Building Transparency

openIMPACT
Monte Carlo Algorithm

April 28, 2022

Acknowledgements

The authors gratefully acknowledge support from the MKA Foundation and other funders of Building Transparency. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

We would also like to thank representatives across government agencies, academia, and industry for their technical feedback and for taking their time in helping us shape this project.

Disclaimer

Some of this work uses data provided by manufacturers, however, reference therein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement or recommendation.

All images in this report were created by Building Transparency, unless otherwise noted.

For more information visit
www.buildingtransparency.org

openIMPACT
Monte Carlo Algorithm

Version: Draft 2
Date: April 26, 2022
Prepared by: Vaclav Hasik, Building Transparency

Table of Contents

1 Introduction

1.1 Purpose of this document

The purpose of this document is to provide a transparent description of the openIMPACT Monte Carlo Algorithm (MCA) and all of its supporting files. This document refers to product category calculations as examples of how the approach works, but the full models and parameters for each product category are documented separately.

1.2 Intended application

The MCA approach described in this document was developed to enable more realistic estimation of the potential variation in environmental impacts of products across production scenarios and regions. It is developed to support the estimation of industry-wide impact ranges to be used in building design tools (e.g., EC3 and Tally) but can also be used for estimating variability and uncertainty of impacts for individual manufacturers.

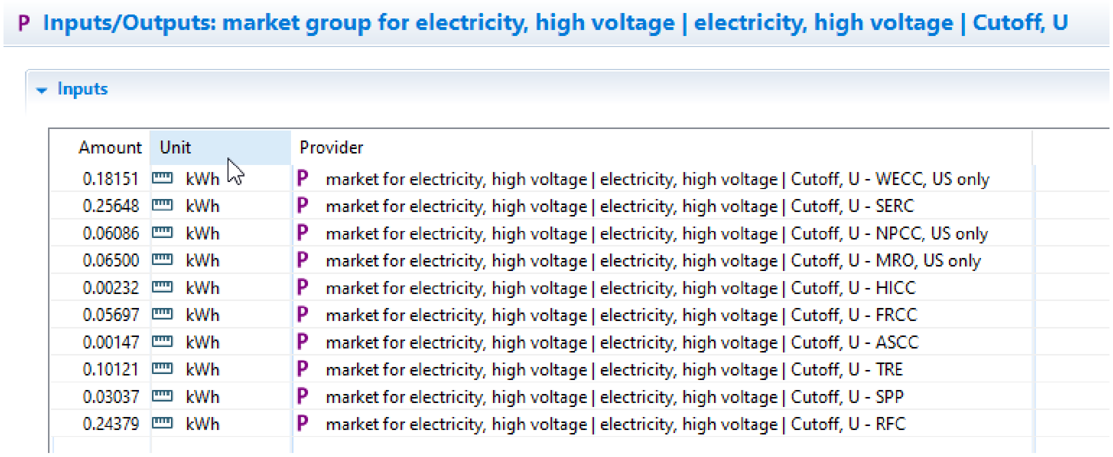
1.3 Problem statement

Life Cycle Assessment (LCA) models are typically set up as deterministic models capturing average conditions in a company or industry. Capturing the range of likely outcomes is seldom done rigorously but is useful to identify areas where product selection can significantly affect outcomes, the parts of the supply chain where transparency is essential, and the uncertainty inherent to estimates based on an average scenario. The openIMPACT MCA uses probabilistic modeling to quantify these ranges in a way that aligns well with published EPDs for actual products.

Uncertainty estimation has been an afterthought in LCA modeling addressed by 1) applying basic uncertainty distributions to input quantities and 2) by using data quality indicators (DQIs). Unfortunately, this approach misses two significant factors which drive variability and uncertainty for most products: 1) the true variation in the particular or regional supply chain by choosing one provider of energy or feedstock materials over another (i.e., switching providers), and 2) the interdependence of some parameters within the LCA models. As a result, these measures offer little insight into either the realistic worst-case scenario or the potential for improvements.

1.3.1 US electricity grid example

The first issue can be demonstrated by looking at the US electricity grid. The typical approach to model US average electricity is to simply list the proportion of each input coming from a particular electricity generating region (see Figure 1), instead of the provider being sampled one-by-one according to its probability of being picked for any single production scenario (i.e., producing any one unit of a product anywhere in the US).



P Inputs/Outputs: market group for electricity, high voltage | electricity, high voltage | Cutoff, U

Inputs

Amount	Unit	Provider
0.18151	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - WECC, US only
0.25648	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - SERC
0.06086	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - NPCC, US only
0.06500	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - MRO, US only
0.00232	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - HICC
0.05697	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - FRCC
0.00147	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - ASCC
0.10121	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - TRE
0.03037	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - SPP
0.24379	kWh	P market for electricity, high voltage | electricity, high voltage | Cutoff, U - RFC

Figure 1. Inputs into an openLCA unit process representing US Average Electricity.

The histogram in Figure 2 shows the difference in the statistical results for the Global Warming Potential (GWP) of the US electricity grid using an average deterministic approach (orange) versus a probabilistic approach (blue). The average approach uses basic uncertainty and DQIs defined in the ecoinvent 3.5 database, while the probabilistic approach shows the distribution of results of choosing one electricity provider at a time based on its probability of providing unit of energy to an end user anywhere in the US at any given time.

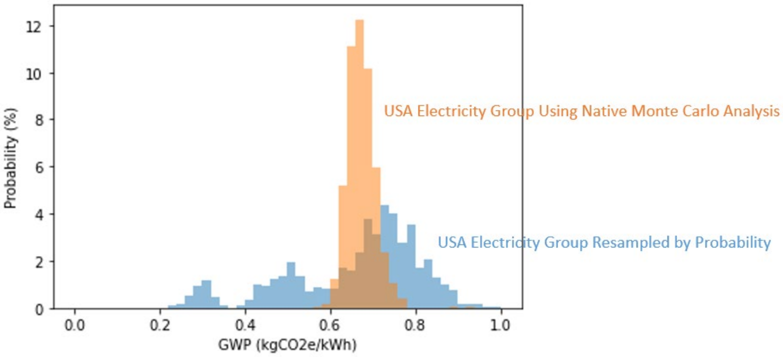


Figure 2. Comparison of statistical results by using an average approach (orange) vs probabilistic approach (blue).

openIMPACT models are set up with 1) single provider inputs for each flow (unless the real-world processes source flows from multiple providers simultaneously), and 2) dependent parameters whenever possible. The openIMPACT algorithm is set up as a Monte Carlo simulation varying providers and parameters according to their probability distributions.

1.3.2 Data quality indicators, variability, and uncertainty

While variability and uncertainty in LCA are considered two separate issues, they both combine to a resulting uncertainty on the point estimate of GWP for a specific product procured for a construction project. LCA models can carry information about the variability of parameters within the model and the representativeness of each quantity and model (referred to as uncertainty). An approach to uncertainty in ecoinvent and other databases is to rate a data set according to data quality indicators (DQIs) that state how representative the data set is of the actual scenario being modeled, and sum up the uncertainties associated with any non-representativeness [1,2]. These DQIs certainly illustrate things an LCA practitioner should seek to maximize, but they do not offer much insight as to how far the actual impact of a specific product might vary from the LCA estimate. For example:

1. Uncertainty does not scale nicely with geographical representativeness.
    a. Inputs such as electricity can be extremely sensitive to small changes in location. For example, the difference between the British Columbia grid (40kgCO2/GWh) and Alberta grid (790kgCO2/GWh) is far greater than the difference between the North American and European averages. Some products (e.g., aluminum, electric arc furnace (EAF) steel) are extremely sensitive to electricity sources, while others (e.g., concrete, basic oxygen furnace (BOF) steel) are minimally sensitive to electricity.
    b. Inputs such as diesel fuel have minimal geographical variation.
    c. In some cases, the location of final product manufacturing is nearly irrelevant because the bulk of emissions come from supply chain inputs that may be sourced from far away.
2. Uncertainty does not scale nicely with technological representativeness.
    a. For example, GWP of blast furnace iron and coal-based direct-reduced iron (DRI) can be quite similar, while coal-based DRI and hydrogen DRI are wildly different.
3. The standard deviation of the sample set is usually not disclosed, which makes the completeness and reliability measures somewhat moot from an uncertainty point of view.

Our study of construction materials has found that most products have just a handful of factors, specific to each product category, that dominate whether a particular product's carbon intensity is relatively high or low. Simulating the possible outcomes for these handful of factors yields useful ranges for the product category that can serve as uncertainty estimates and as guidelines for where supply chain transparency is essential.

2 How it works

The basic approach is to start from a base LCA model, find the major sources of emissions, vary the quantities and input providers (i.e., supply chain options), and observe the effect on the result. Substituting a range of possible sources in a Monte Carlo fashion yields the range of possible outcomes.

2.1 Simulation overview

Before proceeding with a simulation, there needs to be the following set of items:

- openLCA[1] base model (section 2.2.1),
- substitution sheets (section 2.2.2), and
- provider sheets (section 2.2.3).

openLCA then needs to be allowed communication with Python via the openLCA IPC before executing the simulation algorithm.

At a high level, the algorithm goes through the following steps:

1. Select substitution sheet
2. Load all information from the substitution sheet
    a. Main process used to setup a product system
    b. Upstream processes that are to have their providers substituted or parameters adjusted
    c. List of provider sheets used to sample providers
    d. List of parameter information and their distributions or lists of values
3. Setup CSV results file
4. Start Monte Carlo analysis
    a. Select a provider from each provider sheet based on probability
    b. Modify each upstream process with the newly selected providers
    c. Create a product system from the main process (this now includes the full setup of the product system with all linked providers defined in step 4.b)
    d. Select a parameter value for all parameters identified on the substitution sheet
    e. Setup and run openLCA analysis using the product system and parameters defined in 4.c and 4.d
    f. Save results to CSV from step 3 and display histogram with all results up to this point
    g. Delete the product system

Step 4 is repeated as many times as defined by the user. The CSV results files record the providers, parameters, and results for each iteration. The algorithm includes additional steps for finding reference flows, input flows to be modified, and checks for any needed unit conversions.

Note that the IPC client that allows Python integration with openLCA 1.11 has some bugs which currently require the product system to be created and deleted every time there is a change to the underlying processes, hence the need for steps 4.c and 4.g. These steps may be omitted in future versions of openLCA.

---

[1] The MCA approach can be used with any LCA tool, but the current openIMPACT Python script is built around openLCA.

2.2 Supporting files

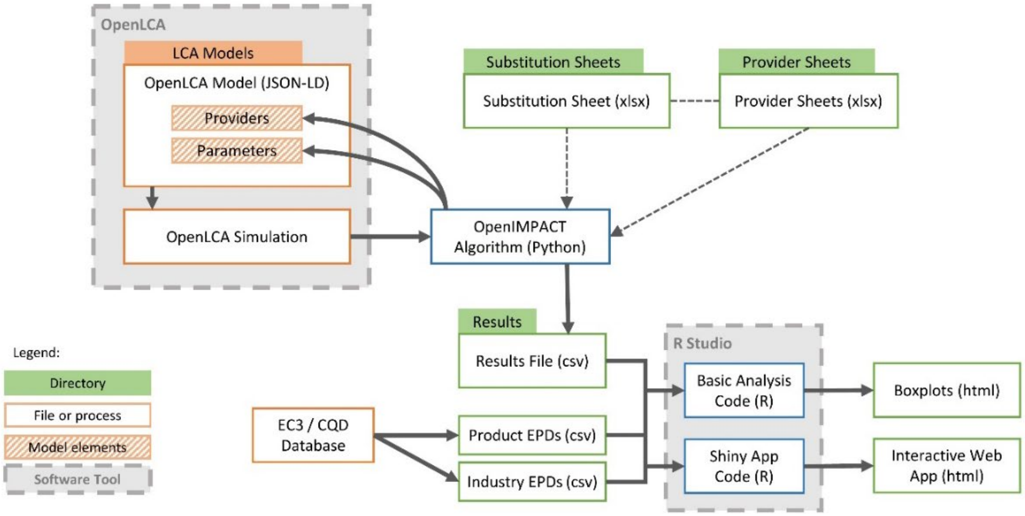


Figure 3. Structure of the openIMPACT scripts and supporting files.

Figure 3 shows an overview of the supporting files, openIMPACT MCA, openLCA, LCA models, results, and analysis. All supporting files and examples of results can be found on buildingtransparency.org.

2.2.1 OpenLCA models

File format: JSON-LD

Our implementation uses openLCA as the calculation engine, therefore it requires an openLCA model to be able to run the simulation. We share our core LCA models developed in openLCA as JSON-LD files. These files can be re-imported into openLCA. Note that the model may need other databases, such as ecoinvent or USLCI, linked to it before proceeding with the simulation.

2.2.2 Substitution sheets

File format: XLSX

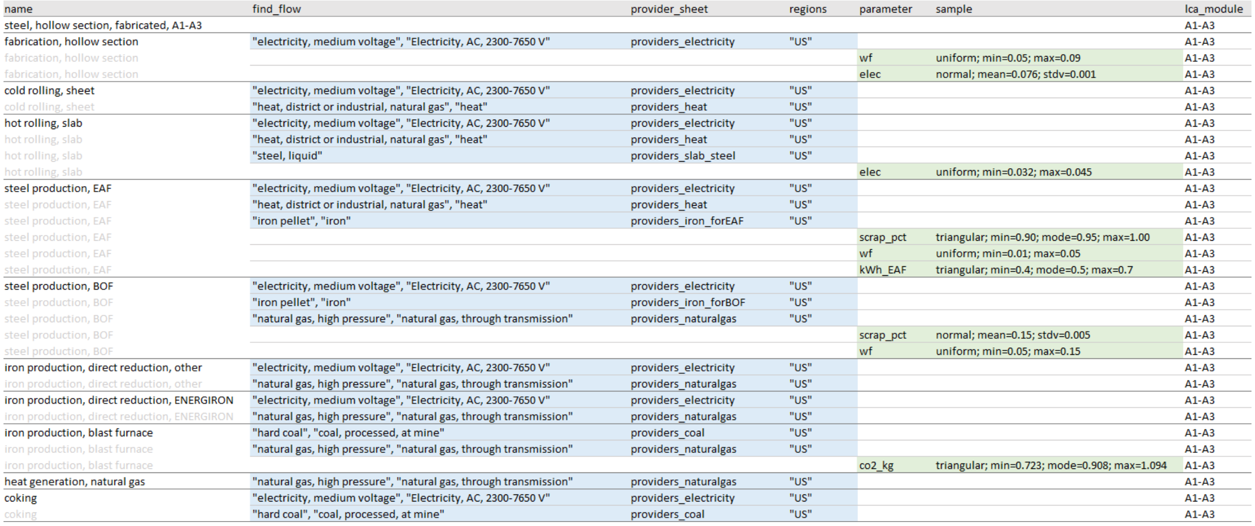
These are data sheets specifying which parts of the LCA model (i.e., the product system) are to be modified using the MCA. Figure 4 shows an example of a substitution sheet, which includes:

1. Name and UUID of each "process" in the product system which is to be modified
2. All possible names of the input flows which are to have their "default provider" substituted – this is how the algorithm selects a specific supply-chain route
3. Name of the "provider sheet" (see section 2.1.3) and the region to use for selecting a new "default provider"
4. Parameter name and the list of numeric samples or sample distribution
5. Tag for which LCA modules each process belongs to – this is currently not used by the algorithm, but will be used in the future for running full cradle-to-grave analyses in a single execution

The full list of column definitions is shown in Table 1.

Table 1. Substitution sheet column definitions.

Column	Description
name	Name of the process to have its flows and provider modified.
location	Location of the process to be modified. Can be empty if the OpenLCA process does not have any assigned location.
uuid	Uuid of the process to have its flows and providers modified.
find_flow	List all the possible flow names to be modified. Normally this may be a single target flow, but if it is being replaced by a flow with a different name, that flow also needs to be listed. This is because once the flow is replaced, the next iteration needs to look for the updated flow name.
provider_sheet	Name of the sheet with the probability information for the particular process and flow.
regions	Some of the provider probability sheets may include more than one region, for example they may include US and Canada, or they may include global data. You can provide data for all regions in the provider sheets and narrow down the scope of the analysis by specifying I list of select regions on this substitution sheet. (NOTE: this is not yet implemented. The regional scope selection is instead harcoded into the script. Look for the specification of "US" or "CA".)
parameter	Name of the parameter in openLCA.
sample	The set of values to be sampled from during parameter redefinition. The inputs can be in the following formats: list; #, #, #, #, # uniform; min=#; max=# normal; mean=#; stdv=# triangular; min=#; mode=#; max=# lognormal; gmean=#; gstdv=#
skip	You can put placeholders in this sheet and mark them as skip "Yes". These rows will be skipped during the analysis.
lca_module	Tag for which LCA module a particular process corresponds to from the perspective of the product system.

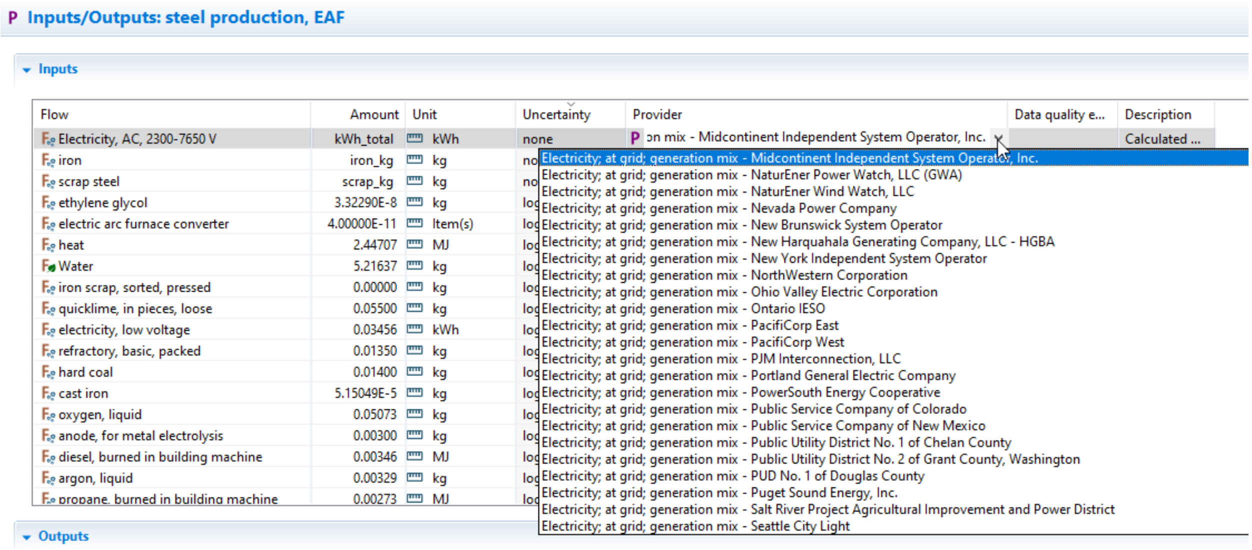


name	find_flow	provider_sheet	regions	parameter	sample	lca_module
steel, hollow section, fabricated, A1-A3						A1-A3
fabrication, hollow section	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
fabrication, hollow section				wf	uniform; min=0.05; max=0.09	A1-A3
fabrication, hollow section				elec	normal; mean=0.076; stdv=0.001	A1-A3
cold rolling, sheet	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
cold rolling, sheet	"heat, district or industrial, natural gas", "heat"	providers_heat	"US"			A1-A3
hot rolling, slab	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
hot rolling, slab	"heat, district or industrial, natural gas", "heat"	providers_heat	"US"			A1-A3
hot rolling, slab	"steel, liquid"	providers_slab_steel	"US"			A1-A3
hot rolling, slab				elec	uniform; min=0.032; max=0.045	A1-A3
steel production, EAF	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
steel production, EAF	"heat, district or industrial, natural gas", "heat"	providers_heat	"US"			A1-A3
steel production, EAF	"iron pellet", "iron"	providers_iron_forEAF	"US"			A1-A3
steel production, EAF				scrap_pct	triangular; min=0.90; mode=0.95; max=1.00	A1-A3
steel production, EAF				wf	uniform; min=0.01; max=0.05	A1-A3
steel production, EAF				kWh_EAF	triangular; min=0.4; mode=0.5; max=0.7	A1-A3
steel production, BOF	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
steel production, BOF	"iron pellet", "iron"	providers_iron_forBOF	"US"			A1-A3
steel production, BOF	"natural gas, high pressure", "natural gas, through transmission"	providers_naturalgas	"US"			A1-A3
steel production, BOF				scrap_pct	normal; mean=0.15; stdv=0.005	A1-A3
steel production, BOF				wf	uniform; min=0.05; max=0.15	A1-A3
iron production, direct reduction, other	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
iron production, direct reduction, other	"natural gas, high pressure", "natural gas, through transmission"	providers_naturalgas	"US"			A1-A3
iron production, direct reduction, ENERGIRON	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
iron production, direct reduction, ENERGIRON	"natural gas, high pressure", "natural gas, through transmission"	providers_naturalgas	"US"			A1-A3
iron production, blast furnace	"hard coal", "coal, processed, at mine"	providers_coal	"US"			A1-A3
iron production, blast furnace	"natural gas, high pressure", "natural gas, through transmission"	providers_naturalgas	"US"			A1-A3
iron production, blast furnace				co2_kg	triangular; min=0.723; mode=0.908; max=1.094	A1-A3
heat generation, natural gas	"natural gas, high pressure", "natural gas, through transmission"	providers_naturalgas	"US"			A1-A3
coking	"electricity, medium voltage", "Electricity, AC, 2300-7650 V"	providers_electricity	"US"			A1-A3
coking	"hard coal", "coal, processed, at mine"	providers_coal	"US"			A1-A3

Figure 4. Example substitution sheet.

The way the substitution sheets are used is similar to the way users select default providers in openLCA. As shown in Figure 5, openLCA allows users to select a default provider (e.g., “Electricity, at grid, generation mix – Midcontinental Independent System Operator”) from a list of providers that produce the flow of interest (e.g., “Electricity, AC, 2300-7650 V”). The selection in openLCA has to be done manually and the list of possible providers includes a complete list of providers that produce the given flow even if the providers are not fully equivalent alternatives. The MCA uses provider sheets (section 2.2.3) to hold a specific subset of possible providers and the market weight of each provider. The substitution sheet holds the information on which provider sheet is to be used for substitution of default providers for a specific flow in a specific process in the product system.

Substitution of providers using this approach is useful for industry-wide analyses, multi-regional analyses, or analyses of products that may use multiple supply chains (e.g., steel fabricator sourcing varying amounts of BOF and EAF steel feedstock).



Inputs/Outputs: steel production, EAF

Inputs

Flow	Amount	Unit	Uncertainty	Provider	Data quality e...	Description
Electricity, AC, 2300-7650 V	kWh_total	kWh	none	on mix - Midcontinent Independent System Operator, Inc.		Calculated ...
iron	iron_kg	kg	no			
scrap steel	scrap_kg	kg	no			
ethylene glycol	3.32290E-8	kg	log			
electric arc furnace converter	4.00000E-11	Item(s)	log			
heat	2.44707	MJ	log			
Water	5.21637	kg	log			
iron scrap, sorted, pressed	0.00000	kg	log			
quicklime, in pieces, loose	0.05500	kg	log			
electricity, low voltage	0.03456	kWh	log			
refractory, basic, packed	0.01350	kg	log			
hard coal	0.01400	kg	log			
cast iron	5.15049E-5	kg	log			
oxygen, liquid	0.05073	kg	log			
anode, for metal electrolysis	0.00300	kg	log			
diesel, burned in building machine	0.00346	MJ	log			
argon, liquid	0.00329	kg	log			
propane, burned in building machine	0.00273	MJ	log			

Electricity; at grid; generation mix - Midcontinent Independent System Operator, Inc.
Electricity; at grid; generation mix - NaturEner Power Watch, LLC (GWA)
Electricity; at grid; generation mix - NaturEner Wind Watch, LLC
Electricity; at grid; generation mix - Nevada Power Company
Electricity; at grid; generation mix - New Brunswick System Operator
Electricity; at grid; generation mix - New Harquahala Generating Company, LLC - HGBA
Electricity; at grid; generation mix - New York Independent System Operator
Electricity; at grid; generation mix - NorthWestern Corporation
Electricity; at grid; generation mix - Ohio Valley Electric Corporation
Electricity; at grid; generation mix - Ontario IESO
Electricity; at grid; generation mix - PacifiCorp East
Electricity; at grid; generation mix - PacifiCorp West
Electricity; at grid; generation mix - PJM Interconnection, LLC
Electricity; at grid; generation mix - Portland General Electric Company
Electricity; at grid; generation mix - PowerSouth Energy Cooperative
Electricity; at grid; generation mix - Public Service Company of Colorado
Electricity; at grid; generation mix - Public Service Company of New Mexico
Electricity; at grid; generation mix - Public Utility District No. 1 of Chelan County
Electricity; at grid; generation mix - Public Utility District No. 2 of Grant County, Washington
Electricity; at grid; generation mix - PUD No. 1 of Douglas County
Electricity; at grid; generation mix - Puget Sound Energy, Inc.
Electricity; at grid; generation mix - Salt River Project Agricultural Improvement and Power District
Electricity; at grid; generation mix - Seattle City Light

Outputs

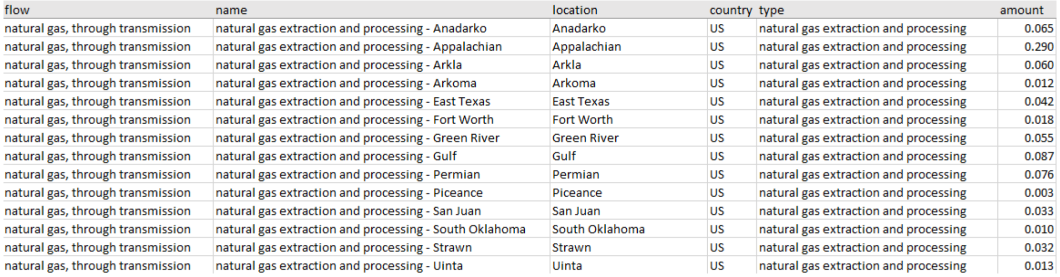
Figure 5. Selection of a "default provider" for a flow in openLCA.

Although this does not currently affect the openIMPACT models, there is one limitation in the current provider substitution approach which stems from identifying targeted input flows only by their name. If there are multiple flows in a process with the same name, their providers will all be substituted with the same provider. The algorithm does not allow substitution of identical flows within a single process with two different providers. This could be improved in the future by adding additional identifiers to the selection of flows that are to be modified.

2.2.3 Provider sheets

File format: XLSX

These are data sheets holding information about all possible providers that can be used as inputs into a particular process, and the probability of each provider being used in each iteration of the MCA. The probability values in the "amount" column can be absolute or percent values representing market share information or other determinant of the probability of a certain provider being selected. The only condition is that the amount is consistent throughout the sheet (i.e., that percentage values are not mixed with absolute values). The MCS algorithm recalculates the probabilities for each subset of values, ensuring that the sum of all probabilities adds up to 100%.



flow	name	location	country	type	amount
natural gas, through transmission	natural gas extraction and processing - Anadarko	Anadarko	US	natural gas extraction and processing	0.065
natural gas, through transmission	natural gas extraction and processing - Appalachian	Appalachian	US	natural gas extraction and processing	0.290
natural gas, through transmission	natural gas extraction and processing - Arkla	Arkla	US	natural gas extraction and processing	0.060
natural gas, through transmission	natural gas extraction and processing - Arkoma	Arkoma	US	natural gas extraction and processing	0.012
natural gas, through transmission	natural gas extraction and processing - East Texas	East Texas	US	natural gas extraction and processing	0.042
natural gas, through transmission	natural gas extraction and processing - Fort Worth	Fort Worth	US	natural gas extraction and processing	0.018
natural gas, through transmission	natural gas extraction and processing - Green River	Green River	US	natural gas extraction and processing	0.055
natural gas, through transmission	natural gas extraction and processing - Gulf	Gulf	US	natural gas extraction and processing	0.087
natural gas, through transmission	natural gas extraction and processing - Permian	Permian	US	natural gas extraction and processing	0.076
natural gas, through transmission	natural gas extraction and processing - Piceance	Piceance	US	natural gas extraction and processing	0.003
natural gas, through transmission	natural gas extraction and processing - San Juan	San Juan	US	natural gas extraction and processing	0.033
natural gas, through transmission	natural gas extraction and processing - South Oklahoma	South Oklahoma	US	natural gas extraction and processing	0.010
natural gas, through transmission	natural gas extraction and processing - Strawn	Strawn	US	natural gas extraction and processing	0.032
natural gas, through transmission	natural gas extraction and processing - Uinta	Uinta	US	natural gas extraction and processing	0.013

Figure 6. Example provider sheet.

2.2.4 Reference sheets

File format: XLSX

These are additional reference tables needed for parametric calculations of some construction products, e.g., table referencing steel gauges and their equivalent thickness. These sheets are currently not integrated into the MCS algorithm and are used only for reference when creating the substitution sheets.

2.3 Results files

2.3.1 Raw results files

File format: CSV

Each iteration of the MCS is saved into a csv file. Each row of this file represents one iteration of the MCS and includes information about the selected providers, selected parameter values, and impact results.

Note that this file tracks all provider selections and parameter substitutions, but some of them may be redundant in a specific iteration. For example, if an iteration of a steel plate model uses a BOF steelmaking route, it will still show modifications done to the EAF process as well. The modifications to an EAF process won't have any effects on the LCA calculations in that iteration but will still be shown in the raw results file.

2.3.2 Comparison data files

File format: CSV

These are csv files with EPD data from EC3. The files include additional columns for identifying industry vs product EPDs, regional information, and label names (i.e., select text to show on the interactive boxplots).

2.3.3 Comparison boxplots

File format: HTML

These are interactive boxplots showing the statistical results of the openIMPACT simulations next to related Industry-wide and Product-specific EPDs. These plots are useful for validation of the openIMPACT models, but also for understanding the state of existing EPD data.

2.3.4 Simulation logs

File format: TXT

The MCS algorithm provides live feedback about the progress of the simulation. This can be useful for understanding how the algorithm works, understanding what steps it took for setting up each iteration of the MCS, for validation and debugging.

2.3.5 Interactive filtering app

File format: web application

Prototype of an interactive filtering app that can be used to quickly assess the statistics of a subset of the data generated using the MCS.

2.4 Script files

openimact_mca.py is the main script for running the openIMPACT MCA.

openimpact_plots.r is a template script for generating interactive boxplots.

3 Requirements

1. Python 3.9 or later
2. Python packages listed at the beginning of the "olca_openimpact.py" code
3. Integrated development environment (e.g., PyCharm)
4. OpenLCA 1.11 or later
5. R and RStudio (if using the openIMPACT R scripts for analysis)

4 References

[1] R. Frischknecht, N. Jungbluth, H.-J. Althaus, G. Doka, R. Dones, T. Heck, S. Hellweg, R. Hischier, T. Nemecek, G. Rebitzer, The ecoinvent database: Overview and methodological framework, Int J Life Cycle Assess. 10 (2005) 3–9. https://doi.org/10.1065/lca2004.10.181.1.

[2] B.P. Weidema, C. Bauer, R. Hischier, C. Mutel, T. Nemecek, J. Reinhard, C.O. Vadenbo, G. Wernet, Overview and methodology: Data quality guideline for the ecoinvent database version 3, Swiss Centre for Life Cycle Inventories, 2013.