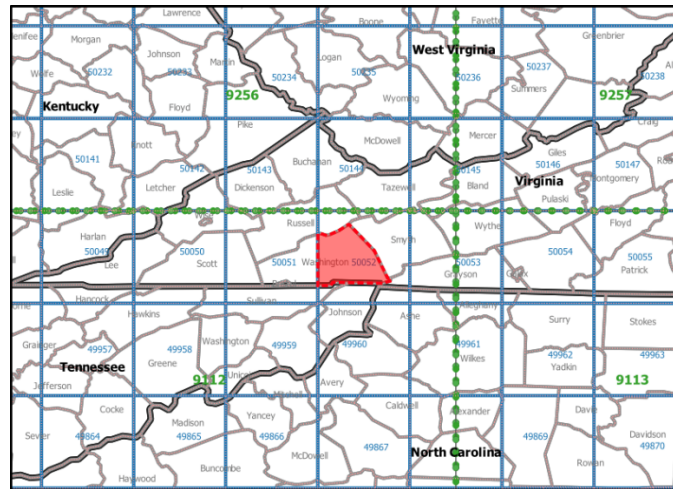


Figure 13. Spatial Intersect of Native-Scale Characterization Factors

Provided by fate models (CFs, green grid), target polygons (T, grey lines) for which a representative CF value is desired, and likelihood polygons (L, blue lines) with data used to inform the aggregated CF calculation. Different likelihood data (e.g., agricultural, population-based, etc.) may be used to calculate different aggregated CFs.

2.2.4 Overall Calculations

Each set of aggregated characterization factors for eutrophication potential is specified by the following:

- substance flow
- compartments: emission compartment, any intermediate transfer compartment, and receiving compartment
- aggregation resolution
- likelihood data: agricultural, nonagricultural, and general

Table 4 summarizes the characterization factors calculated in this study. Each combination of impact category with the aforementioned components leads to nine sets of CFs per combination, as can be seen by the nine columns in the table and is apparent for freshwater eutrophication (one substance, one fate model). There are more combinations for marine eutrophication, as Cosme et al. (2017a) lists three emission compartments and one substance. Given three target aggregation resolutions for each of the three likelihood datasets, there are 27 sets of CFs. Finally, in the coupling of air emissions to marine eutrophication, there are two substances, and the same aggregation and likelihood data. This creates 18 sets of CFs.

Table 4. Summary of Characterization Factors Calculated in this Work

Each row represents a set of CFs (where the set is comprised of all list elements in each row). For example, for freshwater eutrophication, there are 9 sets of CFs: 1 substance ' 3 resolutions ' 3 likelihood datasets. Note that global factors are also calculated, using likelihood data consistent with Payen et al. (in preparation).

Eutrophication Category	Substance Flow	Emission Compartment	Emission Resolution	Transfer Compartment	Receiving Compartment	Receiving Resolution	Aggregation Resolution	Likelihood Data
Freshwater	P	Freshwater	0.5° ' 0.5° cells	-	Freshwater	0.5° ' 0.5° cells	county, state, nation	US EPA
Marine	N	LME	NEWS2 Basins	-	LME	LME	county, state, nation	US EPA
Marine	N	Freshwater, Soil	NEWS2 Basins	-	LME	LME	county, state, nation	US EPA
Marine	NHx, NOx	Air	2° ' 2.5° cells	-	LME	LME	county, state, nation	US EPA
Marine	NHx, NOx	Air	2° ' 2.5° cells	Freshwater	LME	LME	county, state, nation	US EPA
Marine	NHx, NOx	Air	2° ' 2.5° cells	Soil	LME	LME	county, state, nation	US EPA

2.3 RESULTS

Freshwater and marine eutrophication CFs can be found in the Electronic Supplementary Material (ESM 2). The data and models used to develop these calculations are available on EPA's Environmental Dataset Gateway (<https://edg.epa.gov>) and GitHub project repository (<https://github.com/USEPA/lcia-eutrophication>). Characterization factors will also be made available on the FLCAC (<https://www.lcacommons.gov>). Maps have been included in this section (Figure 15-Figure 18) to facilitate greater understanding of the characterization factors and their spatial patterns. These maps focus on the U.S. context; see Payen et al. (in preparation) for global maps.

Table 5 summarizes the statistical dispersion of the CF values, according to the interdecile range. The freshwater and marine eutrophication impact categories are further separated into their emission compartments with corresponding substance flows. These are then each linked to a likelihood category (i.e., agricultural, nonagricultural, or general) and aggregation target (i.e., state or county).

Table 5. Eutrophication Impact Categories, Likelihoods, and Geographic Aggregation Targets, With Associated Characterization Factors Summarized Across the Interdecile Range (IDR)

Agric = agricultural; Nonagric = non-agricultural; IDR = the difference between the first and ninth deciles, i.e., 0.1 and 0.9.

Eutrophication Category	Emission Compartment	Flow	Likelihood	Aggregation Target	0.1	0.5	0.9	IDR / median
Freshwater	Freshwater	P	Agric	State	0.045	0.15	1.2	7.3
Freshwater	Freshwater	P	Agric	County	0.0071	0.12	0.61	5.2
Freshwater	Freshwater	P	General	State	0.045	0.15	1.3	7.9
Freshwater	Freshwater	P	General	County	0.0071	0.12	0.6	5.2
Freshwater	Freshwater	P	Nonagric	State	0.045	0.16	2.9	18
Freshwater	Freshwater	P	Nonagric	County	0.0073	0.12	0.63	5.4
Marine	Freshwater	N	Nonagric	State	0.025	0.05	0.088	1.3
Marine	Freshwater	N	Nonagric	County	0.019	0.05	0.082	1.3
Marine	Soil	N	Agric	State	0.0075	0.012	0.041	2.9
Marine	Soil	N	Agric	County	0.0059	0.0083	0.028	2.7
Marine	LME	N	Nonagric	State	0.12	0.19	0.2	0.44
Marine	LME	N	Nonagric	County	0.073	0.19	0.19	0.63

Eutrophication Category	Emission Compartment	Flow	Likelihood	Aggregation Target	0.1	0.5	0.9	IDR / median
Marine	Air	NH ₃	Agric	State	0.07	0.095	0.16	0.96
Marine	Air	NH ₃	Agric	County	0.071	0.094	0.16	0.89
Marine	Air	NH ₃	General	State	0.07	0.096	0.16	0.95
Marine	Air	NH ₃	General	County	0.071	0.094	0.16	0.89
Marine	Air	NH ₃	Nonagric	State	0.07	0.097	0.16	0.89
Marine	Air	NH ₃	Nonagric	County	0.071	0.094	0.16	0.89
Marine	Air	NO _x	Agric	State	0.071	0.094	0.15	0.8
Marine	Air	NO _x	Agric	County	0.068	0.09	0.13	0.65
Marine	Air	NO _x	General	State	0.07	0.092	0.15	0.83
Marine	Air	NO _x	General	County	0.068	0.09	0.13	0.64
Marine	Air	NO _x	Nonagric	State	0.069	0.09	0.15	0.85
Marine	Air	NO _x	Nonagric	County	0.068	0.09	0.13	0.65

State and county CF values vary based on the spatial nature of the underlying fate models. For example, the freshwater eutrophication model has values that vary by several orders of magnitude. In addition, the model may return a CF value of zero for arid regions of the country. The marine eutrophication model for emissions to freshwater, soil, and LME shows variations across watersheds of up to a factor of ten. In contrast, CFs for NH₃ emissions show less drastic variations, for example, whereas NO_x emissions may vary by 20% from cell to adjacent cell. Overall, there is greater variation among state values for freshwater eutrophication than for marine eutrophication emissions to soil, freshwater, or LMEs. Both groups exhibit greater variation than marine eutrophication CFs from airborne emissions.

The residence time also affects the spatial variation of fate factors and CFs. The freshwater fate factor for phosphorus varies proportionally with the freshwater residence time. According to the underlying freshwater model, greater hydraulic residence time (especially in lakes where the phosphorus may not settle out of the water column) leads to a larger steady-state mass of phosphorus in the system.

Aggregation by likelihood can strongly influence an overall value. However, the variation is typically modest, due to relative sizes of the fate cells and aggregation targets, as well as the inherent fate variation, as previously discussed. Table 6 describes the variation at the state- and country-level.

Table 6. Statistical Variation Across the CONUS Aggregated Characterization Factors, Using Agricultural, Nonagricultural, and General Likelihoods for State-Level and Country-Level Geographic Targets, as Appropriate

Variation is calculated as the difference of the maximum and minimum value (i.e., 0 for CONUS), divided by the median. Note: variation across the following categories could not be calculated, as only two of the three (agricultural and nonagricultural) likelihood values were included: Marine eutrophication; N to soil, N to freshwater, and N to LME.

Impact category and substance	Lower 48 states: max variation	Lower 48 states: median variation	Lower 48 country: total variation
Freshwater eutrophication; P to freshwater	15.4	0.21	1.7
Marine eutrophication; NH ₃ to air	0.15	0.02	0.23
Marine eutrophication; NO _x to air	0.68	0.20	0.22

As seen in Table 5 and Figure 14, differences in likelihood values influence the resulting characterization factors. For example, Illinois (Figure 14) shows a 15-fold difference between the agricultural and nonagricultural CFs for freshwater eutrophication (see also top right panel of Figure 16). The higher CF values associated with the relatively dense population in the Chicago area, compared with the rest of the state, contribute to a nonagricultural CF that is 15.4 times greater than the agricultural CF for the entire state of Illinois. For completeness, all relevant likelihood factors are included for all impact categories and spatial resolutions. Users of TRACI that know the nature and location of an emission may select the appropriate likelihood and location.

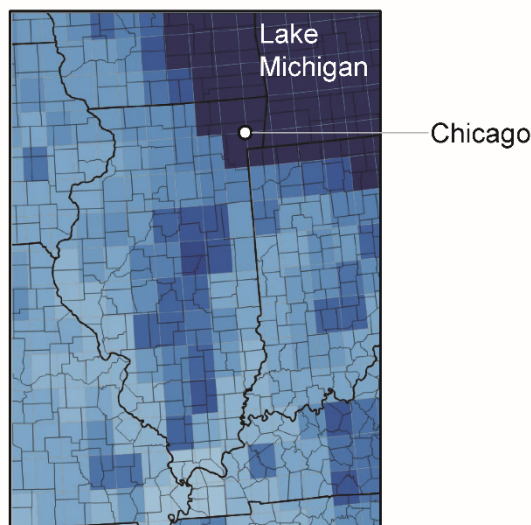


Figure 14. Freshwater Eutrophication Grid for Illinois

Chicago, a population center, is in upper northeastern Illinois. This section of the state is also in the Great Lakes watershed, as seen by the higher (darker) CF values, whereas the remainder of the state is in the Mississippi watershed (with lower, i.e., lighter, CF values).

2.3.1 Freshwater Fate Factors and the Great Lakes

As the Great Lakes are a unique system in the United States and the world, the effect of large lakes on the freshwater eutrophication factor deserves some mention. As noted in Figure 14, there can be sharp gradients in freshwater fate factor from one location to the next, as one moves across watershed boundaries. The following discussion addresses the specific examples from the Great Lakes, but these trends can be observed anywhere there is a lake with a hydraulic residence time significantly larger than nearby systems.

Even among the Great Lakes, there can be a significant difference in hydraulic residence time, and thus the freshwater fate factor. The hydraulic residence time of the Great Lakes ranges from centuries (Superior: 191 years, Michigan: 99) to decades (Huron: 22) to a handful of years (Erie: 2.6, Ontario: 6) (US EPA, 2015). These high hydraulic residence times have a noticeable impact on characterization factors as can be seen for characterization factors for states that drain into the Great Lakes (and specifically Lake Superior and Lake Michigan) in Figure 15.

Therefore, the native (non-aggregated) freshwater fate factors for emissions upstream of Lake Superior or Lake Michigan will be a factor of 20-120 or more greater than the factors for Lake Erie or Lake Ontario. At the aggregated state (and county) level, the likelihood data also play a role. For example, although the Laurentian Watershed of northeast Minnesota drains into Lake Superior, there is relatively little agricultural activity or population in the watershed. Therefore, the state value for Minnesota is more heavily influenced by the freshwater fate factor for the Mississippi River basin (in which native factors are on the order of <0.1 yr). The state of Michigan, however, has areas that drain into Lake Superior and Lake Michigan, and the latter also has agricultural activity and high population density. Therefore, the overall state value for Michigan directly reflects the high freshwater fate factors for Lake Michigan. As a final example, Ohio has dense population and agricultural activity in both the Great Lakes basin and in the Mississippi River basin. Since the fate factors for Lake Erie and Lake Ontario are at least a factor of 20 smaller than the fate factors for the upstream Great Lakes, the state value for Ohio (in the Lake Erie watershed) is modest relative to Wisconsin or Michigan (in the upstream lake watersheds).

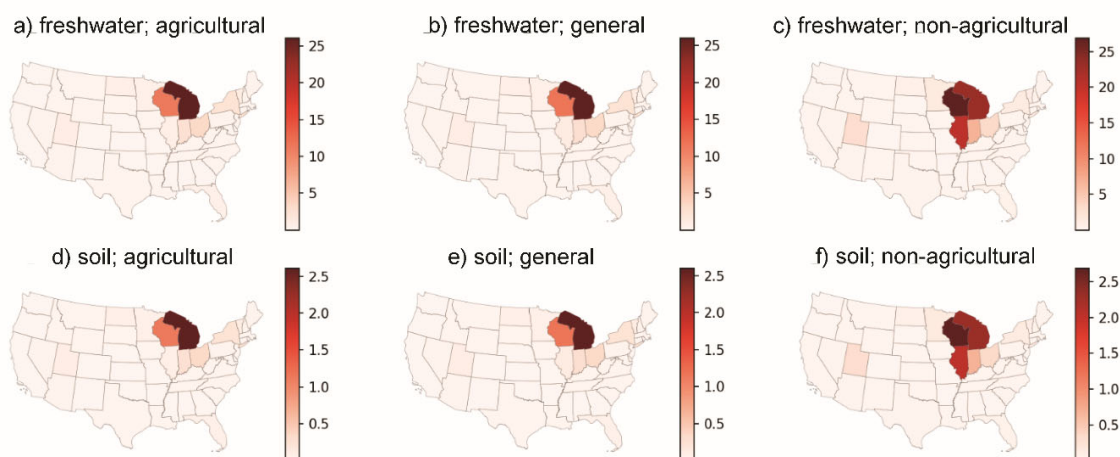


Figure 15. Maps of Characterization Factors for Freshwater Eutrophication to Freshwater (Top Row) and Freshwater Eutrophication to Soil (Bottom Row)

Individual map titles are arranged as [emission compartment]; [likelihood category].

2.3.2 Large Marine Ecosystems (LMEs)

Marine characterization factors are shown in Figure 16. Note that the color gradient is variable for each map, and that CFs are highly variable when comparing individual maps within the figure.

In particular, the characterization factors for emissions of nitrogen directly to marine systems are shown in the center column of the top row. There is a 50-km buffer around each LME, based on the assumption that only emissions within 50 km would be considered direct emissions to a large marine ecosystem. States located outside of this buffer therefore have no fate factor; only states within the 50-km buffer are included. The buffer distance, however, introduces the possibility that a state like Arizona, for example, could have an emission directly to the Gulf of California. The Gulf of California is a relatively enclosed LME, as compared to LMEs off the Atlantic or Pacific coasts. As such, the LME fate factor for the Gulf of California (0.65 yr) is three times greater than the coastal LME of the California Current (0.19 yr)., Therefore, the state CF value for Arizona is correspondingly high.

In the bottom row of Figure 16, air emissions which deposit and influence the Hudson Bay watershed result in higher characterization factors for marine eutrophication than most other air emissions. CFs are also high along the Atlantic coast. These trends are driven by the atmospheric fate modeling of Roy et al. (2012) and the LME fate model of Cosme (2017b). First, atmospheric transport in the United States generally occurs from west to east. All else being equal, a lower fraction of emissions on the West Coast will deposit in the Pacific LMEs than will emissions on the East Coast deposit on Atlantic LMEs. Secondly, the LME fate factor for the Hudson Bay is higher (1.4 yr) than for other LMEs in North America (the California Current, off the U.S. West coast = 0.19 yr, the Gulf of Mexico = 0.16 yr, the Southeast U.S. Continental Shelf = 0.07 yr, the Northeast U.S. Continental Shelf = 0.19 yr, and the Newfoundland-Labrador Shelf = 0.19 yr). Thus, the upper Midwest and Northeast United States have the highest fractions of emissions depositing in an LME (the Hudson Bay) with a higher FF than other LMEs. The use of a fate factor for direct emissions to LMEs should be used with caution, especially in the case of a state-wide value. There should be good evidence to support the fact that the emission occurs on the coast (or within 50 km of the coast).

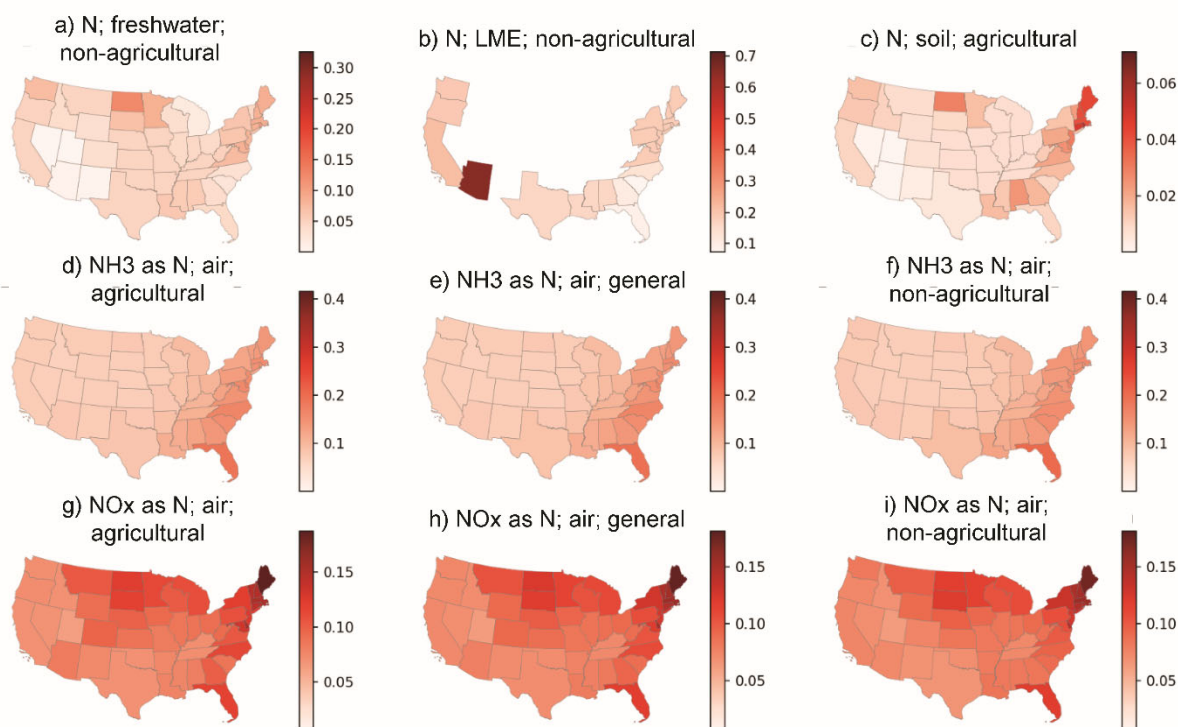


Figure 16. Maps of Characterization Factors for Marine Eutrophication Emissions to Freshwater, LME, and Soil (Top Row) and Air (NH3 and Nox; Middle and Bottom Rows)

Individual map titles are arranged as [substance emitted]; [emission compartment]; [likelihood category].

2.3.3 Application Case: N and P Inventories

To demonstrate the application of CFs in the United States, nitrogen and phosphorus emissions available from the U.S. Environmental Protection Agency (EPA) (Clark et al., 2018; Sabo et al., 2019) were combined in a series of maps which are discussed below. This dataset was chosen because they represent the most comprehensive publicly available N and P inventories for the continental United States (CONUS).

For both Figure 15 and Figure 16 it is important to realize that the CFs are intensive factors (meaning that they are independent of land area) whereas the inventory and impact are both dependent upon land area. This may result in some states (e.g., Texas and California showing up as being highly impacted when, on a land mass basis, they may not be as impacted as other states.

For illustration, the following categories were mapped: human waste, agricultural fertilizer, and urban fertilizer. It should be noted that fertilizer is applied to the soil with the intention of enriching the land, so most of the applied fertilizer is expected to have beneficial impacts on the terrestrial environment by improving the productivity of the crops and/or livestock that are under agricultural development or the urban spaces that benefit from the fertilizer. Runoff is intentionally managed to be at a minimum since runoff represents economic losses to the agricultural operation. Precise calculations of fertilizer runoff are difficult to automate since runoff depends on the slope of the

land, the condition of the soil at the time of application, the rainfall after application, etc. Therefore, a proxy estimate of 10% runoff is incorporated into the CFs of soil application to represent the estimated fertilizer runoff to the freshwater.

Figure 17 illustrates the application case of P at the state-level. The center column are the CFs which were previously described in Figure 14. The P inventory in the leftmost column reflects heavy use of agricultural fertilizer, as expected for the agricultural regions highlighted. Urban fertilizer inventory was difficult to characterize for states with significant agricultural production since in many cases it is not known whether fertilizer purchased is used for agriculture or urban fertilizers, so the left column center figure may have a disproportionate amount of fertilizer attributed to urban applications. These states tend to only report agricultural fertilizer use and therefore rely on county population to scale fertilizer application on non-agricultural land (Gronberg and Spahr 2012). Human waste can be seen in the bottom left figure and is consistent with expectations of population totals in each of these states.

The right-hand column includes the inventory and CFs multiplied for total potential impact. This shows that the influence of the CFs was much more dominant than the variations in inventory per state and that the highest eutrophication would be expected to be resulting from emissions from states which feed into the two Great Lakes with the longest retention times (Lake Superior and Lake Michigan).

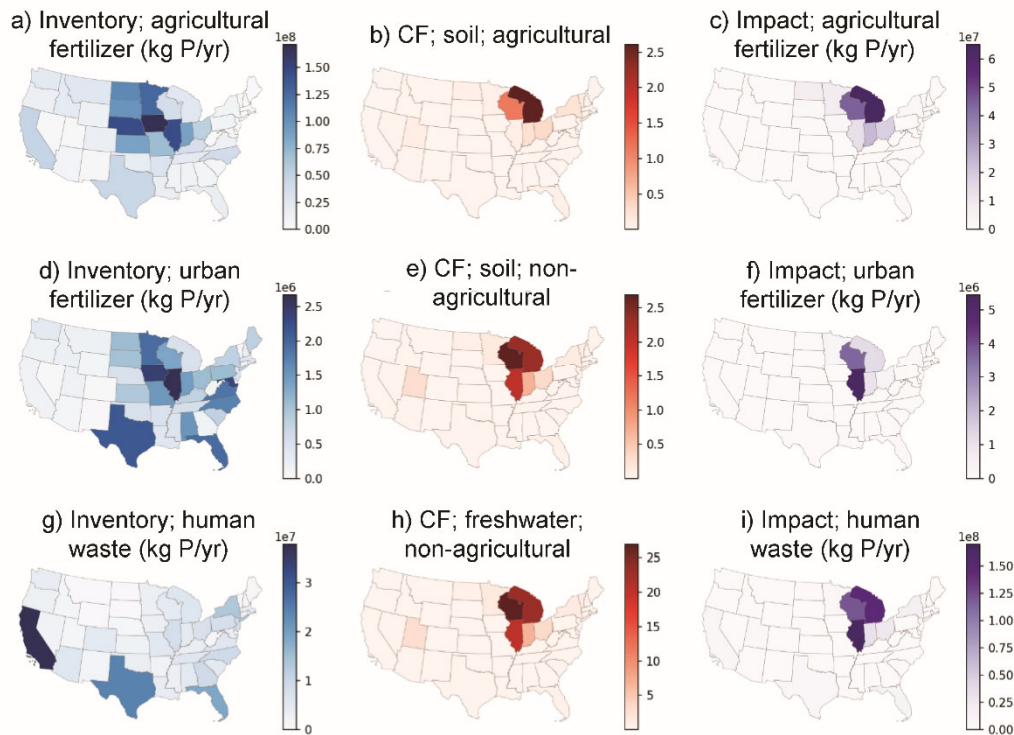


Figure 17. Case Application for P at the State Level: Agricultural Fertilizer (Top Row), Urban Fertilizer (Middle), and Human Waste (Bottom)

Columns represent the national inventory at the state level (left column), the corresponding characterization factor (center column, CF) and the total impact (right column).

Figure 18 illustrates the application case of N at the state-level. The leftmost column (blue) shows the inventory mapping, which follows the expected patterns. For example, human waste is centered around areas of high urban density, while urban fertilizer is found in areas where more suburban practices of lawn fertilizer application, golf centers, and other landscaping trends are popular. Agricultural fertilizer, pesticide, and herbicide use are more prevalent in the Midwestern states, consistent with livestock and grain farmland. The middle column of Figure 18 (red) shows the CFs developed in this research, corresponding to the N inventories (blue). The combination of the inventories and the CFs is shown in the rightmost column (purple) and represents the total potential impact.

For agricultural and urban fertilizers, the largest impacts are dependent upon both the combination of high inventories and the higher level of CFs which are prevalent for emissions in the Midwest and Atlantic states. Human waste impacts are highest in areas of high population, including California, Texas, New York, and a few other coastal states.

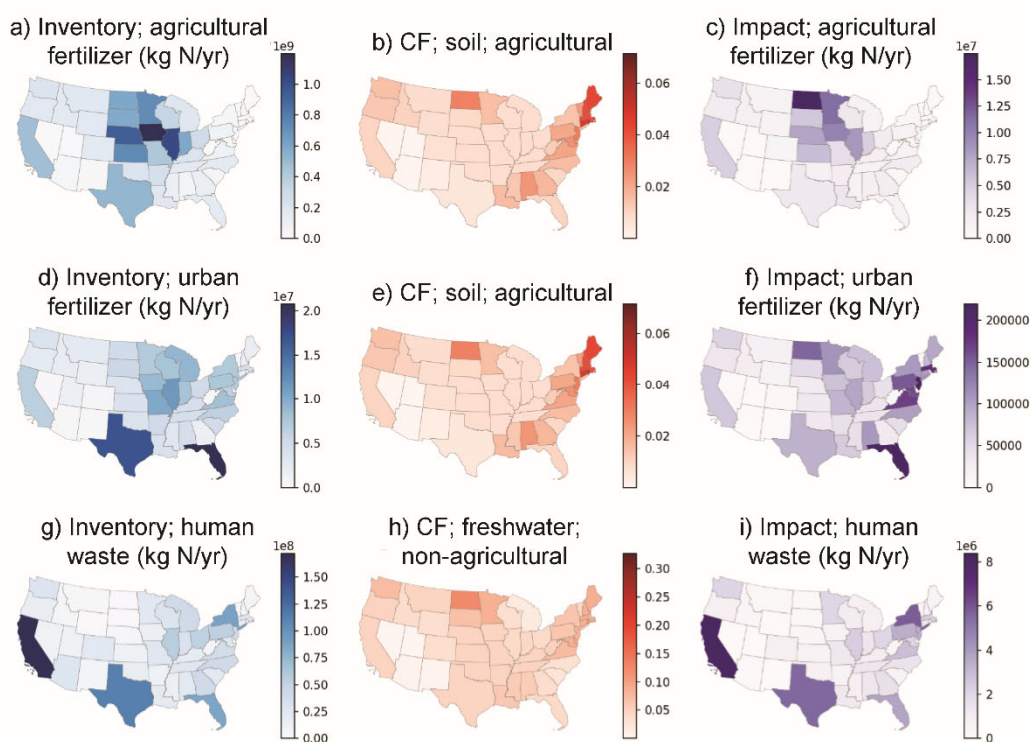


Figure 18. Application Case for N at the State Level: Agricultural Fertilizer (Top Row), Urban Fertilizer (Middle), and Human N Waste (Bottom).

Columns represent the national inventory at the state level (left column), the corresponding characterization factor (center column, CF), and the total impact (right column).

The application case presented here is meant to display the relative influence of CFs and inventories and their level of influence on a CONUS example. The team did not attempt to capture all of the emissions of N or P, but simply demonstrated the use of the CFs developed earlier in the paper.

Although these results are specific to this application, a few states' CFs significantly dominated the calculation of the total potential eutrophication impact. As mentioned previously, the retention times of the Great Lakes were a significant factor in these results. Globally, any large water body (e.g., the African Great Lakes or the Baltic Sea) could similarly influence a country-level CF. The marine eutrophication CFs and the inventories provided a more balanced influence on the potential impact within the US.

2.4 DISCUSSION

This paper builds upon the eutrophication model review presented in Morelli et al. (2018) and its set of recommendations for improvement to the eutrophication potential category in LCA. In the current work, the team presents a preliminary effort to spatially differentiate freshwater and marine eutrophication characterization factors to improve the scientific robustness of the impact category. The freshwater eutrophication fate model was provided by Helmes et al. (2012) while input to the marine eutrophication was provided by Cosme et al. (2017b) with characterization factors derived from NEWS2. In keeping with standard LCIA practice, the team used the assumption that freshwater systems are P-limited and marine systems are N-limited. This assumption should be updated in future work, as co-limitation can occur in marine and freshwater systems.

Fertilizer sources of N and P played a key role in the application study used to evaluate the CF spatial patterns across the country. While some of the eutrophication impacts could be tracked according to agricultural production and lawncare practices, this wasn't the case everywhere and fertilizer inventory in some locations were scaled by community population (as was the case for non-agricultural fertilizer use in a predominantly agricultural region). This secondary population-based estimate introduces uncertainty into the dataset; this uncertainty could be quantified in future efforts, assuming the fertilizer estimates are not improved directly.

In the previous research of Morelli et al. (2018), the authors suggested several improvements to the eutrophication categories, beginning with the separation of freshwater and marine models. Recommendations that would require further research include: (1) refining the spatial and temporal resolutions of the fate and transport models; (2) further developing the residence time estimates for freshwater and marine systems; (3) integrating soil data, such as soil type, to better estimate application and run-off of land-applied nutrients; (4) using hydrodynamic models to expand the marine coverage beyond 66 LMEs. Throughout these improvements, researchers must continue to optimize the increasing spatial data resolution with the practical usability of available environmental models.

2.5 CONCLUSIONS

In this research, the team developed site-specific freshwater and marine eutrophication CFs at a fine scale for the United States and at a larger scale for the rest of the globe. The team also applied freshwater and marine eutrophication factors for the United States for specific examples of agricultural fertilizers, urban fertilizers, and human waste.

Additional research is needed to provide similarly resolved CFs for the entire globe, which would require an expansion and deepening of the underlying data and fate and transport models. Environmental models currently available at higher-level resolutions tend to lack compatibility or

feasibility with the LCA framework, often containing more detailed data than LCA can currently incorporate. However, these models could also replace common assumptions, such as the 10% runoff currently assumed across eutrophication models. Other additions to data could allow for further spatial refinement as well as temporal refinement, especially for agricultural crop cover or seasonal scenarios. Endpoint modeling could also be incorporated once the science has reached a sufficient level of sophistication. Enhanced computing capabilities will allow for even better eutrophication modeling in LCA and extend to other impact categories in the future.

In addition to this method, there is also a eutrophication method under development by the United Nations Environment Program (UNEP) (2019). The co-authors of this paper are involved in the UNEP effort in various capacities and thus have insight into the compatibility of the two methods. While the TRACI method focuses on the contiguous United States and uses data at a finer resolution for the United States while also incorporating global data to produce country-level CFs, the UNEP method offers global CFs at a country resolution, and it uses the same likelihood data as TRACI. In addition, TRACI stops at the midpoint along the cause-effect chain, whereas the UNEP method continues along the chain to the endpoint, thus introducing another level of uncertainty. Both methods are spatially-explicit eutrophication methods.

2.6 REFERENCES

- Bare J (2011) TRACI 2.0: The Tool for the Reduction and Assessment of Chemical and other environmental Impacts 2.0. *Clean Technol Envir* 13:687-696. <https://doi.org/10.1007/s10098-010-0338-9>
- Bare JC, Norris GA, Pennington DW, McKone T (2002) TRACI: The Tool for Reduction and Assessment of Chemical and other environmental Impacts. *J Ind Ecol* 6:49-78. <https://doi.org/10.1162/108819802766269539>
- Bey I, Jacob DJ, Yantosca RM, Logan JA, Field BD, Fiore AM, Li Q, Liu HY, Mickley LJ, Schultz MG (2001) Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *J Geophys Res-Atmos* 106:23073-23095. <https://doi.org/10.1029/2001JD000807>
- Bjørn A, Sim S, King H, Margni M, Henderson A, Payen S, Bulle C (2020) A comprehensive planetary boundary-based method for the nitrogen cycle in life cycle assessment: Development and application to a tomato production case study. *Sci Total Environ* 715:1-16. <https://doi.org/10.1016/j.scitotenv.2020.136813>
- Clark C, Sabo R (2018) National Inventory of Reactive Nitrogen. U.S. Environmental Protection Agency: Environmental Dataset Gateway. <https://doi.org/10.23719/1418976>
- Cosme N, Jones MC, Cheung WWL, Larsen HF (2017a) Spatial differentiation of marine eutrophication damage indicators based on species density. *Ecol Indic* 73:676-685. <https://doi.org/10.1016/j.ecolind.2016.10.026>
- Cosme N, Mayorga E, Hauschild MZ (2017b) Spatially explicit fate factors of waterborne nitrogen emissions at the global scale. *Int J Life Cycle Assess* 23:1286-1296. <https://doi.org/10.1007/s11367-017-1349-0>

- Fekete BM, Vörösmarty CJ, Grabs W (2002) High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem Cy* 16:15-1-15-10. <https://doi.org/10.1029/1999GB001254>
- Gronberg JM, Spahr NE (2012) County-level estimates of nitrogen and phosphorus from commercial fertilizer for the Conterminous United States, 1987-2006. U.S. Geological Survey. <https://doi.org/10.3133/sir20125207>
- Helmes RJK, Huijbregts MAJ, Henderson AD, Jolliet O (2012) Spatially explicit fate factors of phosphorous emissions to freshwater at the global scale. *Int J Life Cycle Assess* 17:646-654. <https://doi.org/10.1007/s11367-012-0382-2>
- Lehner B, Linke S (2015) Derivation of global river network attributes including downscaled runoff and discharge estimates at high spatial resolution. In: AGU Fall Meeting Abstracts, pp H43M-07
- Mayorga E, Seitzinger SP, Harrison JA, Dumont E, Beusen AHW, Bouwman AF, Fekete BM, Kroeze C, Van Drecht G (2010) Global Nutrient Export from WaterSheds 2 (NEWS 2): Model development and implementation. *Environ Modell Softw* 25:837–853. <https://doi.org/10.1016/j.envsoft.2010.01.007>
- Morelli B, Hawkins TR, Niblick B, Henderson AD, Golden HE, Compton JE, Cooter E, Bare JC (2018) Critical review of eutrophication models for life cycle assessment. *Envir Sci Tech* 52:9562-9578. <https://doi.org/10.1021/acs.est.8b00967>
- Norris AG (2003) Impact characterization in the Tool for the Reduction and Assessment of Chemical and other environmental Impacts: Methods for acidification, eutrophication, and ozone formation. *J Ind Ecol* 6:79-101. <https://doi.org/10.1162/108819802766269548>
- Paerl HW (2009) Controlling eutrophication along the freshwater-marine continuum: Dual nutrient (N and P) reductions are essential. *Estuar Coast* 32:593-601. <https://doi.org/10.1007/s12237-009-9158-8>
- Payen S, Ledgard SF (2017) Aquatic eutrophication indicators in LCA: Methodological challenges illustrated using a case study in New Zealand. *J Clean Prod* 168:1463-1472. <https://doi.org/10.1016/j.jclepro.2017.09.064>
- Payen S, Núñez M, Bjørn A, Laurent A, Bare J, Civit B, Golden H, Niblick B, Henderson AD (in preparation) Eutrophication and acidification midpoint characterization factors in the GLAM consensus model: the importance of aggregation scheme. *Int J Life Cycle Assess*
- Rosenbaum RK, Margni M, Jolliet O (2007) A flexible matrix algebra framework for the multimedia multipathway modeling of emission to impacts. *Environ Int* 33:624–634. <https://doi.org/10.1016/j.envint.2007.01.004>
- Roy P-O, Huijbregts M, Deschênes L, Margni M (2012) Spatially-differentiated atmospheric source-receptor relationships for nitrogen oxides, sulfur oxides and ammonia emissions at the global scale for life cycle impact assessment. *Atmos Environ* 62:74-81. <https://doi.org/10.1016/j.atmosenv.2012.07.069>
- Sabo R, Clark C (2019) National Inventory of Phosphorus. U.S. Environmental Protection Agency: Environmental Dataset Gateway. <https://doi.org/10.23719/1504278>

- Schindler DW (2006) Recent advances in the understanding and management of eutrophication. *Limnol Oceanogr* 51:356–363. https://doi.org/10.4319/lo.2006.51.1_part_2.0356
- Sherman K, Alexander L (1986) Variability and management of large marine ecosystems. U.S. Department of Energy. <https://www.osti.gov/biblio/6958679>
- United Nations Environment Programme (UNEP) (2019) Global guidance on environmental life cycle impact assessment indicators. Vol 2. Paris, France
- US EPA (2015) Physical features of the Great Lakes. <https://www.epa.gov/41ustain41ti/physical-features-great-lakes>. Accessed 20 August 2020

3 TASK 2.2 – ADDRESS THE DISCONNECT BETWEEN DATA DEVELOPED IN IMPACT ASSESSMENT METHODS AND DATA FOR LIFE CYCLE INVENTORIES

This chapter contains the content of the following published journal article that was written as part of this project:

Young B, Srocka M, Ingwersen W, et al (2021) LCIA Formatter. *Journal of Open Source Software* 6:3392. <https://doi.org/10.21105.joss.03392>

3.1 SUMMARY

Life Cycle Assessment (LCA) is an established and standardized methodology to comprehensively assess environmental and public health metrics across industries and products (International Organization for Standardization 2006). The EPA is developing an open source LCA tool ecosystem (Ingwersen 2019). The ecosystem includes tools to automate the creation of life cycle inventory (LCI) datasets, which account for flows to and from nature for steps across the life cycle of products or services, and tools for life cycle impact assessment (LCIA) to support classification and characterization of the cumulative LCI to potential impacts. Impacts are expressed via indicators, either midpoint or endpoint, corresponding to different points on the environmental cause-effect chain model (Frischknecht and Jolliet 2016). This paper describes an EPA LCA ecosystem tool ‘LCIA formatter’ that extracts LCIA information from original source methods and converts the data for interoperability with the rest of the EPA LCA ecosystem tools.

3.2 STATEMENT OF NEED

A simplified algorithm for LCA is given in Equation (5), where I are impacts, E are emissions generated (e.g., pollutants) or raw resources consumed (e.g., land, water) per functional unit of product across the entire supporting product system and CF are corresponding characterization factors that quantitatively relate a unit of a flow to a given impact indicator (e.g., global warming, acidification, total land use, etc.).

$$I = \sum(E * CF) \quad \text{Eq. 5}$$

Both E and CF use objects called elementary flows, which are data objects that generally represent a substance (e.g., Ammonia), source or receiving environmental context (e.g., Freshwater Lake), and unit (e.g., kilogram). E will have a total quantity of a given elementary flow from the product system, and the characterization factors in CF are in the form of the indicator unit per elementary flow unit (e.g., kg N-eq per kg Ammonia). E comes from the LCI calculated for the given product under user-defined conditions, and CF is a static dataset that comes from an LCIA method provider.

LCA software generally include LCIA methods to provide impact assessment results for user-created and software-provided LCI, but the elementary flows used in these LCIA methods must be the same data objects as the flows in the LCI data to enable proper impact result calculations. A major challenge to reproducing LCA results and sharing LCA models across platforms is that there is no internationally common list of elementary flows used by either LCI, LCIA, or software

providers. A critical review of elementary flow data used in LCA showed that flows in LCIA methods are the least clearly described and therefore least portable and machine-readable among those from LCI, LCIA and software providers (Edelen et al. 2018). Furthermore, LCA software providers of both proprietary and open source software rarely provide an open process or external peer review of their incorporation of data from LCIA providers (Brightway2 is one exception (Mutel 2017)). This is one potential cause of discrepancies in results across LCA software.

The LCIA formatter is a specific solution to this problem. The LCIA formatter transparently acquires LCIA methods from original provider data portals, maps them to an authoritative flow list, and exports them in common data formats. The LCIA formatter v1.0 uses the FLCAC Elementary Flow List (FEDEFL) as the authoritative system of elementary flows (Edelen et al. 2019). This system has been adopted by federal agencies in the U.S. to share data through the FLCAC ([FLCAC](#)). The automation of this process provided by the LCIA formatter is critical, because as elementary flows used in LCI are updated, the LCIA datasets should also be automatically updated, and vice versa, to facilitate their use in LCA.

3.3 STRUCTURE

The LCIA formatter code is written in Python 3 and created as a standard python package called `lciafmt` that can be installed using pip. The LCIA formatter primarily uses pandas (McKinney 2010) for data wrangling, the Apache parquet format (Apache 2021) for local storage of processed datasets, and `olca-ipc` (Srocka 2020) for writing data in a standard LCA data format. The code is stored on a EPA GitHub [repository](#) and is publicly accessible.

The LCIA formatter accesses LCIA datasets directly from the data provider. These datasets are typically provided as Microsoft Excel or Access files. These are downloaded and saved in a temporary local cache. To support the specific functions necessary to access and parse individual methods, each method is processed within its own Python module. Flow names, indicators, characterization factors, and other metadata are compiled in a [standard format](#). Adjustments are made as needed to improve consistency between indicators and across methods. This includes handling duplicate entries for the same elementary flow and data cleaning (such as cleaning string names, adjusting capitalization, formatting of CAS Registry Numbers). Additionally, the LCIA formatter supports the inclusion of non-specified secondary contexts (emission locations) where none are provided. Where methods provide both midpoint and endpoint categories within a single source, the LCIA formatter parses these indicators for separate use. Midpoint indicators are typically calculated at an intermediate point along the cause-effect model (such as final mass of substance in the environment), and endpoint indicators typically correspond to damage from the elementary flows (such as a disability-adjusted life year, or DALY) (Frischknecht and Jolliet 2016). Therefore, midpoint and endpoint indicators should be evaluated separately. Finally, source flow data are mapped to elementary flows in the FEDEFL (Edelen et al. 2019), through mapping files provided within that [package](#) (Ingwersen et al. 2020). These mapping files correspond to flow names and contexts to a common set of elementary flows generated for LCA modeling by the EPA. Because flows are provided at more detailed contexts than exists in most LCIA methods currently, mapping to the FEDEFL typically expands the number of characterization factors available. Mapped methods are stored locally as parquet files for future access by LCIA formatter or other tools. Additionally, mapped methods can be exported in JSON-LD format for use in LCA software tools such as [openLCA](#).

3.4 AVAILABLE METHODS

The LCIA formatter is structured to easily convert original source data from existing LCIA methods. Version 1.0 of the LCIA formatter converts three commonly used LCIA methods: TRACI 2.1 (Bare 2011), ReCiPe2016 (Huijbregts et al. 2017), and ImpactWorld+ (Bulle et al. 2019) (Table 7), which cover a variety of impact categories. Cumulative LCI indicators based on the FEDEFL are also available as an output of the LCIA formatter.

Table 7. Impact Methods Processed by lciafmt (Young et al. 2021d, 2021c, 2021b, 2021a)

Method	Indicators	Characterization Factors (Source)	Characterization Factors (Mapped to FEDEFL)
TRACI2.1	9 midpoint	33,429	200,686
ReCiPe2016	18 midpoint and 22 endpoint across three perspectives	347,010	2,429,470
ImpactWorld+	18 midpoint and 26 endpoint	87,804	457,983
FEDEFL Inventory Methods	9 indicators	n.a.	5,649

3.4.1 TRACI2.1

The EPA's Tool for Reduction and Assessment of Chemicals and Other Impacts (TRACI) is widely used for LCA across the U.S. Federal Government and by U.S.-based LCA practitioners (Bare 2011). The LCIA formatter accesses the TRACI Excel-based source file and characterizes impacts across 9 midpoint indicators for use in U.S.-focused analyses. When available, flow characterization factors are mapped to the release context with the greatest level of detail. For example, an air emission may be linked to the specific release height and population density of release area. If secondary context data is unknown, the LCIA formatter generates an average factor across the possible contexts to ensure the flow will still be captured in model calculations.

3.4.2 ReCiPe2016

ReCiPe 2016 characterizes impacts across 18 midpoint indicators and three perspectives: Individualist, Hierarchist, and Egalitarian (Huijbregts et al. 2017). The LCIA formatter generates endpoint impacts through a series of midpoint conversion factors provided for each indicator. As is done for TRACI, where characterization factors are not supplied for secondary contexts, an average factor across the possible contexts is generated. This ensures that users that do not specify a secondary context (e.g., emission to air with no indication of population density) can still obtain a characterization factor for a flow.

3.4.3 ImpactWorld+

ImpactWorld+ v1.3 provides characterization factors for 44 indicators at midpoint or endpoint levels (Bulle et al. 2019). ImpactWorld+ v1.3 is downloaded as an Access database and read into

a pandas dataframe using pyodbc. Only elementary flows with global characterization factors are included at this time, as spatially explicit characterization factors (e.g., country) are not yet compatible with the FEDEFL. Context information is added for water scarcity and availability categories. Flowable name is applied as context for land occupation and transformation categories. Context descriptions are provided in the original source for all other categories.

3.4.4 FEDEFL Inventory Methods

The LCIA formatter generates life cycle inventory methods based on groups of elementary flows identified in the FEDEFL. For example, an inventory method for energy resource use represents the summation of all instances of these flows within a dataset. Where necessary unit conversions are applied to achieve a consistent indicator unit.

3.4.5 Valuation

The LCIA formatter also includes a method-agnostic approach to convert indicators (midpoint or endpoint) to monetary values. The primary valuation is based on modified budget constraint modeling (Weidema 2009), updated to USD2014. A DALY is valued as the global average annual potential economic production per capita; the value of ecosystem damages is calculated from the ratio of global population to terrestrial surface area, and it is also validated based on environmental preservation expenditures in selected countries. The conversions between the different ecosystem impact indicators (e.g., PDF.m2.yr and species.yr) are based on the species density calculations from ReCiPe 2008 (Goedkoop et al. 2009).

3.5 APPLICATIONS

The LCIA methods generated by the LCIA formatter for use with the FEDEFL are hosted publicly on the FLCAC for use by LCA practitioners and researchers. These methods support LCAs performed by many parties, including member agencies for the FLCAC such as EPA, U.S. Department of Energy, U.S. Department of Agriculture, U.S. Department of Defense, and others. These methods also enable impact assessment for researchers utilizing the U.S. Life Cycle Inventory (USLCI) Database. LCIA methods from the LCIA formatter are also being used in the [standard format](#) as inputs into other EPA LCA ecosystem tools, including [useeior](#) (Ingwersen et al. 2021) and the Electricity Life Cycle Inventory ([electricitylci](#)) (Jamieson et al. 2020). The system was built to be flexible enough to support creating outputs for LCIA spatially-explicit characterization factors as those become more common. Users and LCA practitioners are encouraged to suggest additional LCIA methods or applications of the lciafmt as described in the package [Wiki](#).

3.6 REFERENCES

- Apache. 2021. “Pyarrow: Python Library for Apache Arrow.” <https://pypi.org/project/pyarrow/>.
- Bare, Jane. 2011. “TRACI 2.0: The Tool for the Reduction and Assessment of Chemical and Other Environmental Impacts 2.0.” *Clean Technologies and Environmental Policy* 13 (5): 687–96. <https://doi.org/10.1007/s10098-010-0338-9>.
- Bulle, Cécile, Manuele Margni, Laure Patouillard, Anne-Marie Boulay, Guillaume Bourgault, Vincent De Bruille, Viêt Cao, Michael Hauschild, Andrew Henderson, and Sebastien

- Humbert. 2019. "IMPACT World+: A Globally Regionalized Life Cycle Impact Assessment Method." *The International Journal of Life Cycle Assessment* 24 (9): 1653–74. <https://doi.org/10.1007/s11367-019-01583-0>.
- Edelen, Ashley, Troy Hottle, Sarah Cashman, and Wesley Ingwersen. 2019. "The Federal LCA Commons Elementary Flow List: Background, Approach, Description and Recommendations for Use." EPA/600/R-19/092. Washington, DC: U.S. Environmental Protection Agency.
- Edelen, Ashley, Wesley W. Ingwersen, Cristina Rodríguez, Rodrigo A. F. Alvarenga, Artur Ribeiro de Almeida, and Gregor Wernet. 2018. "Critical Review of Elementary Flows in LCA Data." *The International Journal of Life Cycle Assessment* 23 (6): 1261–73. <https://doi.org/10.1007/s11367-017-1354-3>.
- Frischknecht, Rolf, and Olivier Jolliet. 2016. *Global Guidance for Life Cycle Impact Assessment Indicators: Volume 1*. UNEP / SETAC Life Cycle Initiative. United Nations Environment Program / Society for Environmental Toxicology; Chemistry Life Cycle Initiative. <http://www.lifecycleinitiative.org/training-resources/global-guidance-lcia-indicators-v-1/>.
- Goedkoop, Mark, Reinout Heijungs, Mark A. J. Huijbregts, An De Schryver, Jaap Struijs, and Rosalie van Zelm. 2009. *ReCiPe 2008: A Life Cycle Impact Assessment Method Which Comprises Harmonised Category Indicators at the Midpoint and the Endpoint Level; Report 1: Characterisation*. 1st edition. Ministry of Housing, Spatial Planning,; Environment (VROM).
- Huijbregts, M. A. J., Z. J. N. Steinmann, P. M. F. Elshout, G. Stam, F. Verones, M. D. M. Vieira, A. Hollander, M. Zijp, and R. van Zelm. 2017. "ReCiPe 2016: A Harmonized Life Cycle Impact Assessment Method at Midpoint and Endpoint Level Report i: Characterization." *International Journal of Life Cycle Assessment* 22 (2). <https://doi.org/10.1007/s11367-016-1246-y>.
- Ingwersen, Wesley. 2019. "LCA Tool Ecosystem." https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=347759.
- Ingwersen, Wesley, Ashley Edelen, Troy Hottle, Ben Young, Sarah Cashman, and Michael Srocka. 2020. "Fedelemflowlist." *GitHub Repository*. GitHub. <https://github.com/USEPA/Federal-LCA-Commons-Elementary-Flow-List>.
- Ingwersen, Wesley, Mo Li, Ben Young, Catherine Birney, and Jorge Vendries. 2021. "Useeior." *GitHub Repository*. GitHub. <https://github.com/USEPA/useeior>.
- International Organization for Standardization. 2006. "Environmental Management – Life Cycle Assessment – Principles and Framework." ISO Standard No. 14040:2006.
- Jamieson, Matt, Wesley Ingwersen, Greg Schively, Tapajyoti Ghosh, Ben Young, and Troy Hottle. 2020. "Electricitylci." *GitHub Repository*. GitHub. <https://github.com/USEPA/electricitylci>.
- McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>.

- Mutel, Chris. 2017. “Brightway: An Open Source Framework for Life Cycle Assessment.” *Journal of Open Source Software* 2 (12): 236. <https://doi.org/10.21105/joss.00236>.
- Srocka, Michael. 2020. “Olca-Ipc.” *GitHub Repository*. GitHub. <https://github.com/GreenDelta/olca-ipc.py>.
- Weidema, Bo Pedersen. 2009. “Using the Budget Constraint to Monetise Impact Assessment Results.” *Ecological Economics*, Eco-efficiency: From technical 47ustain47tion to reflective sustainability analysis, 68 (6): 1591–98. <https://doi.org/10.1016/j.ecolecon.2008.01.019>.
- Young, Ben, Michael Srocka, Wesley Ingwersen, Ben Morelli, Sarah Cashman, and Andrew Henderson. 2021a. “FEDEFL Inventory Methods V1.0.0.” <https://doi.org/10.23719/1522414>.
- . 2021b. “ImpactWorld+v1.3 for FEDEFLv1.” <https://doi.org/10.23719/1522444>.
- . 2021c. “ReCiPe2016v1.1 for FEDEFLv1.” <https://doi.org/10.23719/1522445>.
- . 2021d. “TRACIv2.1 for FEDEFLv1.” <https://doi.org/10.23719/1522413>.

4 TASK 3.1 – STRATEGIC ROADMAP: USEEIO AND DIO ALIGNMENT

4.1 INTRODUCTION

The DoD Sustainability Analysis Guidance (SAG) provides a consistent, practical, and flexible method for conducting life cycle assessments (LCAs) throughout the DoD acquisition process. The method is described in a detailed guidance document (OASD 2016). A sustainability analysis is an overlay on performance; in other words, it identifies the most sustainable alternative among those that meet performance requirements. Results are intended to make relative comparisons. The Sustainability Analysis employs an integrated-hybrid approach that incorporates available primary data from those conducting the analysis with background data from environmentally-extended input-output (EEIO) models. These EEIO models incorporate economic and environmental data from nearly all sectors across the U.S. economy that provide a proxy to assess DoD supply chain activities.

The objective of this task is to develop a Strategic Roadmap that outlines the opportunities to improve EEIO models to provide more accurate and comprehensive life cycle results for products, including defense-unique products such as weapons systems and platforms. The Roadmap was developed in collaboration with the U.S. Environmental Protection Agency (EPA) and includes detailed descriptions of barriers and issues, solutions and resolutions, and expected timelines for the steps of a path forward. The Roadmap also discusses the feasibility of aligning DoD and EPA EEIO models.

4.1.1 Input-Output (IO) Modeling

Economic input-output (IO) models are a publicly available and comprehensive description of exchanges of goods and services between sectors of the economy. In the United States, the Bureau of Economic Analysis (BEA) provides these data (US BEA, 2017). Input-output models can be extended to include environmental information. An EEIO model couples sector-based economic data—referred to as the technosphere, or A matrix—with two additional pieces of information (Figure 19). One addition is environmental data that links sectors to environmental releases and resource use. These data, called the B matrix, are frequently gathered from national statistics. In the United States, the Toxics Release Inventory² (TRI) and National Emissions Inventory³ (NEI) provide most of these data. Finally, the characterization, or C matrix, relates emissions of individual substances or use of natural resources to impacts that are modeled in the system. These characterization factors often come from publicly-available LCA impact models, such as TRACI or Impact World 2002+ (Bare et al. 2003; Jolliet et al. 2003).

² <https://www.epa.gov/toxics-release-inventory-tri-program>

³ <https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei>

The vision of sustainability for DoD is to maintain the ability to operate into the future without decline—either in the mission or in the natural and man-made systems that support it (DoD 2014). At the heart of sustainability is the need to preserve Earth’s natural resource base for future use. DoD depends on these resources—energy, air, land, and water—to conduct its mission. The objective is for sustainability to become thoroughly woven into the everyday fabric of the DoD mission. Just as sustainability directly affects the ability of DoD to perform its mission in the future, so does the Department’s ability to be resilient. The DoD supports sustainable solutions that improve warfighter lethality and make systems available for engagement.⁴

Chapter 3 of the DoD Defense Acquisition Guide (DAG) provides overarching guidance on the systems engineering discipline, its activities and processes, and its practice in defense acquisition programs. The Program Manager (PM) and the Systems Engineer should use this chapter to effectively plan and execute program activities across the system life cycle. Systems engineering (SE) establishes the technical framework for delivering materiel capabilities to the warfighter. SE planning, as documented in the Systems Engineering Plan (SEP), identifies the most effective and efficient path to deliver a capability, from identifying user needs and concepts through delivery and sustainment. Systems engineering event-driven technical reviews and audits assess program maturity and determine the status of the technical risks associated with cost, schedule, and performance goals. The SE approach is implementable under the Common and Accelerated acquisitions models. A Sustainability Analysis, using an LCA method, is a tool to assist in designing systems that are more sustainable—those that use fewer resources over the life cycle, and have fewer human health and environmental impacts, and thus have a lower total ownership cost.

Chapter 4 of the DAG provides guidance for PMs and Program Support Managers (PSMs) to develop and execute successful sustainment strategies, and to document those strategies in a Life Cycle Sustainment Plan (LCSP) that aids program management, communication, and collaboration with critical stakeholders. The purpose of life cycle sustainment planning is to maximize readiness by delivering the best possible product support outcomes at the lowest Operating and Support (O&S) cost. While weapon system sustainment does not actually begin until the first production units are fielded, sustainment planning begins at the earliest stages of the defense acquisition system. Successful post-fielding sustainment performance depends on critical thinking during requirements development and solution analysis.

Environment, Safety and Occupational Health analyses are an integral, ongoing part of the SE process throughout the life cycle. The benefits of early integration of ESOH considerations include:

⁴ Sustainability is not synonymous with the term “sustainment,” which the Defense Acquisition University (DAU) Glossary of Defense Acquisition Acronyms and Terms defines as “translating force provider capability and performance requirements into tailored product support to achieve specified and evolving life cycle product support availability, reliability, and affordability parameters. Life cycle sustainment considerations include supply; maintenance; transportation; sustainment engineering; data management; configuration management; human systems integration (HSI); environment, safety (including explosives), and occupational health; protection of critical program information and anti-tamper provisions, supportability, and interoperability. Initially begun during Materiel Solution Analysis (MSA) phase and matured during the Technology Maturation and Risk Reduction (TMRR) phase, life cycle sustainment planning spans a system’s entire life cycle from MSA phase to disposal.”

- Mitigation of program cost and schedule risks from actions that cause damage to people, equipment, or the environment.
- Reduction of O&S and disposal costs to achieve system affordability.
- Provision of a safe, suitable, supportable, and sustainable capability able to operate world-wide, including opportunities for Foreign Military Sales.

DoD Instruction (DoDI) 5000.02, Enc 3, Section 16 requires programs to use the system safety methodology in Military Standard (MIL-STD) 882 to manage their ESOH considerations as an integral part of the program's overall SE process. This starts with including ESOH management planning in the Milestone A SEP to cover Technology Maturation and Risk Reduction (TMRR) activities and continues throughout the system's life cycle.

The sustainability analysis, using LCA, is a tool to assist the systems engineer in designing more sustainable systems (DoD, 2022).⁵ The streamlined nature of EEIO models allows DoD the ability to focus resources on defense-unique data collection. Collaboration with other agencies and partners on the non-defense-unique data collection and processes allows for efficient use of resource across the federal landscape and instills effective resource management and meaningful innovation. This Roadmap includes a strategy for advancing DoD's ability to assess life cycle-based impacts and costs associated with the acquisition of weapons systems and platforms via use of EEIO models and outlines the path to roll out the strategy.

The Roadmap identifies opportunities to align the DoD and EPA models including:

- Sharing common data—both economic and environmental—when feasible,
- Expanding the scope and accuracy of the models, including additional indicators,
- Potentially capturing more international data,
- Improving spatial resolution, and
- Maintaining the models.

4.2 CURRENT STATE

4.2.1 Defense Input-Output (DIO) Model

The initial DIO model (v1) model is an EEIO of the U.S. model that has been integrated with process-based LCA data extension for various transportation and energy generation processes selected due to DoD relevance. It includes impact characterization factors that can be used to estimate midpoint and endpoint impacts as well as monetary costs of impacts. A schematic of the DIO v1.0 model is presented in Figure 20.

⁵ DoD Guidance (2022) continues that a streamlined LCA process should be used and that sustainability analysis is intended to ensure consideration of important downstream impacts and costs in trade-off and design decisions.

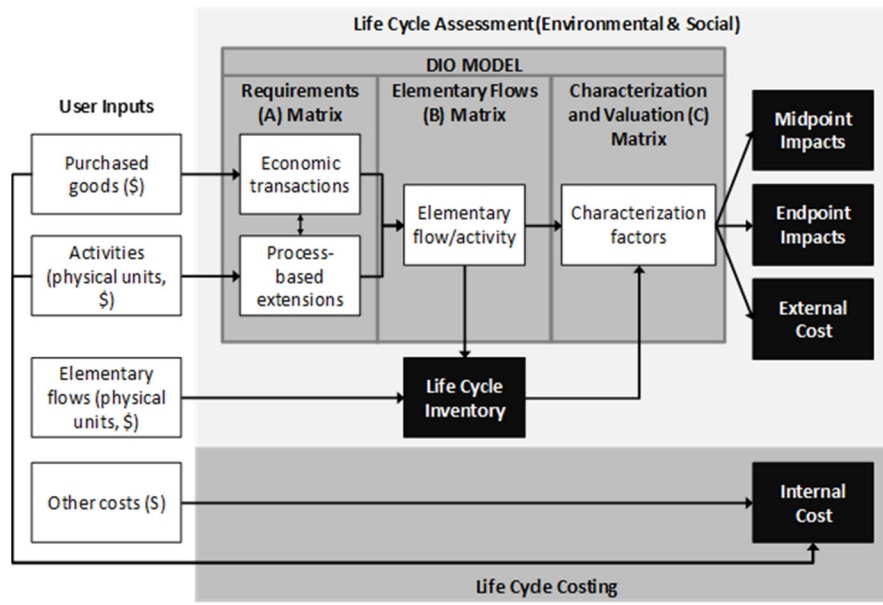


Figure 20. Data Components of DIO

4.2.2 U.S. Environmentally-Extended Input-Output Model (USEEIO)

The EPA developed the USEEIO model in support of its Sustainable Material Management program (Yang et al. 2017). The USEEIO uses the same underlying data from BEA as does the DIO, although the choice of input-output (IO) table is different. The USEEIO is built on producer prices and uses a commodity-by-commodity matrix representing 389 sectors. For the environmental extension of inventory data (including land use, water use, energy, mineral use, emissions of criteria air pollutants, and nutrients), many of the underlying data are the same, although the USEEIO has some additions that the DIO does not, such as nutrient emissions and employment. USEEIO data is largely based on sources from 2013 and includes explicit data quality and extensive metadata. The USEEIO is built using Python and R scripts, which are publicly available; these scripts can start directly with data from the BEA website and are being expanded to allow for more user flexibility. A high-level flow of information and arrangement of these Python scripts is shown in Figure 21.

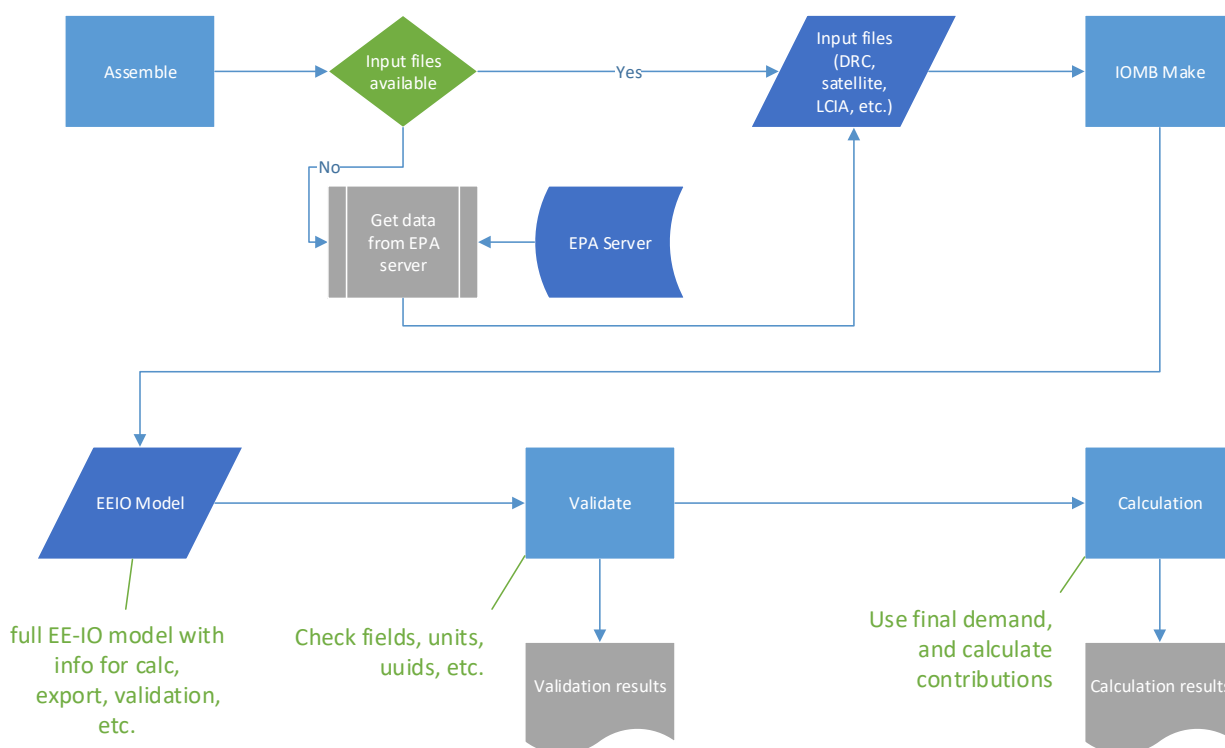


Figure 21. Simplified Structure of USEEIO Model Assembly, Validation, and Calculation

The ability for the scripts to import data from BEA is a significant improvement in terms of future maintenance and updating of the USEEIO, as significant work is not required when new data are released.

The EPA is participating in international efforts to standardize and harmonize inventory flow naming conventions. Ultimately, this will facilitate incorporation of data from variety sources into a consistent dataset. As the USEEIO is inherently connected to these data harmonization efforts, the USEEIO framework and flow naming conventions are becoming a de facto standard.

4.2.3 Comparison Table

A table comparing the design decisions and data sources behind the two IO databases is included in Appendix B.

Briefly, benefits that the DIO brings include the following:

- DIO has higher resolution in elementary flow compartments. For example, DIO identifies emissions to air and soil with urban and rural compartments
- DIO has higher resolution in electricity consumption (at state level) than the default USEEIO model.
- DIO has data for occupational health.
- DIO uses IMPACT WORLD CF's as well as the DOD valuation methods. In contrast, USEEIO uses the TRACI 2.1 method along with some inventory analysis methods that

aggregate resources and emission flows by broader type (e.g., ‘freshwater extracted from the ground’ as ‘water’)

- DIO is hybridized: i.e., it blends the pure input-output approach of USEEIO with some data relevant to specific processes (e.g., military transport), rather than sectors.

4.2.4 Barriers and Issues

The alignment of the two models will require that the following differences in the models be addressed:

- IO model base data: The choice of the BEA Use table (either ‘before’ or ‘after’ sector redefinitions) and the commodity-by-commodity versus commodity-by-industry, represents a conceptual difference between the USEEIO and the DIO. The two models need not necessarily use the same base data, but the needs of users and stakeholders of the two models should be considered when deciding which structure to use. Absent a clear need for different data sources, these two IO models should be the same.
- IO model sector aggregation: Currently, the DIO and USEEIO result in slight differences stemming from sector aggregation.
- Matrix size: Currently, the USEEIO A matrix has 389 sectors; the DIO A matrix has 391. The USEEIO framework does not easily allow for the addition of process flow extensions, such as the DIO data for energy (342 entries for electricity and military energy generation) and military transport (54 entries).
- Harmonization of inventory flow names: Different naming conventions in the two databases could lead to accidental omission of flows or double counting if the databases are combined. For example, flows named “compound X, rural, low stack emission” and “compound X, rural” and “compound X, low stack emission, rural” are treated as different. Therefore, inventory flows in the B matrix must be a harmonized.
- Expansion of inventory flows: Ideally, substances captured in one model should also be captured in the other. If a compound is of interest to the DIO but is not reported in the USEEIO, then it is necessary to make sure that compound appears in matrices from the USEEIO that are used to build DIO.

4.3 STRATEGY

4.3.1 Goals

Broad goals for the alignment of the two models include the following:

- Build an EEIO model in a manner that streamlines IO modeling for DoD without duplication of efforts across the Federal Government.
- Identify opportunities to coordinate efforts between agencies with respect to maintenance of IO and LCA modeling at DoD.
- Identify DoD-specific IO and LCA modeling areas (e.g., worker health, noise, waste flows, material criticality) to be carried forward in future modeling.

4.4 ROADMAP

4.4.1 Short-term

Based on discussions with EPA, partial alignment of the IO portions USEEIO and DIO is impractical. The challenges described above require robust changes to the USEEIO framework to be more flexible with respect to matrix size. In order to align some of the modeling choices modeling teams at the respective Agencies need to come to consensus on the assumptions and related input datasets or develop the flexibility to generate different model versions within the same modelling framework built on acknowledged differences in these assumptions.

The first step towards alignment was determined to be a proof-of-concept: rebuilding the DIO in the USEEIO framework, using current DIO data, even though it is not aligned with USEEIO data. Specifically, the technology, environmental emission, and impact matrices (A, B, C) were used as input data to the USEEIO modeling framework.

This partial alignment required the following steps, which were all completed in Fall 2018:

- Install and run USEEIO framework
- Extract DIO data in matrix form
- Manipulate DIO data to match USEEIO format (e.g., with respect to column headings and delimiters)
- Run USEEIO framework with DIO data

4.4.2 Medium-term

As preparation for a full alignment, the following medium-term steps are envisioned.

- Add flow subcompartments to USEEIO B and C matrices, building on existing value of subcompartment as demonstrated in DIO:
 - Determine which additional flow subcompartments are in DIO and which are associated with unique characterization factors.
 - From the impact methods, get an accurate definition of how those subcompartments are defined.
 - Determine how to link those flows and factors to satellite tables and which data sources in USEEIO.
 - Determine the best method of categorizing emissions in these sources by one of these subcompartments. Develop a method to automate that categorization.
- Recreate the DIO occupational health satellite table in the USEEIO framework
- Standardize inventory:
 - Implement the new FLCAC flow list in the USEEIO framework.
 - Develop a mapping to the IMPACT WORLD and DOD valuation methods LCIA methods.
 - Modify the framework to allow flexibility in selection of impact methods along with model creation.
- Modify the USEEIO framework to allow for an augmentation of the creation of the direct requirements matrix to hybridize by adding in process LCI.

- Identify any other satellite table flows unique to DIO and important to impact results that would otherwise be lost in adoption of an updated USEEIO table. Determine a means to incorporate these flows into an updated USEEIO satellite table.
- Create a noise satellite table and associated impacts methods in a way that can be incorporated into the USEEIO modeling framework
- Determine how to address location-specific information currently incorporated into DIO. Two approaches that have been discussed:
 - Create a multi-regional input-output model (MRIO). The EPA is working currently on a two-region model for a U.S. state and the rest of the United States. Many other forms are possible. MRIO models are much more complex and of larger size than single region models, so very significant changes would be required (for instance in a 50-region model, the A matrix is changed to replicate the flows for each region, so the matrix becomes quite large; what was a single United States activity can become 50).
 - If the A matrix is not changed, then spatial information must be captured elsewhere, such as in the flow name. In general, with the FLCAC flow list development, there have been arguments against using location identifiers in elementary flows.

4.4.3 Long-term

The long-term steps required for alignment are outlined below; the timeline for this alignment is still being discussed.

- Adopt the USEEIO modeling framework into the build stream for EEIO models/model components to meet DOD needs. Contribute to improving the common modeling framework, while separating any proprietary DOD EEIO or other LCA data to be used in an EEIO model from the public framework. To meet this need, build capacity among DOD contractors to build models using USEEIO framework that meet DOD needs.
- Define a process for consensus development on fundamental IO assumptions and any conceptual issues surrounding shared satellite tables. Permit conceptual deviations for EEIO-based applications through adding flexibility to the USEEIO modeling framework.
- Evaluate the compatibility of EEIO models developed with USEEIO modeling framework to work in AcquLCA software or any other DOD software in use or planned that uses an underlying EEIO model. Reconcile any differences through extension of USEEIO modeling framework or revisions to AcquLCA.
- Align DOD plans for new EEIO applications or AcquLCA software updates with EPA's USEEIO model and modeling framework development schedule, to assure these applications and updates are compatible with the most recent EEIO model advances/attributes.
- Identify on an ongoing basis any DOD-specific needs for higher sectoral resolution in an EEIO model (e.g., separating site remediation from waste management activities). Coordinate with EPA to address these needs, in determining whether economic disaggregation or hybridization is the preferred approach, and in performing these updates in a way that can be incorporated into current and future model versions.

4.5 REFERENCES

- Bare, J.C., Norris, G.A., Pennington, D.W. & McKone, T.E. 2003. "TRACI: The tool for the reduction and assessment of chemical and other environmental impacts." *Journal of Industrial Ecology* 6 (3–4): 49–78.
- Bruckner, M., Lloyd, S. & Rak, A. 2016. Using the Defense Input-Output (DIO) Dataset to streamline LCA for the U.S. Department of Defense presented at the LCA XVI Conference, September, Charleston, SC, USA.
- DoD. 2014. *Department of Defense Strategic Sustainability Performance Plan: FY 2014*. <https://www.denix.osd.mil/sustainability/dod-sspp/>.
- DoD. 2022. *Systems Engineering Guidebook*. Office of the Under Secretary of Defense for Research and Engineering.
- Joliet, O., Margni, M., Charles, R., Humbert, S., Payet, J., Rebitzer, G. & Rosenbaum, R.K. 2003. "IMPACT 2002+: A new life cycle impact assessment methodology." *International Journal of Life Cycle Assessment* 8 (6): 324–330.
- OASD. 2016. *Sustainability Analysis Guidance: Integrating Sustainability into Acquisition Using Life Cycle Assessment. Version 5.0*. <http://www.denix.osd.mil/esohacq/home/>.
- US BEA. 2017. *Concepts and Methods of the U.S. National Input-Output Accounts*. <https://www.bea.gov/methodologies/>.
- Ward, H., Wenz, L., Steckel, J. C., & Minx, J. C. 2018. "Truncation error estimates in process life cycle assessment using input-output analysis." *Journal of Industrial Ecology*, 22(5), 1080-1091.
- Yang, Y., Ingwersen, W.W., Hawkins, T.R., Srocka, M. & Meyer, D.E. 2017. "USEEIO: A new and transparent United States environmentally-extended input-output model." *Journal of Cleaner Production* 158: 308–318.

5 TASK 3.2 USEEIO-DIO MODEL

5.1 OBJECTIVE

EPA and DoD have each constructed environmentally-extended input-output (EEIO) models for the U.S. economy that can be used as a standard and complete background life cycle model. A model built on the best aspects of each existing model with new enhancements can provide a robust background LCA model to use for high-level analysis or in the background of detailed product or technology LCA studies as part of Defense Acquisitions. Task 3.2 focuses on the fourth step of the task to improve DoD and EPA input-output models: develop DIO v2.0 that combines elements of USEEIO and DIO and to document the new model.

5.2 BACKGROUND

Within DoD, acquisition refers to the design, engineering, testing, production, operations, maintenance, and disposal of defense systems. Numerous DoD initiatives integrate life cycle thinking and aspects of sustainability into the acquisition process. Further, DoD policy guidance describes sustainable acquisition as “acquisition that minimizes negative impacts on resource availability, climate change, human health, and ecosystem quality” (DoD, 2020a).

The concept of life cycle costing (LCC) was introduced by DoD during the early 1960s and has been required for major acquisition programs since 1971 as the DoD notes (Dhillon, 1989; DoD, 1971; LMI, 1965). Once a DoD requirement (or capability) is identified, resources are allocated, and an acquisition program is established to develop or purchase the needed capability or defense system. This process is managed in accordance with DoD Instruction (DoDI) 5000.02, Operation of the Adaptive Acquisition Framework (AAF), and requires systems engineering techniques at several decision points throughout the AAF (DoD, 2020b). These systems engineering techniques include LCA and LCC estimating (DoD, 2022). DoD policy also requires integration of environment, safety, and occupational health (ESOH) risk management into the acquisition process (DoD, 2020b) and use of a Programmatic Environmental, Safety and Occupational Health Evaluation (PESHE) to document ESOH-related considerations and data (DoD, 2022). Acquisition programs also establish energy key performance parameter (KPP) attributes and thresholds to manage the fuel and electricity demand of military systems, with a focus on reducing associated risks to the DoD’s mission, e.g., logistics or security risks (CJCS, 2015). These and other sustainability-related attributes are analyzed and used to inform acquisition decisions and inform system trade-off analyses.

For DoD, LCA is a tool to assist in designing more sustainable systems—those that use fewer resources over the life cycle, have fewer impacts on human health and the environment, and thus have a lower TOC (DoD, 2022). These costs include often-overlooked “external costs,” which consist of monetized life cycle impacts quantified via LCA and are used to identify tradeoffs with upfront investment in mitigation measures and associated future impacts (DoD, 2020a). As such, LCA can support decision making in acquisition activities, including analysis of alternatives, trade-space analysis, business case analysis, preliminary design, and supportability analysis (DoD, 2022).

Process-based LCA is a bottom-up approach that assesses inputs and outputs at the process-level. The advantages and disadvantages of process-based LCA are well documented (Lenzen 2000). From DoD's perspective, the primary advantage is that process-based LCA provides detailed, actionable results. The primary disadvantages are that process-based LCA is time intensive, requires significant data, requires practitioners with significant LCA expertise, and results in truncation error (e.g., including process data for a few levels of suppliers). DoD does not possess a central LCI repository or significant LCA expertise in its acquisition community.

EEIO analysis is a top-down approach that relates environmental impacts and resource use to monetary transactions within an economy at an industry sector level (Tukker, 2006). Truncation error is reduced since upstream activities throughout an economy are included (Suh et al, 2004). Once an EEIO model is developed, an analysis can be performed in minutes. However, EEIO models are based on sector averages and only include environmental effects for which industry-level data is available. EEIO models have been used for subnational scale analyses in the United States to inform product standard development (Brown et al., 2009) and for purchasing analysis for U.S. Government procurement (SRA International and IERS, 2014). These precedents suggest that sustainable material management-related needs for providing quantitative detail on subnational and organizational scales can be at least partially addressed with an EEIO model (Yang et al. 2017).

Hybrid LCA combines process- and EEIO-based methodologies to leverage the advantages of both approaches. There are three general types of hybrid LCA methods (Suh et al. 2004; Suh and Huppes 2005). Integrated hybrid LCA computationally integrates the physical flows between processes, monetary transactions between sectors, and the links between the two (Crawford et al. 2018). An integrated hybrid LCA modelling approach was chosen because the computational model can be constructed and operationalized for use by acquisition professionals with varying levels of LCA expertise, developed based on currently available data, updated as LCI is generated, and centrally maintained to ensure a consistent approach to LCA within DoD. This allows for a tiered modeling approach, ranging from screening-level EEIO analysis to comprehensive process-based LCA, in which analysts and decision-makers select the appropriate level of analysis given the type of decision, time and budget constraints, and data availability. For example, in early acquisition stages, the integrated hybrid LCA model can be used as a rapid screening tool to identify important risks and prioritize design strategies. In later acquisition stages, it can support detailed LCAs of specific design alternatives.

Integrated hybrid LCA frameworks have been developed and presented for case studies, e.g., building energy systems, computer manufacturing, wind electricity, roads, and wood biofuel (Suh and Lippiatt, 2012; Williams, 2004; Wiedmann et al., 2011; Treloar et al., 2004; Bright et al., 2010). It is not clear if the modelling components were computationally integrated in these studies. To our knowledge, none of these models have been operationalized for general use.

The DIO v2.0 model presented here combines assets of the DIO v1.0 model and the USEEIO v2.0 model (Ingwersen et al., 2022). The DIO v2.0 model may be the first integrated hybrid LCA model to be fully computationally integrated and operationalized for general use. Novel technical aspects have been enabled by the DIO framework, model, and use of USEEIO framework/process. Model preparation leverages the USEEIO framework so that the EEIO model can be updated regularly with the latest available data from EPA. DIO v2.0 including a novel air emissions dataset that

characterizes the physical and demographic context of air emissions and couples these with context-specific impact characterization factors. The model integrates a unique set of LCIA methods, including midpoint and endpoint methods, and combines these potential impacts with monetary valuation of damages.

5.3 METHODS

A complete life cycle model based on the EEIO method can be represented in three core components: (1) a direct requirement matrix, A , containing the inputs of commodity to per unit commodity output; (2) a B matrix, containing the elementary flows per unit of commodity output; and (3) a C matrix, containing CFs relating elementary flows to an indicator unit.

The model can be used with a demand vector, y , of commodity totals to calculate the impacts using Equation 6:

$$H = CB(I - A)^{-1}\hat{y} \quad \text{Eq. 6}$$

where I is an identify matrix and H is the cradle-to-gate impacts of the commodities demanded in y measured in the indicators in C . The USEEIO v2.0 data descriptor (Ingwersen et al., 2022) provides information for additional derivation of the A matrix as well as details on other model components that can be obtained from these core components.

The team created an integrated hybrid model that builds off the base USEEIO through modification and extension to the USEEIO v2.0 model A and B matrices. The team also created a new C matrix with new indicators.

Matrices used to create the A matrix for DIO integrated-hybrid models include:

1. Process-based matrix, A_p , representing inputs and outputs of products in the process-based model.
2. Upstream cut-off matrix, A_u , representing inputs to the process-based matrix from IO sectors. This matrix is created by mapping processes to IO sectors, identifying processes that consume IO commodities from the mapped sectors, and replacing those physical process inputs to monetary values using average commodity prices.
3. The USEEIO A matrix (A_e).

These matrices are combined into an integrated hybrid technology matrix, A_{IH} , using Equation 7.

$$A_{IH} = \begin{bmatrix} A_p & 0 \\ A_u & A_e \end{bmatrix} \quad \text{Eq. 7}$$

The environmental matrix, B , from the IO model is extended to include the environmental exchanges, B_p associated with the commodities in the process-based model, in the form of Equation 8.

$$B_{IH} = \begin{bmatrix} B_p \\ B_e \end{bmatrix} \quad \text{Eq. 8}$$

Table 8. Inputs Into DIO Model A and B Matrices

Data Input	Previously Described in	Used in Model Component
Energy and Transportation Extensions	--	A_p, A_u, B_p
Air Emissions Totals By Industry	--	B_e
USEEIO v2.0.1-411 A matrix	Ingwersen et al. 2022	A_e
National Water Withdrawal Totals By Industry 2015 v1.1 ⁶⁴	Ingwersen et al. 2022	B_e
National Point Source Releases to Ground By Industry 2017 v1.1 ⁵¹	Ingwersen et al. 2022	B_e
National Point Source Releases to Water By Industry 2017 v1.1 ⁵⁰	Ingwersen et al. 2022	B_e
National Commercial Hazardous Waste Totals by Industry 2017 v1.1 ⁶⁹	Ingwersen et al. 2022	B_e
National Land Occupation Totals By Industry 2012 v1.1 ⁵⁸	Ingwersen et al. 2022	B_e
National Employment Totals By Industry 2017 v1.1 ⁷⁰	Ingwersen et al. 2022	B_e
National Greenhouse Gas Industry Attribution Model 2016	Yang et al. 2020	B_e
Commercial Nonhazardous Waste National Totals By Industry, 2014	Meyer et al. 2020	B_e
Commercial Nonhazardous Waste from Construction National Totals By Industry, 2014	Meyer et al. 2020	B_e
Pesticide Release Totals By Sector from USEEIO v1	Yang et al. 2017	B_e
Nutrient Release Totals By Sector from USEEIO v1	Yang et al. 2017	B_e
Mineral Extraction Totals By Sector from USEEIO v1	Yang et al. 2017	B_e
Energy Extraction Totals By Sector from USEEIO v1	Yang et al. 2017	B_e
Value Added By Sector, 2012	Ingwersen et al. 2022	B_e

The totals by sector or totals by industry files go directly into an elementary flow total by industry E matrix, which is normalized to create B_e . The derivation of B_e from E matrix is described in Ingwersen et al. (2022). All data are adjusted to be in producer price in 2012 U.S. dollars (USD).

5.3.1 Process Extensions

All flows described in this section were mapped to (harmonized with) the FEDEFL (Edelen et al. 2022).

5.3.1.1 Energy Use Extensions

A set of custom, pre-defined energy flows were created and included in the A_p matrix to capture the life cycle impacts from using energy commodities (e.g., natural gas, fuel oil, coal, gasoline, diesel, kerosene, other fossil fuels, biomass, and electricity). This includes the use of fuels in external combustion sources, stationary internal combustion engines, military vehicles, and other mobile and stationary sources. It also includes electricity production based on the average generation mix for the United States, each of the 50 states, the District of Columbia, as well as the 8 regions and 136 countries for which the International Energy Agency (IEA) provides statistics. In all, 342 energy use extensions are included in the A matrix. This makes 342 rows (i) representing energy technosphere flows (energy commodity inputs) and 342 columns (j) representing activities (energy use). Each element a_{ij} represents the amount of energy input (i) required for one unit of activity (j), as measured by that activity's reference unit.

Table 9 lists the data sources used for the energy use extensions. For stationary and mobile combustion, heating values were used to estimate the amount of fuel required per unit energy output, and the 2007 retail prices were used to estimate the amount of expenditures from the relevant industries sectors for producing the fuel source. For electricity use, the average generation mix for the United States, each of the 50 states, and the District of Columbia was obtained from the EPA Emissions & Generation Resource Integrated Database (eGRID) (EPA, 2022). The average generation mix for the 8 regions and 136 countries for which the IEA provides statistics was obtained from the IEA Electricity and Heat for 2007.

Table 9. Data Sources for the Energy Use and Transportation Extensions to the A Matrix

Energy Type	Description	Data Sources
Stationary Combustion	Combustion of fuel in fixed devices generally to produce power, generate steam, or provide heat.	US-EPA AP-42; EMEP/EEA 2009 Air Pollutant Emission Inventory; EMEP/EEA 2013 Air Pollutant Emission Inventory Guidebook
Mobile Combustion	Combustion of fuel in equipment not used for the primary purpose of transportation (e.g., military ships, aircraft, and vehicles)	US-EPA AP-42; EMEP/EEA 2009 Air Pollutant Emission Inventory; EMEP/EEA 2013 Air Pollutant Emission Inventory Guidebook; GREET 2013; EPA MOVES 2010 (a & b)
Electricity (U.S.)	Using electricity based on the average generation mix for the United States, the 50 states, and the District of Columbia	US-EPA eGRID (9th edition); EIA Form 1605
Electricity (Non-U.S.)	Using electricity based generation mix for 8 regions and 136 countries for which the IEA provides statistics	IEA Country Electricity and Heat Statistics; EIA Form 1605
Transportation	Transporting freight, fuel, or people.	GREET 2013; EPA MOVES 2010 (a & b)

5.3.1.2 Transportation Extensions

A set of custom, pre-defined energy flows were created and included in the A matrix to capture the life cycle impacts from transporting commodities in the DoD supply chain, as well as provide estimates for downstream operations. In all, 54 transportation extensions are included in the A

matrix. These include the transportation of freight and fuel via truck, train, barge, ocean freighter, aircraft, and the transport of people via car, truck, motorcycle, bus, and motor home. All vehicles included are fueled by petroleum. This adds 54 rows (i) representing transportation technosphere flows (transportation inputs) and 54 columns (j) representing activities (transportation use). Each element $a_{i,j}$ represents the amount of transportation input (i) required for one unit of activity (j), as measured by that activity's reference unit.

Table 9 lists the data sources used for the transportation extensions. Argonne National Laboratory's Greenhouse gases, Regulated Emissions, and Energy use in Transportation (GREET) Model was used to estimate petroleum use per tonne-kilometre (t-km), person-km, or megajoule (MJ). Retail prices from 2007 were used to estimate purchases from the petroleum refining and pipeline transportation industries.

5.3.2 Indicators and Scoring Factors

LCIA, which translates environmental burdens into impact, is not always included in EEIO models. The DIO model integrates LCIA into the dataset. This streamlines the process for completing an LCA, ensures accurate linkages between LCI data and LCIA methods, enables integration of impact categories developed for DoD, and ensures that LCAs for DoD and its original equipment manufacturers are conducted consistently (i.e., using the same impact categories and impact assessment methods).

The DIO model relies on CFs from published LCIA methods. While occupational health and safety (OH&S) is included in the current DIO model, noise pollution has not yet been integrated. The C matrix includes the CFs used to quantify impacts at the midpoint and endpoint level, as well as the valuation of endpoint impacts. The C matrix includes 19 midpoint categories, 22 endpoint categories, and 32 of valuation categories (discussed below).

Element $c_{i,j}$ represents the impact (i), recorded in the appropriate impact units, resulting from one unit of environmental flow (j), recorded in appropriate elementary flow units. For the valuation categories, $c_{i,j}$ represents the monetary value (i), recorded in USD, resulting from one unit of environmental flow (j).

5.3.2.1 Impact Characterization

Table 10 and Table 11 list the midpoint and endpoint impact categories, respectively, and the units and LCIA methods used for each category. The DIO model leverages EPA's LCIA Formatter tool to generate harmonized LCIA methods (Young et al., 2021). IMPACT World+ was selected as the primary LCIA method, as it represents a joint major update to three LCIA methods: IMPACT 2002+, EDIP, and LUCAS (Bulle et al., 2019). This LCIA method also uses state-of-the-art characterization modeling that accounts for spatial variability in local and regional impact categories. This spatially-explicit modelling is used to provide global default CFs used for global impact categories (e.g., climate change) and for local and regional impact categories (e.g., eutrophication) when the location of an environmental release is unknown.

ReCiPe 2016 was used for the impact categories related to resource availability, which includes two midpoint impact categories (mineral use and fossil energy use) and three endpoint categories (mineral use, fossil energy use, water depletion) (Huijbregts et al., 2016). Three midpoint and

endpoint impact categories were developed to account for human health impacts from occupational injuries, illnesses, and fatalities. They are based on the method developed by Scanlon et al. (2013). Incidents of fatal injuries, incidents of long-term disabilities, and person-days of short-term disabilities are translated into weighted disability adjusted life years (DALYs). All injury and illness data were manually cross-walked from Bureau of Labor Statistics (BLS) Nature Codes to the World Health Organization's (WHO) International Classification of Diseases (ICD)-9 diagnosis codes to derive DALY equivalent. BLS Nature Codes, which describe the nature of the injury or illness, do not consistently align to ICD-9 codes, which characterize the treatment used to address the injury or illness. As such, expert judgment was required to pair BLS Nature Codes to ICD-9 codes. This mapping is consistent with the approach used in Scanlon et al. (2013).

Table 10. Midpoint Impact Categories

Impact Category	Units	LCIA Method
Mineral use	kg iron-eq	ReCiPe
Fossil energy use	kg oil-eq	ReCiPe
Water deprivation	m ³ deprived	IMPACT World+
Land use	m ² -yr arable-eq	IMPACT World+
Climate change	kg CO ₂ -eq	IMPACT World+
Human toxicity, respiratory effects, inorganics	kg PM _{2.5} -eq	IMPACT World+
Human toxicity, respiratory effects, organics	kg NMVOC-eq	IMPACT World+
Human toxicity, carcinogens, unspecified	CTUh	IMPACT World+
Human toxicity, carcinogens, unspecified, pesticide residue	CTUh	IMPACT World+
Human toxicity, non-carcinogens, unspecified	CTUh	IMPACT World+
Human toxicity, non-carcinogens, unspecified, pesticide residue	CTUh	IMPACT World+
Aquatic ecosystem toxicity	CTUe	IMPACT World+
Terrestrial acidification	kg SO ₂ -eq	IMPACT World+
Freshwater eutrophication	kg PO ₄ -eq	IMPACT World+
Marine eutrophication	kg N-eq	IMPACT World+

Table 11. Endpoint Impact Categories

Impact Category	Units	LCIA Method
Climate change	kg CO ₂ -eq	IMPACT World+
Resource availability, mineral use	MJ extra	ReCiPe
Resource availability, fossil energy use	MJ extra	ReCiPe
Resource availability, water depletion	MJ extra	Pfister et al.
Human health, respiratory inorganics	DALY	IMPACT World+
Human health, respiratory organics	DALY	IMPACT World+

Human health, cancer, unspecified	DALY	IMPACT World+
Human health, cancer, unspecified, pesticide residue	DALY	IMPACT World+
Human health, non-cancer, unspecified	DALY	IMPACT World+
Human health, non-cancer, unspecified, pesticide residue	DALY	IMPACT World+
Human health, water depletion	DALY	IMPACT World+
Ecosystem quality, terrestrial acidification	PDF-m ² -yr	IMPACT World+
Ecosystem quality, freshwater eutrophication	PDF-m ² -yr	IMPACT World+
Ecosystem quality, marine eutrophication	PDF-m ² -yr	IMPACT World+
Ecosystem quality, water depletion	PDF-m ² -yr	IMPACT World+
Ecosystem quality, water table lowering	PDF-m ² -yr	IMPACT World+
Ecosystem quality, land use	PDF-m ² -yr	IMPACT World+
Ecosystem quality, aquatic ecosystem toxicity	PDF-m ² -yr	IMPACT World+
Ecosystem quality, marine acidification	PDF-m ² -yr	IMPACT World+

5.3.2.2 Valuation of Impacts

Externalities are routinely monetized in cost benefit analysis (CBA) of public policy and private sector projects. For example, agencies developing rules that generate health and safety benefits commonly use the value of a statistical life (VSL) to monetize the benefits from reducing risks of premature mortality (Robinson, 2007). While DoD uses CBA to inform public investment decisions, monetization of life cycle benefits as described in DoD Guidance (2020a) is not often used. Instead, methods such as analysis of alternatives (AoA), cost-effectiveness analysis (CEA), and multi-criteria decision-making (MCDM) are used to quantify and compare nonmonetary benefits to monetary costs and benefits, (Melese et al., 2015). For example, Rohlf and Sullivan (2013) assessed the cost effectiveness of armored tactical vehicle replacement investments based on expenditures per life saved assuming a VSL threshold of \$7.5 million. Naval Sea Systems Command (NAVSEA) et al. (2020) assessed return on investment for diesel engine alternatives to the T-AGS(X) Oceanographic Survey Ship using monetized health benefit, occupational health exposure from primary and secondary particulate matter, and monetized life cycle impacts via LCA.

Outside of DoD, impacts are not routinely monetized in LCA. The International Organization for Standardization (ISO) 14040 LCA standard explicitly states that “there is no scientific basis for reducing LCA results to a single overall score or number, since weighting requires value choices.” As Pizzol et al. (2015) point out, there are challenges associated with using valuation in LCA, including assessment of potential rather than specific (actual) impacts; aggregation of impacts over space and time; and consideration of midpoints—which represent an intermediate state along a well-defined cause-effect relationship between a specific environmental flow and its impact—and endpoints, which represent the more complex and less well-defined relationship between environmental flows and damage to areas of protection, rather than specific targets. However, as the number of environmental and social impacts considered in LCA increases, the ability to semi-quantitatively compare impacts against economic and other costs and benefits will become increasingly important.

The DIO includes valuation as a semi-quantitative comparison tool. As described below, the conversion of human health effects, ecosystem effects, or resource depletion requires additional modelling and subjective choices. For example, costs associated with human health are quantified based on both an individual's willingness to pay for a certain level of well-being and the change in economic productivity associated with increased disability. Thus, the valuation of the impacts described above can be useful as an additional perspective on comparing impacts, but valuation does not objectively capture the true or full cost of impacts. As described in the SAG (DoD, 2020), the costs from valuation of impacts are different from direct and indirect costs (e.g., of purchase, maintenance, and disposition of a system—often referred to as life cycle costs). Therefore, and without exception, the valuation costs described here should never be summed with other costs.

In all, seven methods were used to estimate the external costs from these endpoints. The methods and cost factors are summarized in Table 12 and described here:

1. **Climate change.** The economic damages associated with marginal increases in carbon dioxide emissions are estimated using the social cost of carbon (SCC) values developed by the U.S. Interagency Working Group on Social Cost of Carbon (2013). The SCC is intended to be a comprehensive estimate of monetized damages related to climate change, including changes in agricultural productivity, human health, property damages from increased flood risk, and the value of ecosystem services.
2. **Human health.** Weidema's (2009) budget constraint method was used to estimate the monetary value of a DALY. This method assumes an individual's budget determines the amount they are willing (or able) to pay for a change in life expectancy or life quality, and that an individual's budget is constrained by their income. Since a DALY represents one year of healthy life, the potential annual economic production per capita at full well-being is used as the maximum an average person can pay for an additional life year, or the upper limit for the monetary value of a DALY. Weidema (2009) calculated the potential annual economic production per capita by multiplying the Gross Economic Product (GEP) of the United States by a series of factors to account for the effects of unemployment and underemployment, health and other work-disabling impacts, trade barriers, and missing education on economic production. GEP was assumed to be the sum of the Gross Domestic Product (GDP) and the Gross Household Production (GHP), which was estimated to be half of the GDP. Accounting for exchange rates and inflation, the result is a potential annual economic production per capita of \$129,745 USD₂₀₁₄. This is used as the monetary value of a DALY.
3. **Human productivity.** Human productivity is measured as the loss to society associated with the inability of an impacted individual to contribute productively to the economy. This top-down valuation is also based on human health risks expressed in terms of DALY. Costs are estimated as the per-capita decrease in GDP resulting from disabled life. When normalized to the DALY, this cost is \$40,326 USD₂₀₁₄/DALY.
4. **Ecosystem quality.** Due to the lack of willingness-to-pay values for ecosystem impacts, Weidema (2009) recommends the use of choice modeling to estimate the amount of well-being (i.e., DALYs) we are willing to sacrifice to prevent ecosystem loss (measured in potentially disappeared fraction of species on a meter squared during a year [PDF-m²-yr]). However, a general value for relating ecosystem loss to years of health life lost is not available. Instead, Weidema (2009) considers three possible terrestrial species-area to human well-being (PDF-m²-yr/DALY) exchange rates based on: (1) a relative importance

perspective (i.e., the value society places on ecosystem quality relative to the value placed on human health); (2) a revealed preference perspective (i.e., as demonstrated by policy-makers in political negotiations); and (3) a choice modeling perspective (i.e., as demonstrated in a single geographically-constrained study). The results were compared to environmental protection expenditures as a percentage of GDP in developed countries (as a proxy for the marginal willingness-to-pay for ecosystem protection), and a proxy valuation rate for ecosystem quality was proposed. Accounting for exchange rates and inflation results in a value of \$0.245 USD₂₀₁₄/PDF-m²-yr.

5. **Ecosystem quality, land use.** A revealed preference approach based on averted expenditure is used to determine the value of lost ecosystem services from land use. The U.S. Department of Agriculture's Conservation Reserve Program (CRP) provides cost-share and rental payments to farmers to remove highly erodible land or land in need of conservation buffers from row crop production and plant resource-conserving land cover to conserve and improve soil, water quality, and wildlife habitat. The CRP includes general and continuous sign-up practices, with the latter including land identified as having high-priority conservation issues under the Conservation Reserve Enhancement Program (CREP), previously farmed wetlands and wetland buffers under the Farmable Wetlands Program (FWP), and land selected for Non-CREP high-priority conservation practices. The cost of lost ecosystem productivity directly associated with land use was quantified as the cost of preserving one square meter (m²) of arable land equivalents for one year (yr), which was estimated to be \$0.0273 USD₂₀₁₄/m²-yr based on the cost of conserving highly-productive land included under the CRP's continuous sign-up programs.
6. **Resource productivity, water use.** Water depletion impacts human health (measured in DALYs), ecosystem quality (measured in PDF-m²-yr), and resource availability (measured in MJ extra). The external cost associated with the first two impacts are quantified using the external cost factors for human health and ecosystem quality, as described above. The units for measuring water resource availability, MJ extra, reflect the additional energy required to extract and deliver marginal units of water to future end users. For the resource availability indicators, Weidema (2009) and colleagues (2008) value MJ extra based on current energy prices. Accounting for exchange rates and inflation results in a value of \$0.0070 USD₂₀₁₄/MJ extra.
7. **Resource productivity, fossil energy and mineral use.**

Table 12. Valuation Categories, Values, and Methods

Valuation Category	Value	Units	Valuation Method
Climate change	0.04	USD ₂₀₁₄ /kg CO ₂ -eq	U.S. Social Cost of Carbon
Human health	129,745	USD ₂₀₁₄ /DALY	Weidema (2009)
Human productivity	40,326	USD ₂₀₁₄ /DALY	Weidema (2009)
Ecosystem quality	0.245	USD ₂₀₁₄ /PDF-m ² -yr	Weidema (2009)
Ecosystem quality, land use	0.0273	USD ₂₀₁₄ /m ² -yr	Weidema (2009)
Resource productivity, water use	0.0070	USD ₂₀₁₄ /MJ extra	Weidema (2009)

Resource productivity, fossil energy and mineral use	1	USD2014/USD2014	ReCiPe
--	---	-----------------	--------

5.3.3 Air Emissions Model

As shown in Table 8, an updated air emissions inventory was created for the DIO model. This inventory follows the approach used in the National Criteria and Hazardous Air Pollutant Totals By Industry 2017 v1.1 (Ingwersen et al., 2022) but expands the secondary contexts for more refined impact assessment. Specifically, this study includes new classification routines for two secondary contexts for air emissions available in the FEDEFL (Edelen et al., 2019): stack height and urban/rural location. Facility point-source emissions from the NEI are sorted into four labeled bins for stack height (h): ground ($h < 4\text{m}$), low ($4 \leq h < 25\text{m}$), high ($25 \leq h < 150\text{m}$), and very high ($h \geq 150\text{m}$). For facilities in the NEI and the Toxic Release Inventory (TRI), urban and rural labels are assigned by comparing facility global positioning system (GPS) coordinate data with the boundaries of Census urban areas and urbanized clusters (UAC), which are algorithmically defined at the block-level.^{6,7} With these expanded secondary context labels, flows from facilities are combined as in the prior approach to remove redundant flows based on facility and chemical identifiers within the Standardized Emission and Waste Inventories (StEWI) package.⁸ Additionally and separately, percentages of urban/rural population by county are calculated within Flow Sector Attribution (FLOWSA)⁹ to enable expanded secondary contexts for non-point source emissions. These percentages are derived from a combination of the same UAC shapefiles as the facility classifications plus Census block-level population data, which is only available via the decennial census (e.g., the 2017 UAC shapefile would use 2010 population data).

5.4 RESULTS AND DISCUSSION

5.4.1 Data Records

Data Records are listed in **Error! Reference source not found.** and include the DIO v2.0 model in Microsoft Excel and the useior model object in R Data format, as well as the data product of the new air emissions models, the National Criteria and Hazardous Air Pollutant Totals By Industry 2017 v2.0 in csv and Apache parquet formats.

Table 13. Data Records

Dataset	Note
DIO v2.0	Complete Model
National Criteria and Hazardous Air Pollutant Totals By Industry 2017 v2.0	Environmental Input

⁶ Ratcliffe et al., 2016. Defining Rural at the U.S. Census Bureau, American Community Survey and Geography Brief.

⁷ Census site. United States Census Bureau.

⁸ Young et al., 2022. A System for Standardizing and Combining U.S. Environmental Protection Agency Emissions and Waste Inventory Data. <https://doi.org/10.3390/app12073447>.

⁹ Birney et al., 2022. FLOWSA: A Python Package Attributing Resource Use, Waste, Emissions, and Other Flows to Industries. <https://doi.org/10.3390/app12115742>.

5.4.2 Technical Validation

DIO v2.0 was validated by checking that all the national flow totals by industry could be recalculated within 1% using equation analogous to the EEIO model calculation equation shown above, without the *C* characterization matrix, and after transforming the *B* elementary flow matrix into industry form and making dollar year adjustments. The total U.S. production demand vector is used in the validation. These transformations are shown in equations 28–31 of Ingwersen et al. (2022). The validation check passed for all 203,375 non-zero flow totals by industry. It should be noted that this validation is a check of the EEIO portion of the model only, and not the process-based extensions.

The National Criteria and Hazardous Air Pollutant Totals By Industry 2017 v2.0 was validated through comparison with a dataset computed without the secondary context information added in this model, to see if flowables by specific contexts could be aggregated back into flowables without context. All but 15 of the 563 flowables in this dataset showed an exact match after aggregation or were within 1%.

5.4.3 Code Availability

The datasets were built using the USEEIO modeling framework. DIO v2.0 was assembled using *useeior* v.1.2.1 along with a model configuration and input files, including the energy and process extensions. DIO v2.0 validation was performed using the *useeior* validation functions. The model configuration and input files along with notebooks to generate (DIO.Rmd) and validate (ValidateDIO.Rmd) the model can be found in the [USEPA/DIO](#) repository. The indicators were created using the LCIA Formatter v1.0.0. The National Criteria and Hazardous Air Pollutant Totals By Industry 2017 v2.0 dataset was created in [FLOWSA](#) v2.0.0 using the “CAP_HAP_2017_m2” flow-by-sector method. This method uses a detailed output of facility specific emissions data compiled using [StEWI](#).

5.5 REFERENCES

- Bulle, C., M. Margni, L. Patouillard, A.-M. Boulay, G. Bourgault, V. De Bruille, V. Cao, M. Hauschild, et al. 2019. IMPACT World+: a globally regionalized life cycle impact assessment method. *The International Journal of Life Cycle Assessment* 24: 1653–1674. doi:10.1007/s11367-019-01583-0.
- Crawford, R. H., P.-A. Bontinck, A. Stephan, T. Wiedmann, and M. Yu. 2018. Hybrid life cycle inventory methods – A review. *Journal of Cleaner Production* 172: 1273–1288. doi:10.1016/j.jclepro.2017.10.176.
- DoD. (2020). Sustainability Analysis Guidance: Integrating Sustainability into Acquisition Using Life Cycle Assessment. Version 7.0.
- Edelen, A. N., S. Cashman, B. Young, and W. Ingwersen. 2022. Life Cycle Data Interoperability Improvements through Implementation of the Federal LCA Commons Elementary Flow List. *Applied Sciences* 12: 9687. doi:10.3390/app12199687.
- EPA. 2022. Emissions & Generation Resource Integrated Database (eGRID). <https://www.epa.gov/egrid>

- Lenzen, M. 2000. Errors in Conventional and Input-Output—based Life—Cycle Inventories. *Journal of Industrial Ecology* 4: 127–148. doi:10.1162/10881980052541981.
- Suh, S., and G. Huppes. 2005. Methods for life cycle inventory of a product. *Journal of Cleaner Production* 13: 687–697.
- Suh, S., M. Lenzen, G. J. Treloar, H. Hondo, A. Horvath, G. Huppes, O. Jolliet, U. Klann, et al. 2004. System Boundary Selection in Life-Cycle Inventories Using Hybrid Approaches. *Environmental Science & Technology* 38: 657–664. doi:10.1021/es0263745.
- Tukker, A. 2006. Identifying Priorities for Environmental Product Policy. *Journal of Industrial Ecology* 10: 1–4. doi:10.1162/jiec.2006.10.3.1.
- Yang, Y., W. W. Ingwersen, T. R. Hawkins, M. Srocka, and D. E. Meyer. 2017. USEEIO: A new and transparent United States environmentally-extended input-output model. *Journal of Cleaner Production* 158: 308–318. doi:10.1016/j.jclepro.2017.04.150.
- Young, B., M. Srocka, W. Ingwersen, B. Morelli, S. Cashman, and A. Henderson. 2021. LCIA Formatter. *Journal of Open Source Software* 6: 3392. doi:10.21105/joss.03392.

6 CONCLUSIONS AND IMPLICATIONS FOR FUTURE RESEARCH

Identification, acquisition, and integration of data continues to pose a significant challenge to those who use LCA-based models. As part of Task 1 research, the team developed prototype software capable of generating reusable product system models (PSM) from a standardized product specification and automatically linking these models with existing LCI data from disparate sources. By using F-18 landing gear and electronic circuit board BOMs to more quickly generate unit process that comprise foreground models and link to background data, larger scale models representing complex systems, such as an entire aircraft, would be more easily achieved, aligning with DoD's (2018) Digital Engineering Strategy goal to "incorporate technological innovation to improve the engineering practice." Also, by advancing BOM-to-LCI automation in an open-source format, it is our hope that the community will more easily continue to build on this work. Greater investment of time into refining or improving these prototypes, or into further development of methods to tackle one of the four steps delineated by this study is likely to yield very useful tools that automate the data collection process important to conducting LCA studies. Particularly a prototype could be customized to link BOM-based DoD installation or equipment data with the DIO model. As the DoD SAG (2022a) recommends hybrid modeling, such a tool could help practitioners automatically link their primary data to the DIO model.

By aligning the DIO with the USEEIO model as was done in Task 3 and using EPA's existing model-building infrastructure, this will lower DoD's O&M costs for maintaining a large background dataset that includes all indirect activities in the economy and corresponding data on extractive, industrial, agricultural, transport, and disposal processes that captures their material and energy requirements and environmental emissions and wastes. EPA will continue to update the USEEIO, which gives DoD access to the latest authoritative background LCA modeling information and allows DoD to concentrate resources on development of defense specific process datasets currently lacking to complete LCA studies. Examples of such datasets DoD could focus on development of sustainable aviation fuel, propulsion systems, support infrastructure for decarbonization, materials needed for Military Construction (MILCON) activities, surface engineering and weapons systems sustainment.

An additional benefit as part of Task 3, is that an updated air emissions inventory was created for the DIO and USEEIO models. This advancement will serve as a national resource and increases resolution of impact assessment inferences that can be made by using both DIO and USEEIO. Future collaborations by DoD and EPA may yield additional advancements available for application across the U.S. economy. It is recommended that DoD establish a repository to contain DIO, as its authoritative LCA modeling support tool, and continue to regularly engage with EPA. Likewise, DoD participation in the Federal LCA Commons may spur additional synergies across Federal agencies.

Modeling the potential direct and indirect environmental impacts of chemicals, materials, and products is the aim of LCA studies. Life cycle impact assessment methods produce characterization factors (CF) that link data on emissions and resource use over the life cycle of a product to environmental, human health, climate change, and resource depletion threats. In Task 2, the research team developed new freshwater and marine eutrophication factors that spatially differentiate eutrophication CFs to improve the scientific robustness of fate and transport mechanism characterization. These efforts focused on the contiguous U.S. and update EPA

TRACI's (Bare, 2012) eutrophication potential impact category to provide a state-of-the-art global site-specific model for research and public use.

Related, SERDP and ESTCP investments could continue to advance research to incorporate primary ecological toxicity data to increase resolution of life cycle impact assessment, such as the 2022 PFAS-Free Foam Sustainability Analysis, which incorporated results from across the SERDP and ESTCP Environmental Restoration Program Area to increase the resolution of the USEtox database and create new product-specific CFs. Such investment would allow DoD to meet future requirements like the FY 2021 NDAA requirement to promote and advance the research and development of PFAS-free alternatives to AFFF and prioritize green and sustainable chemicals that do not pose a threat to public health or the environment.

In Task 2, the research team also created the LCIA Formatter tool, which produces standardized datasets of characterization factors optionally using the FLCAC Elementary Flow List (FEDEFL) as a standard flow list. The LCIA methods generated by the LCIA formatter for use with the FEDEFL can be hosted publicly on the FLCAC for use by LCA practitioners and researchers. These methods support LCAs performed by many parties, including FLCAC member agencies such as EPA, Department of Energy, Department of Agriculture, and DoD. The system was built to be flexible enough to support creating outputs for LCIA spatially-explicit characterization factors as those become more common.

Application of the DIO and results of this research will allow DoD to better meet new sustainability requirements as part of defense acquisitions. This potentially affects decision making over a large share of Federal procurement. Fiscal Year (FY) 2023 acquisition funding requested by DoD totalled \$276.0 billion, which includes funding totalling \$145.9 billion for Procurement and \$130.1 billion for RDT&E (DoD, 2022b). The scale of this potential impact is further put into perspective when considering that over half of the entire discretionary Federal budget is spent by DoD (USDT, 2022). This aligns with Executive Order (EO) 14057, *Catalyzing Clean Energy Industries and Jobs Through Federal Sustainability*, Section 102 requirement for the Federal Government “to use its scale and procurement power to achieve... net-zero emissions from Federal procurement.” This research supports incorporation of the social cost of greenhouse gases (SC-GHG) in acquisitions, this work directly supports the EO 14030, *Climate-Related Financial Risk*, requirement to consider SC-GHG in procurement decisions. Also, this research directly supports the FY 2021 National Defense Authorization Act (NDAA) Section 263 requirement for Agencies “to incorporate sustainable chemistry into existing research, development, demonstration [Programs].” As DoD's SAG evolves, it should reflect the updated nature of the DIO model. Also, to increase adoption of this research in support of acquisition and sustainability requirements, SERDP and ESTCP should provide additional guidance to principal investigators for how to deploy this research to estimate impacts, costs, and benefits of their own research.

In summary, the team offers these suggestions for future research.

- Make LCA tools more accessible to a larger body of practitioners by refining and integrating newly-developed product system model prototypes. (Task 1)
- Automate data collection processes important to conducting LCA and life-cycle based studies by utilizing BOM-based DoD installation or equipment data (i.e., primary data sources) with the DIO model. (Tasks 1 and 3)

- Advance life cycle impact assessment methods by updating characterization factors using DoD data (e.g., from SERDP and ESTCP projects concerning environmental restoration, PFAS-free alternatives, installation resilience, or waste reduction). (Task 2).
- Advance life cycle impact assessment methods through the development of new characterization factors using DoD data for defense-relevant impact categories (e.g., noise, chemicals and materials for sustainable energetics). (Task 2)
- Establish a more permanent engagement between DoD and the Federal LCA Commons to ensure continued collaboration and synergies across Federal agencies. (Tasks 2 and 3)
- Partner with SERDP and ESTCP affiliated agencies (i.e., DoD, EPA, DOE) and other executive branch agencies to leverage expertise, data, and models to support DoD life cycle-based studies using the DIO model. (Task 3)
- Establish formal engagement with DoD, DOE, and other agencies that have a stake in the underlying USEEIO model. The USEEIO framework serves as a de facto EEIO standard and will allow further development of domain specific application of EEIO modeling. (Task 3)
- Educate and guide principal investigators to evaluate the life cycle impacts, costs, and benefits of their research. (Task 3)
- Establish the DIO as an authoritative LCA modeling support tool. EPA will continue to update the USEEIO, which gives DoD access to the latest modeling information and allows DoD to concentrate resources on development of defense-specific process datasets. (Task 3)