Where is my footprint located? Estimating the geographical variance of Hybrid-LCA footprints

Arthur Jakobs (✉ arthur.jakobs@indecol.uni-freiburg.de)
Albert-Ludwigs-Universität Freiburg Fakultät für Umwelt und Natürliche Ressourcen: Albert-Ludwigs-Universität Freiburg Fakultät für Umwelt und Natürliche Ressourcen

Simon Schulte (✉ simon.schulte@indecol.uni-freiburg.de)
Albert-Ludwigs-Universität Freiburg Fakultät für Umwelt und Natürliche Ressourcen: Albert-Ludwigs-Universität Freiburg Fakultät für Umwelt und Natürliche Ressourcen

Stefan Pauliuk (✉ stefan.pauliuk@indecol.uni-freiburg.de)
Albert-Ludwigs-Universität Freiburg Fakultät für Umwelt und Natürliche Ressourcen: Albert-Ludwigs-Universität Freiburg Fakultät für Umwelt und Natürliche Ressourcen

Research Article

Keywords: hybrid-LCA, uncertainty, life cycle assessment, environmentally extended input-output analysis, regionalisation, trade statistics, LCA, environmental footprints

DOI: https://doi.org/10.21203/rs.3.rs-1442198/v3

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Where is my footprint located? Estimating the geographical variance of Hybrid-LCA footprints

Arthur Jakobs¹, Simon Schulte¹, Stefan Pauliuk¹

¹ Indsutrial Ecology Freiburg, University of Freiburg, Freiburg, Germany.

Correspondence: Arthur Jakobs, arthur.jakobs@indecol.uni-freiburg.de, Tennenbacher Str. 4, 79106 Freiburg, Germany.

Abstract

Hybrid- Life Cycle Assessment (LCA) mostly fails to fully exploit valuable information from Multi-Regional Input-Output (MRIO) models by aggregating regional supply chains to the lower geographical resolution of process LCA databases. We propose a method for sampling the various individual regions within the aggregated regional scope of LCA processes. This sampling maximises the information content of hybrid LCA footprint results by preserving the regional variance, and it allows for regional price distributions from BACI/UN-COMTRADE international trade statistics to be used to simultaneously improve the accuracy of the hybrid model. This work analyses the impact of regional variance and the use of regional price distributions on the uncertainty of the hybrid footprint results for both the carbon (GWP100) and land use footprint, and compares this to the variance resulting from only price- and only regional variance. We find that the median process footprint intensity increases by $7^{+18}_{-3}\%$ for the GWP100 due to hybridisation, and $90^{+143}_{-23}\%$ for the Land Use footprint. Results show that the magnitude of the footprint uncertainty strongly depends on the product sector of the LCA process and environmental impact considered. In a case study of Swiss household consumption we find truncation error estimates of $8.4^{+9.2}_{-2.7}\%$ for the GWP100 and $36^{+64}_{-14}\%$ for the land use footprint. Our results highlight the importance of regionalisation of process LCA databases, as it has the potential to significantly improve both the precision and accuracy of hybrid LCA models.

KEYWORDS: hybrid-LCA, uncertainty, life cycle assessment, regionalisation, environmental footprints, industrial ecology

1 Introduction

Hybrid Life Cycle Assessment (HLCA) aims to combine the strengths of the detailed Process-based Life Cycle Assessment (PLCA) and more complete Environmentally-Extended Input-Output Analysis (EE-IOA) while trying to mitigate their weaknesses (Suh 2004; Suh and Huppes 2005; Peters 2010). While PLCA offers a detailed description of a
products supply chain, it consequently has a limited system boundary due to the lack of a complete description of the economy at the required level of detail (Suh et al., 2004; Agez et al., 2020). Therefore PLCA calculations mostly underestimate the environmental impacts of the product system under study, a problem also known as the truncation error (Lenzen, 2000; Ward et al., 2018). EE-IOA on the other hand, and more in particular Multi-Regional Input-Output (MRIO) offers a complete picture of economy and it’s supply chains. This completeness however, comes at the expense of detail, with MRIO dealing with aggregated industries and ‘product groups’ as opposed to the individual processes and products in PLCA, making it an impractical tool in cases where a higher resolution is needed (Peters, 2010). Various methods have been proposed to combine these two approaches in the above mentioned HLCA, with some opting to bring detail to the complete (MR-) IO models in the so called ‘matrix augmentation’ and ‘path exchange’ methods, while others aim to complete PLCA supply chains using (MR-) IO data with ‘tiered-’ and ‘integrated hybrid’ models (Crawford et al., 2018 and references therein). In recent years, there have been increased efforts to automate the hybridisation of PLCA databases with EEIO models using both the tiered and path exchange methods (Yu and Wiedmann, 2018; Agez et al., 2020; Stephan et al., 2019; Agez et al., 2022; Crawford et al., 2021), however, so far only Agez et al. (2020, 2022) build a complete global hybrid database using a tiered method to combine the ecoinvent (PLCA) and EXIOBASE (global MRIO) databases.

Both PLCA and EE-IOA have seen a strong trends of increasing regionalisation (PLCA) (Mutel and Hellweg, 2009; Feifel et al., 2010; Frischknecht et al., 2019) or geographic resolution (EE-IOA), (Lenzen et al., 2017; Iowa et al., 2020; Wang et al., 2017; Huo et al., 2022). However, because of the granular nature of the process data, the geographical scope and resolution of PLCA databases have not caught up with the modern global MRIO models that have been developed over the last decade (Tukker and Dietzenbacher, 2013; Peters et al., 2011; Lenzen et al., 2012; Dietzenbacher et al., 2013; Stadler et al., 2018). In practice this means that although a region specific process descriptions are available in large Life Cycle Inventory (LCI) databases such as ecoinvent (Steubing et al., 2016; Wernet et al., 2016), most processes are only covered by generic ‘global’ or ‘rest of world’ (RoW) datasets. Such global or RoW datasets are in turn mostly extrapolations from datasets specific to a particular region (Weidema et al., 2013). As a consequence, the disparity between regional resolution of PLCA and global MRIO models has been a limiting factor to fully exploit the high regional resolution of the latter in complete hybrid LCA databases such as presented in Agez et al. (2020), where the supply chains of LCA processes with an aggregated regional scope are complemented with the sectoral average of the matching regions in the MRIO model. By averaging the supply chain structure and sectoral impacts over various regions, more detailed information is lost. Moreover, Jakobs et al. (2021) showed that the lack of regional resolution increases the variance of the reference product prices, needed to link the monetary MRIO models to the physical supply chains of the PLCA data, leading to a higher variance in the footprint results.

In this paper we present an extension to the hybridisation of ecoinvent and EXIOBASE developed by Agez et al. (2020, 2022), in which we make geographic variance explicit by sampling the individual regions contained in the geographic scope of the LCA processes. In doing so, we preserve the information content contained in the full model
by presenting outcome probability distributions rather than point estimates. Moreover, it also allows us to use regional price distributions with a higher price variance (Jakobs et al., 2021). We analyse the impact of geographic variance (differences between supply chain impacts of a product sector in different regions of the MRIO model) and price variance (the variability of commodity prices used to link the PLCA processes to the background MRIO model) on the footprint intensities (footprint per unit of process output), for both the Global Warming Potential 100 (GWP100) as well as the land occupation footprint. We chose these two footprint indicators because the bulk of land occupation impacts, associated with transformation of natural ecosystems and biodiversity loss (Bruckner et al., 2015; Többen et al., 2018, and references therein), often occurs further up the supply chain and therefore suffers from a higher truncation error in PLCA studies than the relatively well covered greenhouse gas emissions (Agez et al., 2022). As such, land occupation footprints can benefit more strongly from hybrid accounting models (Bruckner et al., 2015).

To illustrate the impact of spatial variance and the use of regional prices on the uncertainty of a consumption footprint, we follow Jakobs et al. (2021) and use the PLCA part of the model of Swiss household consumption of Froemelt et al. (2018). For this illustration case, we analyse the impact of geographical variance, price variance and the combined effect on the GWP and land-use footprint.

This paper is structured as follows: In section 2 we first introduce the tiered HLCA model and the cut-off matrix. We then discuss how the disparity in geographical resolution between the different data affects the model and how we can resolve this (section 2.2) before introducing regionalised prices into the model (2.3). The various data used in this project are discussed in section 2.4 and the Monte Carlo simulation process is introduced in section 2.5. In section 3 the results of the process level footprint intensities (section 3.1) and the consumption basket (section 3.2) are presented. We discuss the results in section 4 before summarising our findings in section 5.

2 Methods & Data

2.1 The Cut-off matrix in ‘tiered’ Hybrid Life Cycle Assessment

HLCA comes in different ‘flavours’ (Crawford et al., 2018), but so far, to the authors best knowledge, only one method has been used to hybridise full PLCA databases (Yu and Wiedmann, 2018; Agez et al., 2020). This method estimates a process’ missing inputs by using the input structure of the corresponding industry in an Input-Output Model, and is referred to as a tiered hybrid approach. In this article we expand the analysis done in Jakobs et al. (2021) which builds upon such a tiered HLCA model PyLCAi
d introduced by Agez et al. (2020, 2022), which hybridises the process life cycle inventory

---

1 We note that Crawford et al. (2018) defines the method used by Yu and Wiedmann (2018); Agez et al. (2020) as an integrated hybrid approach since it can described mathematically as such. However, we will follow the authors of those studies and refer to it as a tiered approach since the downstream cut-off matrix as described by Suh et al. (2004) is not included.

2 https://github.com/MaximeAgez/pyLCAi

---

3
ecoinvent v3.5 (Steubing et al., 2016; Wernet et al., 2016) and the multi-regional input-output database EXIOBASE 3.6 (Stadler et al., 2018). The tiered hybrid model can be expressed mathematically as:

\[
E = (S^{lca} \quad S^{io}) \left[ I - \begin{pmatrix} A^{lca} & 0 \\ C^u & A^{io} \end{pmatrix} \right]^{-1} \begin{pmatrix} y^{lca} \\ y^{io} \end{pmatrix},
\]

where \( E \) is the vector of total environmental impacts associated with the final demand vector \( (y^{lca} \quad y^{io})^T \), \( S^{lca} \) and \( S^{io} \) are the matrices containing the environmental exchanges per unit of output of the processes and IO sectors respectively. \( I \) is the identity matrix of suitable dimensions, \( A^{lca} \) the LCA technology matrix, \( A^{io} \) the IO technology coefficient matrix, and \( C^u \) is the upstream cut-off matrix linking the LCA processes with input-output sectors. \( 0 \) is a matrix of zeros replacing the downstream cut-off matrix which would be used in a fully integrated hybrid approach (Suh et al., 2004). Finally, \( y^{lca} \) and \( y^{io} \) are the final demand vectors for the output of LCA processes and IO product sectors.

As we can see from eq. (1), the main element of such 'tiered' hybrid approach is the upstream cut-off matrix \( C^u \), which connects the complementing inputs \( A^{io}_{IP} \) from product sector \( I \) into the product sector \( P \) corresponding to process’ \( p \) into the latter’s list of inputs. Following Agez et al. (2022) the elements of the cut-off matrix can be be written down as:

\[
C^u_{IP} = \text{Corr}(p, I, P) A^{io}_{IP} \pi_p,
\]

where \( \text{Corr}(p, I, P) \in \{0, 1\} \) is a double counting correction function, ensuring that every input is only included once, either in the PLCA or coming from the IO model, and \( \pi_p \) is the price of the reference product being produced by process \( p \) used to convert the monetary units in the IO model to the physical PLCA units. In the example given by Agez et al. (2022) this could be an input \( C^u_{IP} [\text{euro/car}] \) of R&D services (\( I \)) in the production of a ‘electric passenger car’ (\( p \)), which is estimated from the use \( A^{io}_{IP} [\text{euro/euro}] \) of R&D services (\( I \)) in the ‘Motor vehicles’ sector (\( P \)) multiplied by the price \( \pi_p [\text{euro/car}] \) of an ‘electric passenger car’. In this representation, however, we only consider one generic process for the production of an ‘electric passenger car’ and one ‘Motor vehicles’ sector, ignoring potential regional differences in the supply chains and associated emissions. Multi-regional input-output databases, such as EXIOBASE, do provide such regional supply chain information.

### 2.2 Geographical resolution in HLCA

Hybrid methods, combining input-output and process data have been subject of active research for more than a quarter of a century (Bullard et al., 1978; Treloar, 1997; Lenzen, 2000; Suh et al., 2004; Lenzen and Crawford, 2009; Crawford et al., 2018, among others). Multi-regional input-output models covering the global economy however, have only been around for a much shorter time period with true Global Multi-Regional Input-Output (GMRIO) models only becoming readily available in the last decade (Tukker and Dietzenbacher, 2013; Peters et al., 2011; Lenzen et al., 2012; Dietzenbacher et al., 2013;
These GMRIO models provide regional specific supply chain information for industry- or product sectors along with the sectors’ region specific impact factors. The advent of these GMRIO models did not only improve the accuracy of the calculation of national carbon footprints (Tukker and Dietzenbacher [2013]), it also allowed for the hybridisation of complete PLCA databases, using regional specific supply chain information (Agez et al. [2020]). However, because of the lack of geographical resolution in PLCA databases, this potential has so far not been fully capitalised upon. Due to the fact that many processes in a life cycle inventory databases such as ecoinvent (Steubing et al. [2016]; Wernet et al. [2016]) have a large regional scope, with a relatively small number of processes being country specific (table 1).

### Table 1: Number of non-market processes in ecoinvent 3.5 divided by regional scope.

<table>
<thead>
<tr>
<th>Group</th>
<th>Geograph. scope</th>
<th>Subset</th>
<th># processes</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>16022</td>
<td>100</td>
</tr>
<tr>
<td>Non-market</td>
<td></td>
<td></td>
<td>10962</td>
<td>68</td>
</tr>
<tr>
<td>Single country</td>
<td></td>
<td>All</td>
<td>5644</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unique reference product</td>
<td>1775</td>
<td>11</td>
</tr>
<tr>
<td>Aggregate region</td>
<td></td>
<td>All</td>
<td>5318</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Global</td>
<td>927</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rest of World</td>
<td>2866</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other</td>
<td>1525</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unique reference product</td>
<td>1862</td>
<td>12</td>
</tr>
</tbody>
</table>

In practice, this means that processes with a regional scope encompassing multiple regions or countries present in the MRIO database, are hybridised with a weighted average of the relevant sector in these different regions or countries. The weights are taken as the relative production volume of each region’s sector with respect to the total production volume of the sector in all regions or countries encompassed in the geographical scope of the process. Expanding on the previous example of the production process of an ‘electric passenger car’, let’s say that this process has the regional scope ‘North America’, which encompasses 3 countries present in EXIOBASE: Canada, the USA and Mexico. This means that the input of R&D services into our cars supply chain will be the weighted average of the input of R&D services into the ‘Motor vehicles’ sectors in Canada, the USA and Mexico, where their weight is determined by their production share in the total production (as monetary value) of the ‘Motor vehicles’ sector output in ‘North America’, see figure 1. Formalised, we can write this as:
Figure 1: Example of the construction of the $C^{ui}$ matrix in the MRIO case (eq. 3 graphically). The input of R&D services into the process ‘electric passenger car production’ with geographical scope ‘North America’ is the production volume weighted average of the input of R&D services into the ‘motor vehicles’ sector in Canada, the USA and Mexico, multiplied by the price of an electric passenger car.

$$C^{ui}_{ir,p} = \Gamma(i, p) \sum_j H(p, j) \sum_{s \in S_p} G(p, s, S_p, j) A^{io}_{ir,js} \pi_p ,$$

where $C^{ui}_{ir,p}$ is the input of the sector $i$ (R&D services) from country $r \in \hat{R}$ (where $\hat{R}$ is the set of countries available in the MRIO model) into the supply chain of the process $p$. $\Gamma(i, p) \in \{0, 1\}$ is the double counting correction factor ensuring that if R&D services were already accounted by the dataset for $p$ it is not accounted again. $H(p, j) \in \{0, 1\}$ is a concordance matching process $p$ to MRIO product sector $j$ and $G(p, s, S_p, j) \in [0, 1]$ a geographical concordance matching the regional scope of process $p$ to the different countries $s \in S_p$ of industry $j$ weighting their output relative to the output of sector $j$ in all countries included in the geographical scope or set of countries $S_p$ of $p$. As such $\sum_{s \in S_p} G(p, s, S_p, j) = 1$. The term $A^{io}_{ir,js}$ is the input from sector $i$ in country $r$ into the sector $j$ in country $s$ and $\pi_p$ still is the price of the reference product of $p$ (‘electric passenger car’).

By averaging the regional supply chain information however, we fail to fully exploit the level of regional detail that the current current GMRIO models offer. Moreover, one loses valuable information on the variance $\text{Var}(X)$ of the environmental footprints as a result of regional differences in supply chain structure and direct environmental impacts of an industry or product sector. In this work we analyse the impact of this regional variance.

---

3We note that throughout this paper, we will use the term variance descriptively as level of the variability rather then the mathematical definition ($\text{Var}(X) = E[(X - E[X])^2]$).
on the uncertainty of the environmental footprints. To model this we alter the recipe for the construction of the upstream cut-off matrix $C^u$ to sample the regions $s \in S_p$, instead of taking the average of the regions $s$ in the regional scope $S_p$ of process $p$ as described in eq[2]. This means that instead of calculating the average hybrid contributions for the whole model once, many realisations of the $C^u$ matrix are calculated in a Monte Carlo simulation to obtain a footprint distribution rather than a point estimate. For one realisation of $C^u$ eq[2] then becomes:

$$C^u_{ir,p}(s) = \Gamma(i, p) \sum_j H(p, j) A^s_{ir,js} \pi_p; \ s \in S_p,$$  \hspace{1cm} (4)

with

$$S_p = \{s_1, ..., s_n\}_p,$$

$$P(s|j, S_p) = \frac{X(j, s)}{\sum_{s \in S_p} X(j, s)}.$$

(5)

Here, $P(s|j, S_p)$ is the probability to pick country $s$ given industry $j$ and the country set $S_p$, and $X(j, s)$ is the monetary production volume of the industry $j$ in country $s$. Using our example of ‘electric passenger car’ production in North America this means that in one iterations of our hybrid model instead of taking the average input of ‘R&D services’ into the ‘Motor vehicles’ sector of Canada, the US, and Mexico, weighted by their relative production volume, we now draw one country from the set \{CA,US,MX\} with a probability determined by the relative production volume of the sector ‘Motor vehicles’ in that country to the production volume of the entire set (North America).

To construct a full $C^u$ matrix for one Monte Carlo iteration, the above process of drawing a country and then calculating the $C^u$ matrix elements is repeated for each process, where the set $S_p$ of course only contains one country $s_p$ for a single country process $p$.

2.3 Price variance in HLCA

Eq. (4) contains a fixed price $\pi_p$ for the reference product of a process $p$. However, as [Jakobs et al. (2021)](#) discuss, prices for commodities vary between and within different supply chains as well as being subject to temporal variation. Moreover, [Jakobs et al. (2021)](#) show, using price data from the BACI trade database ([Gaulier and Zignago, 2010](#)), that a price variance for reference products of processes is strongly linked to the regional scope of the process. With processes representing aggregated regions showing a larger price variation than ‘single country’ processes (see figure 2 of the mentioned study). From this it follows that the price distributions for the same reference products for different countries have significant differences between them.

Refining the approach to regional sampling discussed in section 2.2, we further extend eq. (4) to include a region dependent price distribution, which we model based on BACI trade data. For details on the matching of BACI commodities to the reference products of the LCA processes see [Jakobs et al. (2021), section 2.3](#). Eq (4) now becomes:
Figure 2: The distribution of the variation of reference product prices for ecoinvent processes that have been hybridised, have an aggregated regional scope, and for which BACI price data are available (1120 processes). The variation is given as the robust indicator ‘relative Median Absolute Deviation’ (rMAD). The left distribution shows the price variance distribution in the case price distributions have been build according to the aggregated regional scope of the processes, whereas in the right boxplot the price distribution is considered separately for each country within the aggregated regional scope of the processes (36896 price distributions). The boxes show the quartiles and the horizontal lines the 2.5, 16, 50, 84 and 97.5 quantiles from bottom to top respectively.

\[ C_{ir,p}^u(s, \pi) = \Gamma(i, p) \sum_j H(p, j) A_{ir,j,s}^p \pi(p, s) ; \quad s \in S_p , \tag{6} \]

\[ \pi(p, s) = \{\pi_1, \ldots, \pi_n\}_{p,s} , \]

\[ S_p = \{s_1, \ldots, s_n\}_p , \]

still with

\[ P(s|j, S_p) = \frac{X(j, s)}{\sum_{s \in S_p} X(j, s)} , \tag{7} \]

and now

\[ P(\pi_p|\pi(p, s)) = \frac{v(p, s, r)}{\sum_r v(p, s, r)} . \tag{8} \]

where \( \pi(p, s) \) is the set of prices for which the reference product of \( p \) is exported from country \( s \), and \( P(\pi_p|\pi(p, s)) \) the probability of drawing a price \( \pi_p \), given the physical volume \( v(p, sr) \) for which reference product \( p \) is exported from country \( s \) to country \( r \). Returning to the example of the process of ‘electric passenger car’ manufacturing in North America from before, we first pick a country from the set CA,US,MX as described in eq. 4 and then instead of multiplying this by a fixed price for ‘electric passenger cars’ a price is drawn from the export price distribution of the reference product (‘electric passenger’) car of the drawn country.

The use of region specific price distribution does not only offer a theoretical improvement of the hybrid LCA method, the price variance of the reference products indeed decreases if the price distribution for processes is regionalised. Figure 2 shows the distribution of price variability for the 1120 reference products of hybridised processes with
an aggregated regional scope, for which BACI trade data is available. The price variability is calculated using the robust metric relative Median Absolute Deviation (rMAD) (Gauss, 1816; Hoaglin et al., 2000) of the individual volume weighted price distributions. The rMAD is analogous to the coefficient of variation but is robust against outliers. The left box plot shows the distribution in the case where the price distribution for the entire regional scope is used (e.g. North America from our example). The right box plot shows the distribution of rMADs if for the same processes we consider the ensemble of regional price distributions (e.g. CA, US, and MX separately). The boxes show the quartiles of the distribution and the horizontal bars the 2.5, 16, 50, 84, and 97.5 quantiles (from bottom to top). The median price rMAD value using aggregated regions is 0.32, which decreases to 0.14 when the 1120 regionally aggregated price distributions are split into 36896 regional price distributions. We note that although the price variance per model realisation of drawing a country, drawing a price and calculating the impacts do decrease, this does not necessarily mean that the variance of the environmental impacts decreases as well. Moreover, as it captures potential correlations between commodity prices and environmental impacts of sectors in different regions, this may amplify the variance of the environmental impacts of the process considered.

As the BACI trade database does not provide price data for all reference products, we follow Jakobs et al. (2021) and use a combination of ecoinvent prices and proxy data to estimate price distributions for the reference those commodities and services without a BACI price. See Jakobs et al. (2021, section 2.4 and table 2) for details on this process.

2.4 Data

In this paper we work with a hybridisation of the process inventory database ecoinvent 3.5 (Steubing et al., 2016; Wernet et al., 2016) and the global MRIO database EXIOBASE 3.6 for the year 2012 (Stadler et al., 2018). The price data comes from the BACI trade database (Gaulier and Zignago, 2010). The reference year 2012 was chosen for classification consistency between the different data sources. The hybridisation methodology builds upon the hybrid model introduced by Agez et al. (2020). Finally, we follow Jakobs et al. (2021) and use the PLCA part of the Swiss household consumption model by Froemelt et al. (2018) as a case study to illustrate the effect regional and price variance discussed on a consumption based environmental footprint. The model is based on the Swiss household budget survey (Bundesamt für Statistik, 2013, HBS 2009-2011) corrected to our reference year 2012. Using this model allows for comparison of the analysis in this paper to the analysis of price variance only in Jakobs et al. (2021). For more details on the model as used in this study we refer the reader to Jakobs et al. (2021) and Froemelt et al. (2018).

2.5 Running the Model

Using the method of rebuilding the upstream cut-off matrix by probabilistic sampling of process regions and reference product prices described in sections 2.2 and 2.3 we perform a Monte Carlo simulation calculating the hybrid impact multipliers for all processes.
\[ M_{\text{hybrid}} = C^\text{EXIO} \ S^\text{EXIO} \ L^\text{EXIO} \ C^u, \] (9)

where \( M_{\text{hybrid}} \) is a \( K \times N^{\text{proc}} \) matrix of \( K \) hybrid impact multipliers for \( N^{\text{proc}} \) processes, representing the ‘missing’ impacts per unit of output of the reference product of the processes. \( C^\text{EXIO} \), \( S^\text{EXIO} \), \( L^\text{EXIO} \) are the characterisation matrix, the stressor matrix and the Leontief-inverse matrix for EXIOBASE respectively, and \( C^u \) is the upstream cut-off matrix. The hybrid multiplier matrix \( M_{\text{hybrid}} \) can then be used to calculate the hybrid impact of the supply chain of individual processes, or for a production or consumption basket.

The \( C^u \) matrix is constructed as described in section 2.2, where the double counting correction being applied is the so called STAM (for ‘similar technological attributes method’) introduced in Agez et al. (2019) and implemented in the hybridisation package pylcaio (Agez et al., 2020).

As described in section 2.2, the probability of drawing a particular country (e.g. Canada) for a process (e.g. electr. passenger car production) with an aggregated regional scope (e.g. North America), is proportional to the relative production volume of the (relevant) product sector (e.g. motor vehicles) in that country (Canada). However, the MRIO product sector usually encompasses more products than just the reference product, so it may occur that the reference product (electric passenger car) is not being exported from a particular country (Canada), i.e. that we do not have a BACI record for this reference product (electr. passenger car) in this country (Canada), but the production volume of the IO product sector (motor vehicles) in this country (Canada) is non-zero. In such a case we manually set the production volume of the product sector (motor vehicles) in this country to zero for this process (electr. passenger car) and rescale the relative production volumes of the other countries (Mexico, USA) within the regional scope of the process.

We note that although in our approach prices are sampled from a regional specific distribution, thereby capturing any potential correlations between sectoral impact and commodity price, the regions and prices are sampled independently for each process of the ecoinvent database, meaning that cross-process price correlations are not captured. As discussed in Jakobs et al. (2021), it is highly likely that commodity prices are correlated, however, modelling such correlations is a very complex problem due to the many different and opposing effects at play. They therefore fall outside the scope of this work.

In reality, the geographical location of processes within a given supply chain will also be correlated. E.g. for certain trade relations, physical proximity of the two trading ‘processes’ is required. As these correlations are on a sub-industry level however, they can be modelled for individual supply chains, but in a global model remain too complex to model correctly without additional data. Moreover, we expect the impact of these type of correlations to be small compared to the geographic variance analysed in this study.

### Convergence of Monte Carlo simulations

The Monte Carlo simulations have been run for each scenario with 10,000 runs. Because of the nested sampling in the case of variable geographic scope and regional prices it is hard to predict the number of runs needed to reach convergence of the simulation. We therefore analysed the convergence of the mean and median as well as the coefficient of

10
variation (CoV) and the relative median absolute deviation (rMAD) for both individual processes as well as in the Swiss household consumption case. In the latter case both the mean and the median reach convergence after less than 2000 runs and the CoV and rMAD had both converged after around 4000 runs. For individual processes we analysed the convergence of the rMAD which we use for our analysis, which converged after around 2500 runs.

3 Results

Here we present the results of our analysis of the variability of hybrid footprints due to regional differences in a sector’s environmental impact. To isolate and compare the effect of geographic variability and price variability, we present the results of three modelling ‘scenarios’: (I) where we keep the ‘geography’ of processes fixed but vary the price (as in Jakobs et al. [2021]) (Fix-Var), (II) where we vary the ‘geography’ and use region specific price distributions (Var-Var), and (III) where we vary the country used for hybridisation but keep the price fixed (Var-Fix), using the mean price from the distributions in scenario (I). In section 3.1 we first look at the variance in the process footprint intensity defined as the process’ supply chain impact for one unit of output of the reference product. We then present the results of an illustration case of the same Swiss Household consumption study used in Jakobs et al. [2021] to show the impact of regional variability in supply chain impacts and using regional prices on a consumption basket study in section 3.2.

In the following sections the uncertainty ranges are, unless specified differently, given in Percent Point (pp) where (-x pp, +y pp) shows the 2.5-97.5 percentile range with respect to the median.

3.1 Process level footprints

Figure 3 shows the distributions of the variance of the processes’ ‘hybrid’ Footprint Intensities (FI) i.e. the relative variability in the truncation error correction per unit of output, for the impact categories GWP100 and Land Use. Only hybridised processes that have a (regionalised) price distribution from BACI available have been included here to allow for a consistent comparison between the three different scenarios (Fix-Var, Var-Var and Var-Fix, see above). This process sample is divided in different subsets according to the regional scope of the processes and the product sector they belong to. The regional scope has been divided into ‘All aggregate regions’ and ‘Single Countries’, where the former is further subdivided into ‘Global’, ‘Rest of World’ and ‘Other aggregate regions’. The division into product sectors is based on the ‘central product classification’ (CPC) section to which the processes’ reference products belong. These sectors are as follows: section 0 ‘Agriculture, forestry and fishery products’, section 1 ‘Ores and minerals; electricity, gas and water’, section 2 ‘Food products, beverages and tobacco; textiles, apparel and leather products’, section 3 ‘Other transportable goods, except metal products, machinery and equipment’, and section 4 ‘Metal products, machinery and equipment’. For each subset the distribution of the processes’ relative Median Absolute Deviation (rMAD) is given. The distributions are summarised as the 2.5%, 50%, and 97.5% quantiles with the coloured lines, and the 16%-84% quantile range represented by
Figure 3: The distribution of the variability of process ‘hybrid’ footprint intensities, given as the robust indicator relative Median Absolute Deviation (rMAD) for all hybridised processes with ‘regionalised’ BACI price data. The results are shown for impact categories Global Warming Potential 100 and Land Use. The processes have been divided into subsets of different regional scope and product type (CPC section, see main text for the full names) where the number of processes in each subset is given in the second column. The three different colours show the results of the three different MC runs: fixed geography-variable price, variable geography-variable price, and variable geography-fixed price. The coloured lines show the 2.5%, 50%, and 97.5% quantiles of the distribution, and the black lines show the 16%-68% quantile range.
We find that the variance of footprint intensities stemming from the geographic variance within product sector (Var-Fix, orange) is on the same level as the variance due to price variance (Fix-Var, blue) when we consider all processes in the sample, for both the GWP100 (median rMAD 0.15 vs 0.16 respectively) and the land use impact (median rMAD 0.19 vs 0.17 respectively). Given that the price variance and geographic variance are independent, they amplify one another, which we observe in the significantly higher variances in the Var-Var scenario (green) for both impact categories (median rMAD 0.28, 0.32 all processes for GWP100 and Land Use respectively). Further we find that for global (GLO) and rest of world (RoW) processes the geographic variance seems to lead to a slightly higher impact variance than the price variance (0.23 vs. 0.21 (GLO) and 0.21 vs. 0.16 (RoW) for the median GWP100 rMAD), an effect that is more pronounced for the land use footprint intensity (0.29 vs. 0.22 (GLO) and 0.29 vs. 0.17 (RoW) for the median Land Use rMAD). The fact that the geographic variance between product sectors in different regions leads to a non-zero footprint intensity variance for the subset of ‘single country’ processes is due to the fact that the processes in this group still include processes with an aggregated regional scope in their PLCA supply chains. These processes in turn are impacted by the geographic variance in their upstream (IO) supply chain. For these ‘single country’ processes however, this is of course only a secondary effect, which can be seen in the significantly lower rMAD values for this subset (0.03 and 0.06 respectively for GWP100 and Land Use).

Considering the various product sections we find that the ‘Agriculture, forestry and fishery products’ (CPC section 0) has the lowest median variance of all sectors in all three scenarios with the Var-Var (green) scenario being on the same level as the ‘single country’ subset with (0.14 vs. 0.13 GWP100 and 0.19 vs. 0.17 Land Use respectively). As Jakobs et al. (2021) point out, this might partially be explained by the fact that 142 of 242 processes in CPC section 0 are ‘single country’ processes. Naturally however, the geographical variance (Var-Fix, orange) has a much higher spread, given that this subset does also contain processes with an aggregated regional scope. This in turn of course leads to a higher spread in the Var-Var scenario compared to the ‘single country’ processes. We also see a very wide distribution of variances for processes of CPC section 1 ‘Ores and minerals; electricity, gas and water’ for all scenarios with little difference between the GWP100 and Land Use impacts. Moreover, for both impact categories we see high median variance in the Var-Var scenario (green) with a rMAD value of 0.36 and 84th percentile of around 0.8 and 97.5th percentile greater than 0.9. For this section 88 processes have an aggregate regional scope (12 GLO, 49 RoW, 27 other) and 53 are ‘single country’ processes. For both section 1 and section 2 (‘Food products, beverages and tobaccos; textiles, apparel and leather products’) however, we see a strong increase in variance when considering both geographic- and (regional) price uncertainty.

Overall, the results in figure 3 show how important it is to have regional (single-country) data, have more accurate prices and move regional variability to higher tiers in the supply chain, which in combination drastically reduces the uncertainty in the IO part of a hybrid supply chain.

A different way to look at the impact of regional and price variance and the use of regional price distributions on the footprint intensities of hybridised processes, is to look at the relative increase of the process FI’s when hybridisation is taken into account. Figure
4 shows the median increase in footprint intensity due to hybridisation (truncation error correction) for both the GWP100 and Land Use impact categories. The processes are grouped into the same subsets as in figure 3 and the errorbars show the corresponding median uncertainty (2.5-97.5 quantiles) on the full hybrid FI. Here the ‘increase’ in FI is defined as the median IO impact for all MC runs per process divided by the process’ PLCA impact. As in figure 3 the results are shown for the three different ‘scenarios’. For the GWP100 FI’s we find that the price variance has a higher impact on the FI uncertainties than the regional variance, and, as with the relative variability presented in figure 3, price and geographic uncertainty mostly amplify each other in the fully regionalised scenario (Var-Var), leading to an increased uncertainty range of (-3pp,+18pp), up from (-2pp,+7pp) in the the Fix-Var case (numbers are for the full set of processes: subset ‘All’). We note however, that the full regionalisation of the hybridisation (Var-Var), also leads to an increase of the median truncation error correction of about 10-20% for the full set of processes (All), depending on the modelling scenario used as reference (Var-Fix or Fix-Var). A clear stand out is the subset of CPC section 1 (‘Ores and minerals; electricity, gas and water’) processes, which, in the fully regionalised scenario (Var-Var, green), not only has the highest median GWP100 FI increase of 14%, but features by far the largest median uncertainty range of (-4pp,+49pp), despite not showing particularly large uncertainty ranges in either the Var-Fix or the Fix-Var case.

For the Land Use FI’s we find that median truncation error correction is indeed much higher than for the GWP100, with typical values of around 100% but large differences between the different subsets. The large magnitude of the truncation error correction however, means that although the relative variance decreases for the Land Use FI’s, with a relative uncertainty of (-25%,+158%) with respect to the median compared to (-43%,+280%) for the GWP100 (Var-Var case), the absolute range of possible FI’s is very large indeed, with the median truncation error correction for the full set of processes being 90-143% in the regionalised case (Var-Var). Like with the GWP100 FI’s, CPC section 1 shows the largest increase with 178% and also has the largest uncertainty range in the Var-Var scenario (green) of (-30pp,+256pp) although the difference compared to other subsets is not as large compared to the GWP100 FI’s with the next largest truncation error estimate being CPC section 4 (‘Metal products, machinery and equipment’) with an FI increase of 160% (-33pp,+204pp) in the fully regionalised scenario (green). Unlike the GWP100 FI’s we see that the truncation error estimate for Land Use of both CPC section 0 (‘Agriculture, forestry and fishery products’) and CPC section 2 (‘Food products, beverages and tobaccos; textiles, apparel and leather products’) are very small compared to the other product sections. This is not surprising given the large direct land use impacts of these sectors compared to potential impacts in higher tiers of the supply chain. However, we find that when regional variability and regional price distributions are considered, the truncation error is still non-negligible with 3% (-1pp,+8pp) and 7% (-3pp+25pp) for CPC section 0 and 2 respectively. ‘Single country’ processes also show a small truncation error estimate with a low variance, which might be again be partially explained by a large fraction of the processes in this subset (182 out of 418) coming from either CPC section 0 (142) or CPC section 2 (40).
Figure 4: The hybrid footprint intensities (FI) for both GWP100 (top) and Land Use (bottom), relative to their PLCA counterpart for all hybridised processes with a BACI price available. The processes have been grouped into the subsets according to their geographical scope and product sector (see figure 3 for the sample size of each subset, and the main text for names of the CPC categories). The bars give the median increase in FI due to hybridisation (truncation error) and the error bars show the median uncertainty range (2.5%-97.5% quantiles). Each subset shows the results for the three ‘region-price’ scenarios Fix-Var, Var-Var, and Var-Fix.
Figure 5: The distributions of the Monte Carlo simulations for the average Swiss household consumption, presented as a relative increase with respect to the PLCA only footprint for both the global warming potential (left) and the land use footprint (right). The different colours show the different modelling scenarios described in section 3, and the vertical dashed lines show the median of the respective distribution. The x-axes have been cut to include 99% of the distributions.

3.2 The effect of regional variance on a consumption basket

The results for both the carbon- (GWP100) and Land Use footprint for 1 month of an average Swiss household consumption basket are shown in tables 2 and 3, respectively. The main difference to the results in previous section 3.1 is that here the processes’ output, and with that their impacts, are scaled to satisfy the final demand of Swiss household consumption, therefore no longer relying on the unit final demand as defined in ecoinvent. Results are presented for the different modelling scenarios and for different subsets of processes, to be able to differentiate between the different effects, and access their relevance on the overall result. The subsets are as follows: ‘All’ show all processes providing a non zero output to fulfil the Swiss household consumption demand, all the other groups are subsets of this set of processes. ‘BACI Hybrid’ are those processes that have been hybridised and for which we have a (regional) price distribution from the BACI trade data, and ‘Hybrid’ processes are all processes that have been hybridised. We note that this subset covers the complete ‘hybrid part’ or the ‘truncation error correction’ of the total footprints. The ‘(BACI) hybrid aggregate regions’ subsets consists of the processes that have been hybridised with an aggregate regional, and for the subset of those with a BACI regional price (BACI). The last two subsets are then those processes that have been hybridised and have a single country regional scope with and without a BACI price. Finally, the results are presented for the PLCA part- and the HLCA part of the footprint separately, with the latter one showing the uncertainty results of the Monte Carlo simulations as the 2.5 and 97.5 percentile interval. The percentage column shows the percentage points increase of the PLCA footprint due to hybridisation, e.g. the relative magnitude of the truncation error correction.

Carbon Footprint (GWP 100)

We find that the overall footprint truncation error correction for 1 month of average Swiss household consumption for the GWP100 also increases when regionalisation of supply chains and price distributions is considered (compared to including price variance only) (table 2), the contribution from hybridisation however, remains relatively small though
Table 2: The GWP100 HLCA footprint results for the average 1 month Swiss household consumption. The footprints are presented for the three different simulation ‘scenarios’ for each different subset of the processes in the supply chain. The first column shows the subset, the second column provides the number of processes in the subset, the simulation scenario is provided in the third column, the footprint of the PLCA only part is presented in the fourth column and the sixth and seventh column show the truncation error estimate in absolute units and relative to the PLCA impact in column 5. The uncertainty ranges indicated in the latter two columns indicate the 95% confidence interval of the Monte Carlo simulation results.

<table>
<thead>
<tr>
<th>Subset</th>
<th>N_{proc}</th>
<th>Simulation run</th>
<th>PLCA CO$_2$eq</th>
<th>HLCA$_{Baci}$ CO$_2$eq</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Geography – Price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sim. run</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>12065</td>
<td>Var – Var</td>
<td>1144</td>
<td>96$^{+105}_{-31}$</td>
<td>8.4$^{+9.2}_{-2.7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>1144</td>
<td>83$^{+26}_{-19}$</td>
<td>7.2$^{+2.3}_{-1.6}$</td>
</tr>
<tr>
<td>Var – Var</td>
<td>144</td>
<td></td>
<td>1144</td>
<td>77$^{+71}_{-21}$</td>
<td>6.7$^{+6.2}_{-1.8}$</td>
</tr>
<tr>
<td>Fix – Var</td>
<td></td>
<td>Var – Var</td>
<td>144</td>
<td>58$^{+82}_{-21}$</td>
<td>40.0$^{+57.1}_{-14.6}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>144</td>
<td>41$^{+25}_{-11}$</td>
<td>28.8$^{+7.2}_{-7.0}$</td>
</tr>
<tr>
<td>BACI hybrid</td>
<td>1151</td>
<td>Var – Var</td>
<td>144</td>
<td>39$^{+46}_{-10}$</td>
<td>27.0$^{+32.1}_{-7.0}$</td>
</tr>
<tr>
<td>Hybrid</td>
<td>3794</td>
<td>Var – Var</td>
<td>493</td>
<td>96$^{+105}_{-31}$</td>
<td>19.5$^{+21.3}_{-6.4}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>493</td>
<td>83$^{+26}_{-19}$</td>
<td>16.8$^{+16.5}_{-3.8}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Var</td>
<td>493</td>
<td>77$^{+71}_{-21}$</td>
<td>15.6$^{+14.4}_{-4.3}$</td>
</tr>
<tr>
<td>Fix – Var</td>
<td>114</td>
<td>Var – Var</td>
<td>114</td>
<td>53$^{+82}_{-21}$</td>
<td>46.3$^{+105}_{-18.1}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>114</td>
<td>37$^{+25}_{-11}$</td>
<td>32.2$^{+21.0}_{-9.0}$</td>
</tr>
<tr>
<td>BACI hybrid agg. reg.</td>
<td>813</td>
<td>Var – Var</td>
<td>114</td>
<td>94$^{+36}_{-10}$</td>
<td>30.0$^{+40.2}_{-8.6}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>114</td>
<td>70$^{+26}_{-19}$</td>
<td>40.5$^{+51.0}_{-15.1}$</td>
</tr>
<tr>
<td>Hybrid agg. reg.</td>
<td>1573</td>
<td>Var – Var</td>
<td>206</td>
<td>84$^{+105}_{-31}$</td>
<td>40.5$^{+51.0}_{-15.1}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>206</td>
<td>70$^{+26}_{-19}$</td>
<td>33.9$^{+12.5}_{-9.0}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Var</td>
<td>206</td>
<td>64$^{+71}_{-21}$</td>
<td>31.1$^{+34.6}_{-10.0}$</td>
</tr>
<tr>
<td>Var – Var</td>
<td>287</td>
<td>Var – Var</td>
<td>287</td>
<td>4$^{+2}_{-1}$</td>
<td>15.2$^{+5.6}_{-2.9}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>287</td>
<td>0$^{+0}_{-0}$</td>
<td>15.7$^{+0.0}_{-0.0}$</td>
</tr>
<tr>
<td>BACI hybrid sin. reg.</td>
<td>338</td>
<td>Var – Var</td>
<td>30</td>
<td>4$^{+2}_{-1}$</td>
<td>15.2$^{+5.6}_{-2.9}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>30</td>
<td>0$^{+0}_{-0}$</td>
<td>15.7$^{+0.0}_{-0.0}$</td>
</tr>
<tr>
<td>Hybrid sin. reg.</td>
<td>2221</td>
<td>Var – Var</td>
<td>287</td>
<td>13$^{+3}_{-2}$</td>
<td>4.4$^{+1.2}_{-0.7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>287</td>
<td>13$^{+3}_{-2}$</td>
<td>4.4$^{+1.2}_{-0.7}$</td>
</tr>
</tbody>
</table>
Table 3: The Land Use HLCA footprint results for the average 1 month Swiss household consumption. The footprints are presented for the three different simulation ‘scenarios’ for each different subset of the processes in the supply chain. The first column shows the subset, the second column provides the number of processes in the subset, the simulation scenario is provided in the third column, the footprint of the PLCA only part is presented in the fourth column and the sixth and seventh column show the truncation error estimate in absolute units and relative to the PLCA impact in column 5. The uncertainty ranges indicated in the latter two columns indicate the 95% confidence interval of the Monte Carlo simulation results.

<table>
<thead>
<tr>
<th>Subset</th>
<th>Nproc</th>
<th>Simulation run</th>
<th>PLCA ([\text{m}^2])</th>
<th>HLCA (_{\text{Baci}})</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Geography – Price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>12065</td>
<td>Var – Var</td>
<td>330 (117_{-20}^{+210})</td>
<td>35.5(^{+63.5}_{-13.9})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Fix</td>
<td>330 (106_{-7}^{+102})</td>
<td>32.0(^{+30.9}_{-10.9})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>330 (105_{-6}^{+107})</td>
<td>31.7(^{+32.8}_{-8.9})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Var</td>
<td>166 (67_{-28}^{+184})</td>
<td>40.4(^{+110.8}_{-17.1})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Fix</td>
<td>166 (51_{-19}^{+102})</td>
<td>30.8(^{+61.4}_{-14.1})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>166 (56_{-15}^{+75})</td>
<td>33.8(^{+45.1}_{-9.2})</td>
<td></td>
</tr>
<tr>
<td>Hybrid</td>
<td>3794</td>
<td>Var – Var</td>
<td>242 (117_{-20}^{+210})</td>
<td>48.5(^{+66.8}_{-19.0})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Fix</td>
<td>242 (106_{-7}^{+102})</td>
<td>43.8(^{+32.2}_{-15.0})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>242 (105_{-6}^{+107})</td>
<td>43.4(^{+34.1}_{-12.2})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Var</td>
<td>75 (60_{-28}^{+185})</td>
<td>80.2(^{+245.8}_{-37.7})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BACI hybrid agg. reg.</td>
<td>813</td>
<td>Var – Fix</td>
<td>75 (44_{-19}^{+102})</td>
<td>59.0(^{+135.6}_{-25.1})</td>
</tr>
<tr>
<td></td>
<td>Hybrid agg. reg.</td>
<td>1573</td>
<td>Var – Var</td>
<td>107 (100_{-20}^{+210})</td>
<td>94.0(^{+196.4}_{-42.8})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Fix</td>
<td>107 (89_{-36}^{+102})</td>
<td>83.5(^{+35.5}_{-10.9})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>107 (88_{-15}^{+107})</td>
<td>82.4(^{+99.9}_{-27.4})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Var</td>
<td>91 (7_{-2}^{+2})</td>
<td>7.4(^{+2.3}_{-1.7})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Fix</td>
<td>91 (7_{-0}^{+0})</td>
<td>7.4(^{+0.0}_{-0.0})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>91 (7_{-2}^{+2})</td>
<td>7.4(^{+2.3}_{-1.7})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Var</td>
<td>135 (17_{-3}^{+5})</td>
<td>12.3(^{+3.5}_{-1.9})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Var – Fix</td>
<td>135 (17_{-0}^{+0})</td>
<td>12.3(^{+0.0}_{-0.0})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fix – Var</td>
<td>135 (17_{-3}^{+5})</td>
<td>12.3(^{+3.5}_{-1.9})</td>
<td></td>
</tr>
</tbody>
</table>
not insignificant with 8.4% of the total PLCA estimated footprint, particularly given the skewed uncertainty range of (-2.7pp,+9.2pp) for the 95% confidence interval. Leading to a maximum truncation error of more than 15% in using the conservative STAM double counting correction strategy. Looking at the Fix-Var scenario results, we see a truncation error correction of $6.7_{-1.5}^{+6.2}$% which is smaller than $11_{-1.5}^{+6.2}$% found for the same 'scenario', which is due to the fact that in this analysis we omitted the ‘non-functional-flow-based hybridization’ introduced by Agez et al. (2022) (in order not to introduce more assumptions and restrict the analysis to the two effects studied here). Therefore the total number of hybridised processes in the Swiss household consumption model is smaller than in Jakobs et al. (2021). This smaller set of hybridised processes however, covers roughly 87% of truncation error correction and 90% of the uncertainty found in Jakobs et al. (2021).

Looking at the contributions of the different variances, we find that on the overall carbon footprint (GWP 100), the geographic variance (Var-Fix) scenario plays a smaller role than the price variance (Fix-Var), (-1.6pp,+2.3pp) compared to (-1.8pp,+6.2pp). We note however, that as with the FI in section 3.1, the price- and geographic variances are modelled separately in these two scenario’s. Hence in the fully regionalised scenario (Var-Var) these two sources of variance amplify each other, leading to the uncertainty mentioned above of (-2.7pp,+9.2pp). Moreover, the truncation error correction itself also increases by 17-25% in the Var-Var scenario compared to the scenarios considering only one source of variance.

Comparing the total truncation error correction $96_{-31}^{+105}$ kg CO$_2$eq to the PLCA impact of the subset of processes that is hybridised, we find that the relative truncation error correction is as high as $20_{-6}^{+21}$% in the fully regionalised case (Var-Var). Considering only those processes that have an aggregate regional scope and regional price distributions, (BACI hybrid aggregate regions), we find that although this subset only constitutes 10% of the PLCA footprint at 114 kg CO$_2$eq, the truncation error, at $53_{-21}^{+81}$ kg CO$_2$eq, is $46_{-18}^{+72}$%, which constitutes a relative uncertainty of $(-39\%, +153\%)$. We note that the uncertainty range in the Var-Fix scenario for the (BACI) hybrid processes with an aggregate regional scope, covers the total variance for this scenario of the (BACI) Hybrid group and indeed full sample. Looking at the hybridised single region processes, we see that their hybridisation contributes relatively little to the overall footprint ($\sim$1%) and its uncertainty, although their relative uncertainty, stemming solely from the price variance is still high with $(-15\%, +27\%)$ for all hybridised single region processes.

**Land Use Footprint**

Looking at the Land Use footprints (table 3), we find that at $117_{-46}^{+210}$ m$^2$, the truncation error correction constitutes $36_{-14}^{+64}$% of the total land use footprint of $447_{-19}^{+210}$ m$^2$. Considering again only the hybridised processes, we find that the upstream supply chain impacts stemming from the IO part constitute $40_{-19}^{+87}$%. A reason that an impact category such as Land Use has a larger truncation error than the carbon footprint can be attributed to the fact that most CO$_2$ emissions are associated with energy use, which is used in every step of the supply chain and therefore relatively well covered by a PLCA model. Land Use impacts on the other hand are linked to relatively few processes (e.g. agricultural and mining activities), so if these activities are ‘cut-off’ because they appear further up the
supply chain, this can have a significant impact on the attributed Land Use footprint. One might expect a similar effect in other concentrated impact categories such as water use or the material footprint.

In contrast to the carbon footprint results, we find that the variance of the Land Use footprint stemming from the geographic variability (Var-Fix) is of the same magnitude as the price variance (Fix-Var): \((-11\text{pp}, +31\text{pp})\) compared to \((-9\text{pp}, +32\text{pp})\). Not surprisingly, these variances also amplify when both considered, leading to an overall relative truncation error uncertainty of \((-39\%, +179\%)\). We also observe again that the biggest contribution to the truncation error correction and its variance comes from regionally aggregated processes.

Finally, figure 5 shows the full distributions of the Monte Carlo results for the full Swiss household footprint (subset ‘all’ in tables 2 and 3), as a relative increase with respect to the PLCA only footprint. The different colours represent again the different modelling scenarios, and the dashed vertical lines indicate the median of the respective distributions. The x-axes of the distribution plots have been cut to include 99% of the distribution, as the distributions include some extreme values due to outliers in the price distribution. However, we observe strongly peaked although positively skewed distributions for each modelling scenario. Also visually apparent from these distributions is that for the Land Use footprint, the Var-Fix and Fix-Var scenarios show very similar distributions, whereas for the GWP 100 the Var-Fix distribution is more strongly peaked compared to the Fix-Var distribution. Overall, as also shown in tables 2 and 3, the Land Use footprint distributions are much wider in every scenario.

4 Discussion

Environmental accounting has become a major tool in designing policies to reduce the environmental impact of the economic activities at the (inter-) national, regional, city, or even company or institutional level. And as the reliance on the models needed for the environmental accounting grows, so does the need to understand these models and their uncertainties, both to be able to better interpret the results and to improve upon them. In this analysis, we have proposed an improvement to an existing hybrid LCA method and analysed the uncertainties previously not captured. Here we discuss the implications of our findings and the limitations and further possible improvements of the method.

The results in section 3 show that regionalising supply chains has the potential to significantly alter the outcomes of a hybrid study, for both the process level FI’s (section 3.1) and in the illustration case of the Swiss Household consumption (section 3.2). Even for an impact category that is relatively well covered by PLCA databases such as GWP100, the impact results have a large range of outcome possibilities (uncertainty) once regionised supply chains and prices are considered. Information of this uncertainty is crucial to provide a robust ranking of the contribution of different consumption categories in the total footprint and for discussing the relative importance of individual commodities. Jakobs et al. (2021) discussed how regionalisation of process inventory databases leads to a decrease in price variance of the reference products, which in turn leads to smaller uncertainties in HLCA studies. In this analysis we hybridise each process with an aggregate regional scope as if it were a single country process (Var-Var scenario), and
thereby on average decrease the price variance of reference product as compared to the full price distribution for the aggregated region (see figure 2). However, in the full Monte Carlo analysis, this does not impact the overall price variance for this process, as we still sample the full price distribution, yet merely in different ‘sub-distributions’. Moreover, these ‘sub-distributions’ now get multiplied with different hybrid multipliers (eq. 9), ‘increasing’ the overall variance significantly. We note that this uncertainty is always there, whether one chooses to ignore it by only considering a single-point estimate or to make it explicit by performing a stochastic analysis. The variance coming from the lack of geographical resolution of processes however, can be resolved by regionalising or spatially disaggregating process inventory databases, at least to the resolution of the current generation of global MRIO tables. To accomplish this, additional data sources can be used such as the BACI trade statistics (Gaulier and Zignago 2010), regional production statistics from Eurostat, the Food and Agricultural Organization of the United Nations, or electricity mix data from the International Energy Agency. Moreover, as Patouillard et al. (2018) point out in their review and practical recommendations paper on the spatial dimension of life cycle assessment, focus should be placed on processes contributing mostly to the overall uncertainty of the relevant footprints of a given reference flow. To this end, the methodology capturing HLCA specific uncertainty introduced in this paper could be implemented in LCA software such as Brightway2 (Mutel 2017) such that global sensitivity analysis methods as developed by Kim et al. (2022) or Cucurachi et al. (2021) could be applied to aid targeted spatial disaggregation efforts.

**Double counting correction**

One of the main challenges of tiered (or integrated) hybrid LCA is the avoidance of accounting flows of goods and services, and with that the associated environmental flows, more than once. In practice this is done through a double counting correction factor that nullifies inputs from the MRIO table if they already have been accounted for in the process inventory (see eq. 2). For the hybridisation of large systems or entire databases such as done here, this correction has to rely on a set of heuristics which of course will not be accurate for all individual processes in the database. In this analysis we employ the so-called STAM double counting correction (Agez et al. 2019), which is a rather conservative approach erring on the side of underestimating the truncation impacts rather than overestimating them. Jakobs et al. (2021) show that in case of less stringent double counting correction schemes the general impact of hybridisation increases and naturally, with it the variance. This means that the impacts and variances found in this study are also rather conservative estimates, given the data and assumptions.

**Capitall inclusion**

PLCA and EE-IOA traditionally deal differently with the attribution impacts from capital creation. In PLCA the impact of capital investment is propagated to the end product or service (e.g. the impact of any product will include a tiny piece of the impacts needed to produce the power plant generating the electricity needed for its production), whereas in EE-IOA, capital investments are typically accounted separately in the year in which they occur. Capital investments however are a well know driver of resource use and en-
environmental impacts and as such the endogenisation of capital investments has been an active topic of research recently (Södersten et al., 2018a,b; Font Vivanco, 2020). Agez et al. (2022) included the endogenisation of capital investment into the hybrid LCA, and showed that the capital endogenisation in the MRIO part of the hybrid model, increases the additional impact due to hybridisation about 50% for the GWP 100 and Land Occupation IW+ impact factors (Agez et al., 2022, figure 2). As the endogenisation of capital in MRIO models comes with its own set of assumptions and limitations, we have chosen for this study not to include the endogenisation of capital investment in the MRIO model. This allows us to focus on the uncertainty stemming from regionalisation and price variance. We note however, that the methodology introduced in this analysis can be readily adapted to include the capital add-on matrix approach presented in Agez et al. (2022) in the fully regionalised approach.

Illustration case of Swiss household consumption

In section 3.2 we analysed the effect of geographical variance on the footprint of an illustration case of the average monthly Swiss household consumption. The ecoinvent database contains an unrepresentative large number of Swiss process inventories as compared to specific inventories for other countries. Therefore, we investigated the appropriateness of the choice of Swiss household consumption to illustrate the effect of regional variance on the footprint results. To do so we created a counter factual scenario by replacing any final demand of a given product from a Swiss process, to the equivalent process in Germany when available or the equivalent ‘Europe’ or ‘RoW’ processes. We found that although the number of Swiss processes in the total supply vector did not change much (281 in the real case of Swiss household consumption and 277 in our counter factual scenario), by design, they occur at different stages in the supply chain. We found that changing to these counter factual scenario’s has little impact on the truncation error and its uncertainty for both footprints in the Var-Fix and Fix-Var scenarios (table 4). In the Var-Var scenario it leads to slightly increased truncation error estimates of 9.9% for GWP100 (up from 8.4%) and 43.7% for Land Use (up from 35.5%), however, the uncertainties of truncation error estimate increase strongly to (-4.5pp,+21.2pp) up from (-2.7pp,+9.2pp) for GWP100 and (-21pp,+88pp) up from (-14pp,+64pp) for Land Use. This shows that the results of HLCA studies focussing on product systems or consumption baskets in other regions than Switzerland will likely show even higher variances due to regional price and process variability. Of course the exact magnitude of the variability also depends strongly on the product system or consumption basket under study.

Hybridisation uncertainties in relation to PLCA footprint uncertainties

In this analysis we focused on the uncertainties in a database-wide tiered HLCA model, and presented the uncertainty in relation to the PLCA impacts. In doing so we treated the PLCA footprint as a point estimate, whereas in reality the PLCA impacts have their own uncertainties. To put the uncertainties in the hybrid model into perspective against the uncertainties in the PLCA part only we used Brightway2 LCA software framework (Mutel, 2017) to perform a stochastic calculation of the PLCA impacts for the Swiss household
Table 4: The truncation error correction with uncertainty in the case of a counter factual scenario where all products and services of the Swiss household consumption final demand vector coming from Swiss processes have been replaced by the equivalent product or service from German, European, or RoW processes. The last two columns give the factual scenario of Switzerland for comparison. The results are shown for the different variance scenarios and for the overall GWP100- and Land Use footprints.

<table>
<thead>
<tr>
<th>Impact</th>
<th>Simulation run</th>
<th>N\textsubscript{proc}</th>
<th>HLCA %</th>
<th>N\textsubscript{proc}</th>
<th>HLCA %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Geography – Price</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var – Var</td>
<td>12124</td>
<td>9.9\textpm21.2</td>
<td>8.4\textpm4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GWP100</td>
<td>Var – Fix</td>
<td>6.0\textpm1.9</td>
<td>12065</td>
<td>7.2\textpm2.3</td>
<td></td>
</tr>
<tr>
<td>Fix – Var</td>
<td>5.5\textpm5.6</td>
<td></td>
<td>6.7\textpm1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Var – Var</td>
<td>43.7\textpm21.4</td>
<td>35.5\textpm33.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land Use</td>
<td>Var – Fix</td>
<td>29.8\textpm29.3</td>
<td>12065</td>
<td>32.0\textpm30.9</td>
<td></td>
</tr>
<tr>
<td>Fix – Var</td>
<td>29.5\textpm8.7</td>
<td></td>
<td>31.7\textpm32.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The truncation error correction with uncertainty in the case of a counter factual consumption case, which considers all flows, both technosphere- and biosphere flows, as random variables, with uncertainty distributions based on the pedigree uncertainty information in the ecoinvent database (Muller et al., 2016). The results for the GWP 100 and Land Use impacts for the PLCA part of the model of Swiss household consumption are 1235\textpm131 kg CO\textsubscript{2}eq and 368\textpm58 m\textsuperscript{2} respectively, which present the median and 95% confidence interval. The first thing to note here is that when doing a full stochastic analysis, the median (as well as the mean) values are substantially higher than their static counterparts presented in section 3.2. This discrepancy arises from the implementation of the pedigree approach, where many exchanges (flows) are modelled with a lognormal distribution where the static exchange values are set to be the median of the lognormal distribution. As in a uncertainty analysis, it is the mean that gets propagated, which in the case of a lognormal is larger than the median, the stochastic result therefore is higher than the static result. For our point of discussion however, we are interested in the relative uncertainties how they compare to the uncertainties arising from geographic and price variance in the hybrid model. The relative (2.5%, 97.5%) uncertainty for the GWP 100 of the PLCA part is (-9pp,+11pp), comparing this to the (-3pp,+9pp) uncertainty of the fully regionalised hybrid approach (Var-Var), we see that whereas the PLCA uncertainty, which considers uncertainties in all exchanges, is relatively symmetric, the geographic and price variance are strongly positively skewed, with the upper bound error being of similar magnitude as the PLCA uncertainty. For the Land Use footprint, the relative uncertainties for the PLCA and hybrid (Var-Var) parts are (-16pp,+37pp) and (-14pp,+64pp) respectively. Here we see that both distributions are positively skewed, with a similar lower bound uncertainty, but that the 97.5 percentile is about a factor of two higher than the PLCA part. Of course we need to keep in mind that we are not comparing the same uncertainties here, with the PLCA uncertainties covering the exchange uncertainties, and the variance in analysed in this paper only dealing with model uncertainties and excluding flow uncertainties within the IO model. However, the simple
comparison of the magnitude goes to show that these model variances are significant for the complete uncertainty budget, even for the GWP 100 impact category, for which the truncation error is relatively small.

5 Conclusion and outlook

Although the uptake of HLCA is being made easier with the advent of readily available HLCA databases (Agez et al., 2022; Crawford et al., 2021), the acceptance of HLCA as a main stream tool for environmental assessment is still limited. As with any (new) method though, the credibility of HLCA methods hinges on the thorough assessment of the uncertainties of the model. In this paper we analyse the impact of one such source of uncertainty, the regional variance in supply chains, on HLCA results for GWP100 and Land Use impact. Moreover, we present the current best practice to deal with the disparity in regional resolution between PLCA- and global MRIO databases, maximising the information content of the resulting footprints by using all information available in the data. In practice this means that instead of averaging regional upstream supply chain information in MRIO models in case of low regional resolution processes in PLCA databases, we propose to employ a stochastic analysis instead where the various regions in the upstream supply chain are sampled. This simultaneously allows to use regional price distributions, thereby increasing the accuracy of the model.

We find that regional variability in upstream supply chains has non-negligible effect on the footprint intensity of processes with an aggregated regional scope. The magnitude of the resulting footprint uncertainty however, depends on the product sector of the process and the environmental impact considered. With the difference between product sectors being more pronounced for the Land Use impact than for the GWP100, highlighting the importance of hybrid product system models for determining product-specific and complete land use footprints. We also find that although the regional price variability is on average lower than the inter-regional price variance for a given commodity, using regional price distributions while considering regional supply chain variance, strongly enlarges the overall uncertainty of the footprint results.

As with the uncertainty due to price variance, the positive skew of the footprint uncertainties due to regional process/sector variance, highlight the importance of regionalisation of LCA, as it means that the lower limit of the truncation error is close to the median truncation error estimate, however the upper limit can be up to multiple times the median estimate.

In concrete applications and for given reference flows, practitioners need to identify those processes that contribute particularly much to uncertainty due to their regional variance and regional price differences. Regionalisation of those processes then improves the precision of the footprint estimates, and this procedure can be facilitated by implementation this hybrid LCA framework, including uncertainty, into LCA software such as Brightway2.

This work builds on the hybridisation methodology and open source software package developed by Agez et al. (2020, 2022). To further support such cumulative research, the software developed for this analysis is available as an open source package on the GitHub repo [https://github.com/jakobsarthur/Geographic_Variance].
Acronyms

EE-IOA  Environmentally-Extended Input-Output Analysis
FI  Footprint Intensities
GMRIIO  Global Multi-Regional Input-Output
GWP100  Global Warming Potential 100
HLCA  Hybrid Life Cycle Assessment
LCI  Life Cycle Inventory
MRIO  Multi-Regional Input-Output
PLCA  Process-based Life Cycle Assessment
pp  Percent Point
rMAD  relative Median Absolute Deviation

Acknowledgements

The authors would like to thank Chris Mutel for his advice on the computational handling of the data in the model and Carsten Dorman for his valuable input and discussion regarding the statistical analysis. Finally the authors thank Thomas Wiedmann for his constructive comments.

Funding

AJ was supported by Swiss National Science Foundation grant number 407340.172445 as part of the project "Open Assessment of Swiss Economy and Society" which in turn was part of the National Research Program "Sustainable Economy: resource-friendly, future-oriented, innovative" (NRP 73) (http://www.nrp73.ch). SS receives funding from the Eva Mayr-Stihl foundation.

Supporting Information

The following supporting information can be found online in the supporting information section:

S1  Data behind figures 3, 4, and 5.
Data and code availability

The python code used in this analysis, is available in the GitHub repository [https://github.com/jakobsarthur/Geographic_Variance_HLCA](https://github.com/jakobsarthur/Geographic_Variance_HLCA). The data behind the figures is available in the SI. The hdf5 files containing the full results of the Monte Carlo simulations will be shared upon request.

Conflict of Interest Statements

The authors declare that they have no conflict of interest.

Authors’ contributions

All authors contributed to the research, analysis and manuscript and read and approved the final manuscript.

References


Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial1Figuredata.xlsx