Open Risk White Paper

Sustainable Public Procurement

Part II: Making Use Of Environmentally Extended Input-Output Data

Authors: Philippos Papadopoulos

September 29, 2025

www.openriskmanagement.com

The open future of risk management

Summary

Adopting Open Data and Open Source tools can be a catalyst in the sustainability transition and Green Public Procurement (GPP) is a particularly opportune domain to explore this potential. In the first paper in this series we outlined key information flows and tasks that are relevant in this context. Here we dive deeper into how one can leverage a particularly promising source of economic and environmental impact information: Environmentally Extended Input-Output (EEIO) databases. We will discuss some of the challenges that must be addressed to effectively use EEIO tools in GPP, especially in European context. The use cases in focus are: how to produce an overall inventory of direct and indirect emissions for different procurement categories, and how to differentiate between green and non-green products. Some practical problems that must be tackled are: the consistent linkage of procurement data sets to EEIO databases, and the disaggregation of EEIO sectors to more granular of green / non-green products. We discuss potential approaches and their pros and cons. We analyze the specific challenge of creating a CPV-NACE mapping that would link demand driven procurement product taxonomies with supply driven economic activity classifications. We outline a methodology for disaggregating EEIO business sectors to constituent product categories with differentiated environmental impact while preserving the accounting constraints satisfied by the aggregate databases.

Further Resources

- Equinox and Leontief are open source projects that facilitate sustainable portfolio management and EEIO calculations respectively. Sustainability.town is an online demo website illustrating relevant workflows.
- The Open Risk Manual is an open online repository of information for risk management developed and maintained by Open Risk.
- The Open Risk Academy offers a range of online courses around risk management and sustainable portfolio management, utilizing the latest in interactive eLearning tools.
- More content in our Open Risk White Papers and the Open Risk Blog.

About Open Risk

Open Risk is an independent provider of training and risk analysis tools to the broader financial services community. Our mission is captured by the motto: The open future of risk management. Learn more about our mission and projects at: `www.openriskmanagement.com`

1 On the use of EEIO data in Green Public Procurement

"To tell a new story let's start with a new picture of the whole economy." - Kate Raworth, Doughnut Economics.

Green Public Procurement (GPP) represents a major opportunity [1] to catalyze the transition of economies towards more sustainable patterns. It is both a major component of all economies and is less subject to the constraints of private sector procurement decisions. Yet GPP faces various challenges, including the availability, quality and cost of factual environmental impact information and tools on which procurement authorities can base decisions [2]. In the first paper [3] of a series dedicated to *Sustainable Public Procurement* we reviewed mechanisms that can help address pain points in data flows through the use of *open standards*, *open data* and *open source* tools [4]. We framed the overall task of GPP as an instance of *Sustainable Portfolio Management* [5], defined as: tools and practices for the effective management of *portfolios of contracts* with due consideration of sustainability objectives. We outlined the information processing pillars that are relevant, and formulated the task of *attributing* greenhouse gas emissions to an existing procurement portfolio.

In the second paper of this series we dive deeper into the potential roles of one particularly relevant source of economic and environmental impact information: statistical economic databases termed **Environmentally Extended Input-Output** frameworks (frequently abbreviated as EEIO). Such tools carry significant promise towards facilitating sustainable portfolio management and Green Public Procurement more specifically and they are already being used in sustainable finance domains. A few pioneering studies of use in public procurement have already been produced 1.3. We will discuss here some of the challenges towards expanding the use EEIO tools in GPP.

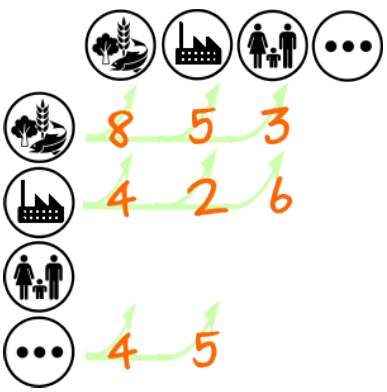


Figure 1: EEIO tables and the associated economic models track exchanges between the various actors comprising the economy. They have become a powerful tool supporting environmental and economic policies in various contexts. They provide unique insights into the economic relations between different production and consumption sectors (optionally also countries) by tracking monetary exchanges on a statistical basis (via surveys). Monetary based EEIO tables can be further augmented with physical (material and energy flows) that associate the environmental and social impact of production and consumption, per country and sector/product category. These impacts can be GHG emissions, natural resource usage (such as water or land) etc. revealing unique views into the structural factors affecting sustainability. Nevertheless, there are important issues to address before such approaches can be fully embedded in decision making.

The two main use cases that are our focus here are: how to produce a reliable overall emissions *attribution* schedule for a procurement portfolio and how to *differentiate* between green and non-green products. The development and embedding of concrete applications in production environments is by no means trivial and there are several concrete issues that must be addressed. The objective in this paper is to discuss some immediate challenges: i) how one can confidently connect contractual procurement data to EEIO databases (the linkage or mapping between procurement product classifications and economic activity classification systems) and ii) how to tackle the (currently) limited granularity of the EEIO paradigm. Our context is the European GPP landscape and the corresponding EEIO databases. Similar considerations apply to many other regions.

1.1 Linking GPP portfolios to environmental impact metrics

The meaning of environmental impact from procurement activities is intuitively obvious but precise definitions that are both practical to implement at scale and generate the right incentives is far from trivial. Modern economies are built on complex networks of interdependence between economic actors, extended supply chains that cross country boundaries, reliance on diverse infrastructures (transport/logistics and energy grids being particularly relevant) and, not least, a vast multitude of products and services. The quantification and attribution of environmental impact needs transparency of methodology and adequate data quality if it is to support decision-making. Weaknesses such as over or under-estimation of impact, biases and blind spots may undermine both perception and the effectiveness of GPP.

The journey of information linking real world impact to procurement portfolios is long and context dependent. Comprehensive methodologies towards GHG emissions inventories have been provided by organizations such as the IPCC [6], which help compile the *direct emissions from primary sources.* The sources are classified according to the Common Reporting Format (CRF). For GHG emissions tracking, the bulk of the statistical work is done at national level with submissions to UNFCCC. All signatory nations report greenhouse gas (GHG) inventories under the United Nations Framework Convention on Climate Change (UNFCCC). They record emissions and sinks of greenhouse gases by categories. A sketch of the European situation is given in Figure 2.

Each inventory source category (CRF/NFR source code) is assigned to a emitting production activity (i.e. NACE grouping) and/or private household consumption activity. This mapping of economic production activities to business sectors (NACE) enables the linking of environmental impact data to other facets of the economy. This is done by statistical agencies. The methodology for these maps may differ by country and may not easily accessible. The European Environment Agency (EEA) and EuroStat coordinate the collection of national GHG inventories and eventually publish them mapped to economic sectors as classified by NACE. This mapping is presently *coarse*, e.g., the latest EEA/EuroStat Air Emissions Accounts provides only 65 distinct estimates (out of the ten times larger total number of sector/product classes). This complex information flow underlies the construction of environmental extensions to IO databases and thus forms the readily available information tools for adapting EEIO for the purposes of GPP.

In the European context, the CPV classification services public procurement communities by covering a large number of products, but, as we will see, many of these extra codes are not used and an official mapping to NACE is missing. As there is no official statistical mapping specifically for public procurement activities, solutions must be based on indirect mapping via NACE, an approach which we will discuss in 2.

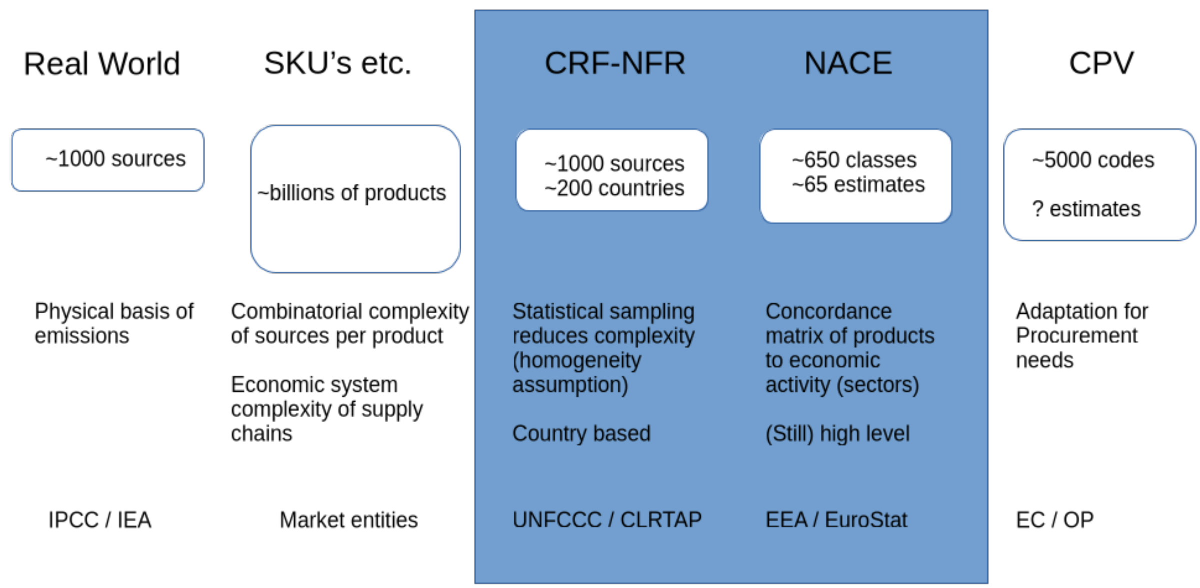


Figure 2: Schematic representation of the information landscape and associations that are implicitly needed for the estimation of direct GHG emissions. From left to right we have the information processing journey that is required to keep track of diverse environmental impacts. The top row indicates what type of labeling or statistical classification is involved. The last row indicates which entity is primarily responsible for the scheme. The fundamental environmental impact intensity parameters needed to characterize economic production processes are not dramatically many. In other words, the distinct number types of *emissions generating processes* is finite, of the order of a thousand. Yet the number of possible products and the different production possibilities exhibit explosive *combinatorial complexity* that exceeds billions of distinct components, products and production contexts. This complexity can be tamed by aggregation and *statistical sampling*, which is the key added-value of EEIO databases. Yet the granularity of environmental impact estimates, while growing, is currently rather limited.

While the above flow serves as the basis for establishing measured environmental impact, it does not in itself address all information requirements in any concrete sustainable portfolio management context. A key piece of additional information is the so-called *inventory boundary.* It has been recognized early on that significant environmental impacts from products may be occurring anywhere in their upstream or downstream supply chain. Most notably the **GHG Protocol** initiative addresses such *indirect emissions from supply chains* [7, 8, 9, 10]. While there is currently no version of the GHG Protocol that explicitly covers the GPP context, the core tenets of such an approach are reasonably easy to articulate: The central economic entity in the case of GPP is a public authority pursuing procurement of goods, works and services in the open market (central and local government procurement activities would be the typical instances). Impact metrics can be grouped according to the three main *Scopes*[1].

- **Scope 1** emissions comprise emissions from economic activities that are undertaken by the public authority itself using procured goods. As an example this would be from the use of fossil fuels in internal combustion engine (ICE) powered vehicles serving the public authority's needs. The extent of such activities might be inferrable from procurement contract information, but their environmental impact is not causally linked to the procurement *spending amount* (in this case, the cost of the vehicles). The impact is rather derived from the subsequent *use* of products or services (e.g., the number of kilometers traveled per year, the number of heating hours etc.) including if appropriate final disposal or recycling emissions. The economic activity indicators for own use must be sourced separately and can subsequently be mapped to EEIO sectors/products under the *producer responsibility approach* [11] which multiplies direct sectoral emission intensities with usage indicators.

- **Scope 2** emissions are a special case of direct emissions, with a focus on the consumption of *electricity*. This subcategory is singled out in the GHG Protocol because of a complicating factor: the environmental impact from electricity use in not uniquely defined by the underlying product technology but depends on the *energy mix* of the applicable grid(s) used, in this instance by the procurement authority. Continuing with our example, if public procurement was for electric rather than ICE vehicles, the direct emissions of using these vehicles would link to the energy mix of the grid available. To capture this effect in EEIO context one requires *sufficient resolution of the energy sector*, in particular an accurate representation of the energy mix of the electricity grid being used.

- **Upstream Scope 3** emissions comprise the impact from activities that will be performed by contractors to fulfill the procurement contract. In turn those can be split into the *direct emissions* of the contractors (e.g., the energy use for the manufacture of parts or the assembly and delivery of vehicles) and the *embodied emissions* in the goods or services that were required to deliver the above, thus including energy costs of the supply chain of the manufacturer. In EEIO context contractor and further upstream supply chain emissions are captured naturally via standard methodologies [12]. In this case the spending amount under the procurement contract links naturally to the monetary exchanges recorded in the EEIO database (keeping in mind the annualized nature of the latter). Physical quantities, to the extent they can easily be extracted from documentation could also provide complementary quantification methods as large EEIO databases track material flows as well.

The relative weight and importance of these three different scopes varies by product and service and must be evaluated for every category.

1.2 The granularity challenge and means to tackle it

A typical question for GPP implementers would be: For a given procurement proposal (technical specifications, award criteria etc.), what will be the environmental impact? While entirely reasonable, this requirement highlights the *mismatch* between the potentially very detailed and complex descriptions of products versus the small number of distinct environmental impact estimates (currently typically less than a hundred).

One approach to address the mismatch is the so-called *process-based Life Cycle Assessment*, sometimes also called the bottom up approach to LCA. It works by pursuing a detailed analysis which links technical specifications, labels or other requirements to fundamental emission intensities via a *process inventory* of energy and material requirements, which are then mapped to official data and methodologies from IEA, IPCC, etc., via the ISO 14040 (2006) and ISO 14044 (2006) standards. The primary difference (and a conceptual weakness of process-based LCA) from EEIO is the ambiguity of *system boundaries*. This is termed the *truncation error* and it denotes the proportion of impact (attributed value) that is not covered by the system boundaries of the LCA. Truncation errors can occur when, for whatever reason, e.g., cost, complexity, data quality, errors etc. some part of the environmental impact is ignored.

The complementary (EEIO-LCA) approach to constructing a life-cycle analysis avoids the truncation error by defining inventory boundaries from a macroscopic perspective, essentially by grouping low level processes into discrete production sectors (e.g. NACE). In contrast to the process-based LCA approach this can be seen as a top-down approach to inventory modeling, where top-down refers to how contributions are grouped and sampled, while the fundamental process intensities are obviously still the same. A key strength of EEIO is that it offers 100% completeness, i.e., no truncation error by construction. At least in theory, no emission pathway is at risk of being missed. Another important aspect

[1]Downstream Scope 3 emissions are less generally applicable in the GPP context, but they play an important role in corporate context

of EEIO is that, while not a trivial procedure, it can be made consistent with the impact measures as reported by national accounts [13]. This is relevant for linking the effectiveness of GPP to national targets. Another benefit is that other types of environmental impact (and even social considerations) are easier to integrate in EEIO-LCA, with minimal alterations of the overall design. This is achievable with the machinery of so-called *satellite accounts* that can associate various environmental and social metrics to the economic input-output network.[2]. Last but not least, the official status of statistical agency publications, the transparency, accessibility and usage in different policy contexts helps reinforce EEIO frameworks as reliable sources of impact information.

Yet while in principle EEIO databases can be made very granular, current EEIO databases suffer from relatively limited granularity. The resolution of many products and services is low and does not reflect the heterogeneous nature of many sectors. As mentioned, EEIO databases solve the problem of cost and exponential combinatorial complexity effectively by making use of *statistical sampling*. The statistical approach is made possible by the assumption of *homogeneity* within pools of essentially similar economic exchanges. This results in potential *aggregation error* if one tries to use the average estimate instead of the specific product. Other potential issues are the *timeliness* of data, data *confidentiality* affecting some sectors etc.

All-in-all, EEIO approaches deliver good *screening-level tools* that provide an overall schedule of environmental impact, but need additional steps before being used for more granular decision making. One strand of thinking to address the granularity challenge has been the use of process-LCA as a base method and augment with EEIO data where required, so that one can utilize the benefits of EEIO while minimizing its limited granularity. This is the so-called *hybrid-LCA*. Process-LCA data are used as a starting point and combined with EEIO-LCA to capture missing flows [14]. The even larger complexity of such an approach, combined with the still evolving data quality landscape and potentially important methodological differences warrant caution: In a comparison reported in [15], while they find that there is good agreement between process-LCA and EEIO-LCA for certain sectors, more than half of the matched products differ by more than a factor 2.

Historically an alternative strand of thinking has been to proceed in the opposite direction, starting with the macro database and enhancing with more granular information where needed. This is the so-called *IO disaggregation procedure*. The IO literature since its earliest times devoted substantial attention in this direction. The earliest proposed method dates back to [16], who first framed the disaggregation problem. Given the obvious practical utility, a massive literature followed since, with an important milestone being the work of [17][3]. A more recent contribution by [18] is relevant for our purposes, as they bring into the discussion EEIO use cases and consider IO balance constraints that were ignored in earlier literature. In Section 3 we expand on this approach and we illustrate the computational workflow with an example in 4.

LCA and the art of the feasible

- Pure top-down approaches (EEIO) offer a solid foundation for high-level impact analyses but have difficulty addressing detailed questions at the discrete product level.
- Pure bottom-up approaches (process-LCA) offer (potentially) the requisite detail, but are hampered by high cost and overall completeness questions (truncation error).
- Hybrid-LCA are bottom-up approaches augmented with EEIO. They aim to address the truncation error, but at the cost of more methodological complexity.
- Top-down disaggregation approaches can expand (to some extent) the granularity of EEIO by injecting a small number of additional information.

♡

1.3 Case Studies of using EEIO in GPP analysis

In recent years a number of analyses have demonstrated the use of EEIO in GPP context. We briefly enumerate them here.

[2] One caveat is that all such calculations are adopting a stylized *linear assumption*, where impact is always assumed proportional to level of economic activity

[3] These early studies were informed by computational constraints of earlier era's

1.3.1 Government of Canada

The objective of a recent study [19] was to identify the most emission-intensive procurement categories and develop targeted interventions. The work estimated the embodied carbon footprint of the goods and services procured by Canada's central procurement organizations (the Public Service and Procurement Canada and Shared Services Canada) over the period 2016 – 2020. Embodied emissions have been calculated using **openIO-Canada**, an open source model for environmentally extended input-output analysis, which uses spending to estimate the embodied carbon in goods and services procured. In the analysis, emissions related to the *use* and *end of life* phase (i.e., Scope 1/2 emissions) were not counted. Namely they are considered as already measured as part of other government activities and operations (e.g. the operation of facilities and fleets). The estimates were based on average emissions given the amount spent in each procurement category. The approach did not measure the effect of green public procurement choices versus conventional choices.

1.3.2 Government of the Netherlands

A similar study in the Netherlands[20] aimed to gain insights into the environmental impact of the Dutch central government procurement. The two LCA methods discussed above have been combined in part: EEIO based input-output analysis and process-based life cycle analysis. The EEIO approach provided the complete picture of the impact (based on market averages per product group) and made use of the EXIOBASE database. To address the distinction between sustainable and non-sustainable products or services, targeted product-specific life cycle analyses were subsequently used. This was done for two product groups, energy carriers and construction.

1.3.3 A Europe-Wide POC study by Open Risk

In the context of proof-of-concept project to assess the environmental footprint of the *entire* European public procurement activity (at least above a certain spending thresholds), we undertook a study using the public procurement data of the TED database. The approach and methodology were documented in a series of posts:

- Part 1 - Overview of the Public Procurement TED dataset
- Part 2 - Identification of Entities involved in procurement
- Part 3 - Attribution of GHG Emissions using the CPV classification
- Part 4 - Green Public Procurement as Sustainable Portfolio Management

Ultimately data quality issues with the public spend data available on TED prevented from drawing firm conclusions. Nevertheless a broad picture can be obtained after certain data cleaning and data imputation steps.

2 The CPV classification and the NACE mapping

We discuss now in some detail the **Common Procurement Vocabulary**, or CPV classification, as it plays a crucial in public procurement in European context. The aim is to sketch the challenges and opportunities in the context of linking procurement data to EEIO databases. The CPV is a classification system for public procurement aimed at standardizing product references used by authorities to describe the subject of procurement contracts (what is being asked for from the market). The main objective of introducing the CPV was to enhance transparency, efficiency and competition. The idea is to provide a common basis for formulating procurement needs and making it easier to identify opportunities for suppliers. Assigning a CPV code aims to surface opportunities which might otherwise be hidden behind linguistic barriers and/or definitional vagueness. While the precise procurement needs are spelled out in textual form inside technical specifications and other documentation, the CPV code provides the headline perspective. Ceteris paribus, this reduction of discovery barriers would also apply in the context of Green Public Procurement, but as the CPV was not designed with GPP in mind this is at present not possible.

The CPV classification has an important *statistical function* and this is the role of most interest for us here. Enhanced discovery of tenders can in principle be served by other knowledge management tools, such as *keywords, tags* and *search engines* [4], the statistical understanding of procurement activity can be served better with the appropriate utilization of CPV codes. As all contracts need to have CPVs assigned, this allows statistical data collection on the *who, what, when*

[4] analysis of search patterns shows that bidders actually do use *keyword search* of tender documents more often than searching by CPV codes!

and how of public procurement activities. The linkage and use of EEIO databases in GPP context relies precisely on this important statistical role of the CPV.

The CPV is by far not the only classification used in procurement. The CEN Workshop Agreement (CWA) [21] contains an analysis of the main classification systems and catalogs used in Europe in both private and public sectors, covering similarities and differences between four important classifications. The four systems analyzed were CPV, eCl@ss, GPC and, UNSPSC. The CPV was found to offer fairly comprehensive coverage. An interesting aspect of the eCl@ss classification (that is missing in CPV) is the use of *standardized attributes for class descriptions.* This suggests that a semantically richer model than what is offered by the current CPV, one that would integrate information about underlying processes and properties may enable in the future far more accurate matching to environmental impact measurements. But for now Europe-wide GPP applications must content with the current CPV system.

2.1 Brief History of CPV Development and Legal Status

The CPV classifies requirements as *goods* (supplies), *works* (which means principally construction works) or *services.* It was originally based on the nomenclature Classification of Products by Activity (CPA). The CPA is a *six-digit* code system to provide a product classification for Europe. It was chosen as a base for CPV because it was consistent across Europe and adapted to its industrial structure. In turn the CPA was built on the basis of two classifications: the Central Product Classification (CPC), an international nomenclature developed by the United Nations, and the International Standard Industrial Classification (ISIC) which is a nomenclature promoted by the United Nations to classify economic activity. However, the CPA was deemed not detailed enough for public procurement purposes. It was deemed not able to provide adequate descriptions of the works, supplies and services that public authorities were procuring.

The first version of the CPV was published in 1993 and has been revised extensively since then. Among others, there was a change from the old *materials-driven structure* to a *product structure.* Work on the preparation of the CPV nomenclature culminated with the European Commission's adoption of Recommendation 96/527/EC which led to Regulation 2195/2002. This was subsequently modified by Regulations 2151/2003, 213/2008 and Regulation 596/2009. The use of the CPV codes became mandatory in 2006. Article 23 of the 2014/24 Directive further mandated that any references to nomenclatures in the context of public procurement shall be made using the Common Procurement Vocabulary (CPV) as adopted by Regulation 2195/2002.

The classification is thus already for decades embedded in EU law and procurement practice, providing also an invaluable historical record of the evolution of procurement needs and practices[5]. Accordingly, EU contracting authorities *must* assign CPV codes to the contracts they wish to procure in the market and this categorization is routinely included in the publication of public procurement notices in the Official Journal of the European Union and via TED (Tenders Electronic Daily). On various occasions the CPV classification has been reviewed to assess its ongoing suitability. Studies on behalf of the European Commission evaluated the functioning of the current CPV and examined scenarios for future improvements both from a technical and legal standpoints. Besides the official CPV related documentation, the information in this section is derived from a number of such publications [21, 22, 23].

2.2 The CPV Structure

The CPV comprises a *main vocabulary* and a *supplementary vocabulary.* The main vocabulary is currently made up of *9,454 terms* referring to goods, works and services. Each code is made up of eight digits and a wording that describes the type of works, supplies or services forming the subject of the contract. The first eight digits make up the body of the code, and a ninth digit is added as verification but carrying no further semantics. The CPV tree structure is organized as follows:

- The first two digits identify the Divisions (XX000000-Y). The first two digits form a block and correspond to a total of possible 99 divisions (45 actual).
- The first three digits identify the Groups (XXX00000-Y) - 272 actual groups
- The first four digits identify the Classes (XXXX0000-Y) - 1002 actual classes
- The first five digits identify the Categories (XXXXX000-Y) 2379 actual categories
- Each of the last three digits provides a greater degree of precision within each category (sub-categories).

[5] A historical analysis of trends would also be valuable in GPP context

The nature of the categories at different levels (divisions, groups etc.) is not defined explicitly but only inferred from their role in the hierarchy. The CPV lacks any additional structured description for codes. The meaning is solely to be derived from the accompanying code label. The labels for goods (supplies) contain words or word combinations such as the principal name of the product and the type it belongs to. Codes corresponding to services and works contain information on the type of service provided and for what, for whom or by whom the service is provided.

2.2.1 The supplementary vocabulary

The supplementary vocabulary comprises two levels and allows additional data or qualitative elements, completing the description of the object of the contract. In total, there are 903 supplementary vocabulary items. These items are made up of an alphanumeric code with a corresponding wording, making it possible to be more precise regarding the specific nature of the goods being procured. While the supplementary vocabulary could in theory be enhanced to identify products with specific environmental impact characteristics, the reviews of CPV usage cited have found that this part of the classification it is not much used, if at all.

2.3 CPV issues and impact on GPP

There are various known issues with the CPV classification and its usage that may impact the utility of the CPV classification in GPP context, in particular the linkage with EEIO data. The historical reviews of CPV structure and CPV usage have surfaced a range of issues, the following being an indicative list:

- The hierarchical tree structure of the CPV is not always consistent (e.g. codes are not mutually exclusive). While the coverage of the CPV is generally complete it is unbalanced: different divisions, groups and classes contain very different number of single elements.
- In roughly 10–20% share of tenders the code applied does not correctly describe the works, supplies or services procured. Assigned codes might be too general or too narrow. More than one code may be used in the standard forms for the publication of public procurement notices (the first one being the titular one).
- The level of classification detail provided is not fully used in practice and also not necessary. Many codes are never used. The CPV is mandatory only for above-threshold procurement (but is used on a voluntary basis also for below-threshold procurement).
- The CPV is currently used only for publication and identification of tender notices, but not used in other phases of procurement.
- New green products or services may not fit neatly into existing CPV codes but there is no defined process for maintaining (updating) the CPV except by revising the existing regulation.

Some of the issues with the CPV structure are intrinsic to this type of knowledge management tool: a hierarchical taxonomy is not always the best approach (especially when used in isolation) when aiming to categorize a large body of items. The prevalence of *faceted taxonomies* on practically all modern e-commerce platforms and the general support for *full text search* is a hint of the additional functionalities that must served in this context. Besides data quality issues (the degree to which the CPV code is representative of a procurement activity, *including its footprint*), an important issue from a EEIO perspective is the absence of a formal classification that unambiguously maps the CPV to other existing economic classifications. The most relevant such concordance for our purposes would be the NACE classification of economic activities. While these issues do not prevent high-level applications (given the limited granularity of current EEIO databases), their importance would increase with deeper levels of the hierarchy.

2.4 The NACE classification

The NACE classification plays a critically important role in the utilization of EEIO in GPP context. As it forms the basis of the official classification of economic activities in Europe, economic and environmental data in EEIO databases typically adopt this classification system. The NACE system has its own complex history and relation with other classifications, summarized in 3.

The NACE classification is relevant to understand the *supply side* of procurement. Namely it classifies the economic entities that are providing goods and services. It is thus (potentially) available as an *ex-post data point*, once an award has been made, on the basis of the commercial identity of the contractor, e.g., as captured in chamber of commerce registrations. This data point can be an important validation mechanism next to the ex-ante CPV codes sourced from

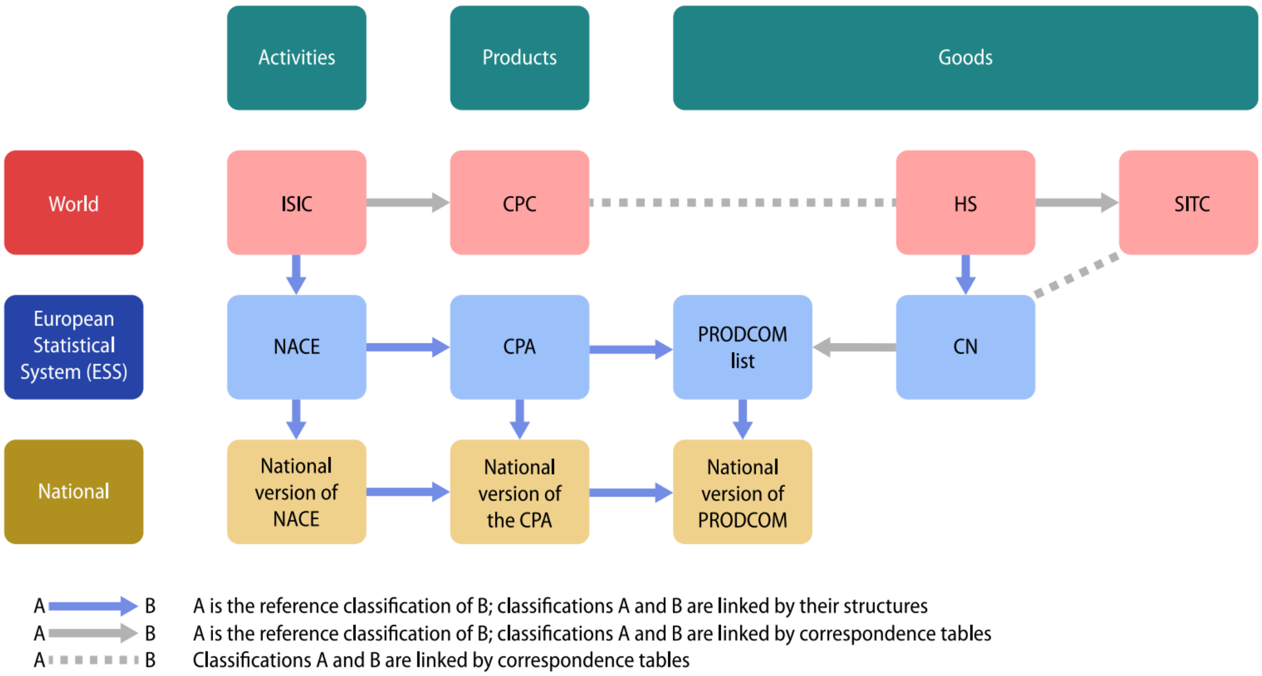


Figure 3: Overview of international and EU classification systems of economic activities. *Goods* are a subset of *Products*, while Products are aligned with *Activities*, except not necessarily on a 1 to 1 basis. The CPV classification is not formally linked to any of the above, despite having common conceptual roots.

procurement data. The NACE 2 classification has the following structure (NB: this structure has slightly changed in the recent NACE 2.1 update):

- Level 1: 21 sections identified by alphabetical letters A to U;
- Level 2: 88 divisions identified by two-digit numerical codes (01 to 99);
- Level 3: 272 groups identified by three-digit numerical codes (01.1 to 99.0);
- Level 4: 615 classes identified by four-digit numerical codes (01.11 to 99.00).

Associating CPV codes to NACE codes, formally by providing a *concordance table* between the NACE and CPV classifications enables linking procurement spending data (i.e., CPV-tagged tenders) with industry classifications (NACE) and thus for EEIO analysis. A simple CPV-NACE concordance matrix is not expected a-priory, given that the CPV has been adapted explicitly for public procurement needs by diverging from CPA. For example the CPV has *narrower scope* than NACE, as it caters specifically to public procurement needs within a single phase of the procurement lifecycle. CPV is further demand driven, procurement-focused versus the supply driven or economic activity-focused NACE categories. Not all NACE categories need be associated to CPV codes, only production activities that are relevant to public procurement. On the hand every item on the CPV classification *is* provided through the economic activities of some entity, which will be typically classified somewhere on the NACE hierarchy. Many products may be produced by the same sector but the same product might also be produced by different sectors. Finally, the CPV classification is far more granular than NACE. The implication of all the above is be that there will be many-to-many relations, and exact equivalences are not the norm. Presented side by side the hierarchical trees are shown in table 1.

3 The disaggregation challenge

Subject to the correct mapping of CPV to NACE, the attribution of average impact per sector (both direct and indirect) proceeds without major conceptual obstacles[6]. We will discuss now how to evaluate the impact of green procurement

[6]Obviously there are many potential data quality challenges

Level	NACE 2.0	CPV	Emissions Data
1	21 Sections	**N/A**	Complete
2	88 Divisions	45 Divisions	Partial
3	272 Groups	272 Groups	**N/A**
4	615 Classes	1002 Classes	**N/A**
5	**N/A**	2379 Categories	**N/A**
6	**N/A**	5756 Sub-categories	**N/A**

Table 1: Side-by-side comparison of the NACE and CPV hierarchies. There is no 1:1 correspondence of taxonomy layers. The Level 1 (sections) are used in NACE but are missing in CPV, whereas Levels 5 and 6 are populated in CPV but missing in NACE. In general CPV is the more populated classification so multiple CPV codes will map to one NACE code. Some CPV categories may be spread across NACE categories. The limited availability of emissions data at granular sector levels means that the lack of a concordance table is at present not as critical but it would become a bottleneck already at the next level of granularity.

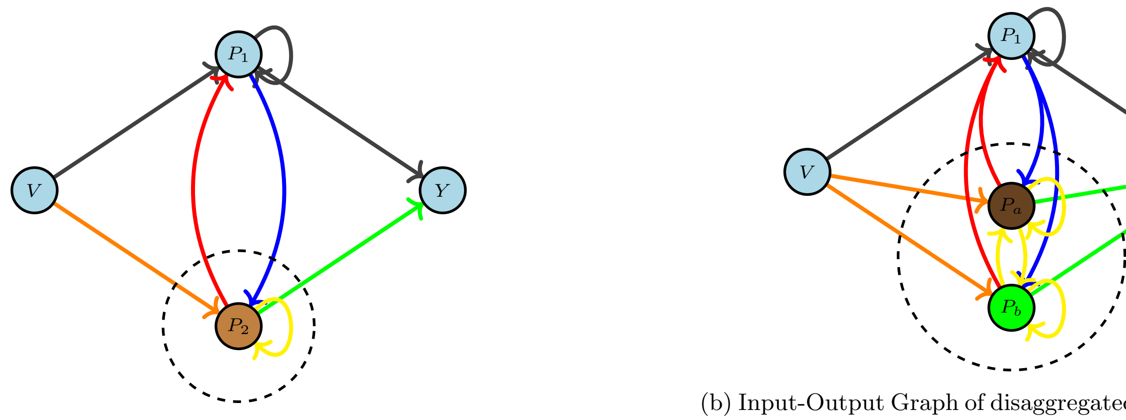
policies, a frequent requirement in the roll-out of GPP. Answering such questions stumbles on the aggregate nature of standard EEIO databases that has been described already in section 1. While in principle future highly-disaggregated EEIO systems may address GPP needs more directly, the current gap between high-level reported sectoral averages and granular product-level procurement suggests that alternative approaches might be opportune.

For each sector, a prerequisite for capturing differentiated environmental impact is obviously the *market availability* of green products. The discussion and methodology presented below thus applies to product categories where green alternatives are already available in the market and have a measurable market share[7]. The essence of the challenge is that the *assumption of homogeneity* that enables sampling approaches and compilation of EEIO databases treats all currently available products and technologies within a category as essentially identical. To "undo" this averaging requires that in addition to the input aggregated EEIO data, we introduce *additional data points*, but in a manner that is as consistent as possible with the aggregate EEIO data. Without it being an exhaustive list, some reasonable requirements from a disaggregation method would be:

- Utilize and preserve as much as possible the integrity and structure of the official EEIO databases, including all *balance constraints.*
- Avoid complex and opaque disaggregation, as it may face challenges with user acceptance. The traceability of assumptions and the ability to challenge them or override them if they do not match market developments is important.
- It should be possible to leverage diverse sources of information, provided they are consistent with the main body of the EEIO data. Older EEIO data capturing previous market structures and physical flow data compiled alongside monetary data are prime candidates.
- Use less processed EEIO tables such as *Supply-and-Use tables* (SUT) and make as few modeling assumptions as possible [24]).
- Avoid (again, as much as possible) economic model-dependent assumptions such as the long-term stability of technical coefficient matrices, in other words work preferentially with measured transactions rather than derived quantities.
- Anticipate use cases where we might want to work with Ghosh type models (analysis of downstream supply chain impacts) rather than exclusive focus on the Leontief approach which is suitable only for upstream supply chain estimates.

In this section we assume as a starting point the availability of an EEIO database that is directly usable for impact analysis and we explore how to disaggregate a specific sector. We present first general mathematical expressions and we will use a simpler example in the next section 4 to illustrate the concepts and calculation flow more concretely. Pictorially the task of disaggregation is depicted in figure 4.

[7] New green products with negligible market share can be accommodated within an EEIO framework but lean more on the what-if analysis capabilities of IO *models* as opposed to relying on existing, measured data.



(a) Input-Output Graph of aggregated product categories (P_1, P_2) with an average associated impact estimate. The economic neighborhood of the product to be disaggregated (P_2, brown circle) is indicated by the dashed circle.

(b) Input-Output Graph of disaggregated products (P_a, P_b), where the green/non-green distinction of their respective environmental impacts is now explicit. After the disaggregation we have a darker brown (heavier impact) and a green (lighter) category of sub-products.

Figure 4: Side by side the graphs of an Input-Output graph (or network) and the refinement that is required to achieve the desired disaggregation into green/non-green products. Value added (such as labor) is entering the dashed circle (orange arrow) from the V node. Final demand (consumption) is exiting the circle (green) into the Y node. Public procurement is *part* of that final demand node. Production requires inputs from other sectors (blue) and in turn *may* be used as input by other sectors (red). Finally, it is possible that there is *self-use* by a sector (e.g. energy to produce energy), presented as the yellow self loop. These arrows are *weighted*, as they encode the size of monetary transactions. When disaggregated, the weights of each color must be preserved. In addition, the total output and input (the sum of all arrows entering and leaving a node) must balance both before and after the split. These constraints still leave open considerable freedom in the determination of the relative weight of different exchanges. Pinning down the disaggregated structure requires new data inputs and/or assumptions.

The key quantitative input from an EEIO database is a set of jointly derived numerical *matrices* that represent the economy under consideration:[8]

- $\bar{Z}$, is the economic **transactions matrix**, representing exchanges (e.g. in Euros) between producing sectors (whether those operate in the primary, secondary or tertiary segments).
- $\bar{Y}$, is the **final demand matrix**, representing the purchases for consumption by households and the public sector. Public procurement spending (per year) is precisely in this accounting element.
- $\bar{V}$, is the **value added matrix**, representing auxiliary inputs into the producing sectors. This is principally labor inputs in the corporate sector.
- $\bar{f}$, is the environmental stress **intensity matrix**, quantifying impact per unit of production.

All of these matrices, except the environmental stress intensities $\bar{f}$, are expressed in monetary values. As noted, frequently there is availability of physical unit equivalent flows and those might have important uses in GPP, but we do not digress on this option. By convention, we will mark with a bar the matrices that are verbatim copies from the database and we will use the same letter without a mark to denote the disaggregated version (which aims to distinguish green from non-green products).

Under the one-to-one sector/product mapping typical in symmetric EEIO systems we will talk about *products* rather than sectors, as the objective is to investigate the disaggregation into green/non-green products. Without loss of generality we will focus on only one product, which is to be split into its green and non-green version. More products can be handled iteratively. We further assume that the product of our focus is represented by the last index.

[8]So-called Multi-Regional databases will include a number of countries - which helps address the question of domestic versus foreign environmental impacts. For simplicity of presentation we do not track here the geographic distinction. There is no loss of conceptual generality: The product index is assumed "flattened" to cover each of the sector/products produced in different countries. In practice there are extra data quality implications when working with large MRIO's, as it is not easy to compile such databases from heterogeneous statistical sources.

The objective is to obtain a disaggregated set of matrices Z, Y, V, f that will be used to inform us on the more detailed impact of procurement activities. The precise manner of such disaggregation can only be prescribed once there is an investigation of what relevant additional data sources are available. This will in general depend on the contracting authority context (the data available in their own systems, any other statistical information they have access to etc.) and the product category. In any case, though, we split the additional data required for a useful disaggregation of the EEIO into *primary and secondary data sets* to highlight the role they play in the disaggregation. This split is motivated by the fact that the primary data set affects immediately the direct emissions attributed to production (the Scopes 1 and 2), whereas the secondary data set affects the only the embodied emissions profile (the Score 3 upstream effects). Another difference between the two is that primary data can be specified directly from market research, whereas the secondary data must satisfy the same accounting integrity conditions as the aggregated EEIO system.

3.1 Primary Data

The primary data set concerns the availability, market share and environmental impact properties of alternative (but functionally equivalent) products that serve the same procurement need. The differentiation in environmental impact may come from any part of a product life cycle. Minimally we need the *market share* of sustainable versus standard versions of products and the environmental impact difference (*direct emission intensities*). Thus on the one hand the ratio of volumes produced respectively of each sub-product and the relative improvement of impact intensity by the green product. Without such a primary dataset it is difficult to conceive any data-driven methodology for disaggregation.

3.1.1 Disaggregation of Total Input / Output

Total output $\bar{X}$ is the row sum of transactions matrix $\bar{Z}$ and final demand $\bar{Y}$. It is a column vector of dimension $n+1$. As the name implies, it represents the total production of goods and services in the economy. Equality of total input and total output must be preserved also under disaggregation if the system is to continue representing the economy faithfully. The total output vector plays a fundamental role in disaggregation approach discussed here: the proportion of green/non-green sub-products for a given category is a key input and will be used repeatedly. We can apply this procedure for each NACE industry sector/product iteratively. Only sectors with material relevance in GPP context need be disaggregated. As discussed already, this screening follows from an overall attribution to sectors. Absolute high-impact sectors where different sub-categories have materially different emissions intensities are the obvious focus areas.

Computationally, the connection between the two systems is obtained via suitable *grouping* and *disaggregation* matrices, which in turn derive from the market share data. The *grouping matrix* produces aggregated quantities when left-multiplying a disaggregated object. It is a matrix S of size $(n+1) \times (n+2)$, or $(n+1)$ rows and $(n+2)$ columns:

$$S = \begin{bmatrix} \mathbf{I}_n & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \tag{1}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix with unit values along the diagonal. The S matrix has all column-sums equal to one:

$$\sum_{i}^{n+1} S_{ij} = 1, \text{ for } j = 1 \text{ to } n+2 \tag{2}$$

The aggregation of the total output vector from the disaggregated version reads:

$$\bar{X} = SX \tag{3}$$

$$\bar{X}_i = \sum_{j}^{n+2} S_{ij} X_j, \text{ for } i = 1 \text{ to } n+1 \tag{4}$$

which resolves into the following expressions which simply state that we have a number of sectors that don't change (the *common sectors*) and the joint output of green/non-green products must match the total output in that category. For clarity, instead of generic index notation, we will denote with a the disaggregated non-green product and with b its green alternative.

$$\bar{X}_i = X_i, \text{ for } i = 1 \text{ to } n \tag{5}$$
$$\bar{X}_{n+1} = X_a + X_b \tag{6}$$

The inverse operation, constructing a disaggregated system from an aggregated one requires the weight matrix that associates two sub-categories of products with their corresponding aggregates. This matrix includes the weights expressing the market share of each sub-category of product within the overall market for this sector.

$$W^T = \begin{bmatrix} \mathbf{I}_n & 0 \\ 0 & w_a \\ 0 & w_b \end{bmatrix} \tag{7}$$

where

$$w_a = X_a/\bar{X}_{n+1} \tag{8}$$
$$w_b = X_b/\bar{X}_{n+1} \tag{9}$$
$$\tag{10}$$

The weights (shares) of the sub-products must sum to one ($w_a + w_b = 1$) which follows from $X_a + X_b = \bar{X}_{n+1}$. As stated, these market shares are exogenous data that must be provided by studying the sector/product that is to be disaggregated. Disaggregation is achieved via matrix multiplication:

$$X = W^T \bar{X} \tag{11}$$
$$X_i = \sum_{j}^{n+1} W_{ij}^T \bar{X}_j, \text{ for } i = 1 \text{ to } n+2 \tag{12}$$
$$\tag{13}$$

which resolves to

$$X_i = \bar{X}_i, \text{ for } i = 1 \text{ to } n \tag{14}$$
$$X_a = w_a \bar{X}_{n+1} \tag{15}$$
$$X_b = w_b \bar{X}_{n+1} \tag{16}$$

3.1.2 Environmental Impact Disaggregation

The direct environmental impacts intensity is an additional critical piece of primary data we need. These are associated with the production processes of each product. They are generally known from fundamental process analysis (IPCC, IEA) and related databases. The aggregated database includes only one such impact intensity vector for each environmental stress and sector modeled.

$$\bar{f} = \begin{bmatrix} \bar{f}_1 \\ \vdots \\ \bar{f}_{n+1} \end{bmatrix} \tag{17}$$

These are converted to absolute direct emission values per product by multiplying with the total output vector (production in Euros is multiplied by impact per Euro):

$$\bar{E}_i = \bar{f}_i \bar{X}_i \tag{18}$$

We assume that specific intensities f_a, f_b can indeed be sourced for the sub-products of interest, with f_a being the emissions intensity of the non-green product (higher than the average $\bar{f}_{n+1}$) versus f_b the lower intensity of the green product. The result would be an augmented intensities vector:

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \\ f_a \\ f_b \end{bmatrix} \tag{19}$$

The disaggregated total impact per product follows by multiplication with the product output:

$$E_i = f_i X_i \tag{20}$$

3.1.3 Disaggregation of Final Demand and Value Added

The final demand and value added matrices are $\bar{Y}, \bar{V}$ and Y, V respectively. The disaggregation of final demand and value added can be based on further market research data *or* by adopting the same proportionality assumption, namely that the two sub-products exhibit with respect to these exchanges the same proportions as with their overall production output[9].

Under the assumption of proportionality to total output we get:

$$Y = W^T Y \tag{21}$$

$$V^T = W^T V^T \tag{22}$$

In words the above assumptions are stating that the ratio of purchases of the two sub-products by end-consumers (households and public sector) is the same as the respective purchases of intermediate producers (companies). Similarly, value added inputs (like labor) are also weighted by the production proportions. Clearly if there are data that indicate those economic relations to be materially different, it is important to use adjusted proportions.

3.2 Secondary Data

The above primary data set does not incorporate any information about the market structure of *production inputs*. These are captured by the transaction matrix Z. To the degree that upstream Scope 3 emissions are important we are still in the dark as to how get differentiated estimates (gren/non-green split). The additional degree of fidelity and utility is provided by incorporating a resolved model for the economic structure (the shape of the supply chain) for the various subtypes of products. This secondary data set is more subtle (and potentially more difficult to source). It concerns estimating *supply chain differences* in the production of green versus non-green products. Data sourcing strategies might involve inferring the market structure of non-green production from EEIO datasets of earlier years (when the green products were not yet available in the market).

In the absence of such data an assumption that is sometimes done is that the upstream supply chain of environmentally friendly products is similar to generic products. This then leads to a disaggregation on a proportional basis, i.e. reflecting again the market shares of production. The modeling bias of this assumption depends on the nature of the sustainable product (the degree to which it shares supply chains with the generic products) and its relative market share. The smaller the market share, hence the smaller its influence in the current average estimates, or the more radically different its supply chain from older technologies, the more likely that the assumption will introduce quantitative discrepancies with ground truth. NB: this *same technology assumption* does not immediately obviate the split into green/non-green product. Their impact difference might be entirely due to differences in what we called primary data. It is in theory possible that an industrial sector can produce two products with different environmental impacts even while utilizing more or less the same inputs. For example the improvement might be via intangible changes in underlying production technology: an improvement in furnace burning efficiencies or a more aerodynamic vehicle design etc.

[9]For the purposes of GPP we will also need to disaggregate the portion of final demand that is due to public procurement. We leave that exercise for later, as it is of a different nature from the green/non-green disaggregation we discuss here.

3.2.1 The Economic Transactions Matrix

The aggregate transactions matrix $\bar{Z}$ is of dimensions $(n+1) \times (n+1)$, where n is the number of products that will remain aggregated (common sectors) and $n+1$ is the index of the product to be split into green/non-green sub-products. The disaggregated matrix Z describes the same economy as $\bar{Z}$, the only difference being that product $n+1$ is to be segmented into two distinct sub-categories (a, b). The disaggregated matrix Z (with elements Z_{ij}) has dimension $(n+2) \times (n+2)$. Each one of the last two sub-categories of Z will correspond to the products with materially different environmental footprint. The n products that are not disaggregated are referred to as the common products, while the sub-categories created from the disaggregated class are referred to as the *new products*. Our objective now is to investigate in more detail what kind of new information is needed, and what constraints and assumptions are reasonable to make towards the disaggregation of matrix $\bar{Z}$ into matrix Z. Keeping track of the number of variables, the matrix Z has $n^2 + 4n + 4$ elements in total, whereas $\bar{Z}$ provides only $n^2 + 2n + 1$ data points. Because the common sectors in the matrix Z are unchanged we have the conditions:

$$\bar{Z}_{ij} = Z_{ij}, \text{ for } i, j = 1 \text{ to } n \tag{23}$$

Those fixes n^2 values of the matrix Z and leave unconstrained the remaining $4n+4$ values. These must be fixed either with new input data or reasonable assumptions. The aggregated system satisfies the *accounting balance conditions* that total output is equal to total input.

$$\bar{X}_i = \sum_j^{n+1} \bar{Z}_{ij} + \bar{Y}_i \tag{24}$$

$$= \sum_j^{n+1} \bar{Z}_{ji} + \bar{V}_i, \text{ for } i = 1 \text{ to } n+1 \tag{25}$$

$$\tag{26}$$

We require that those continue being satisfied in the disaggregated form:

$$X_i = \sum_j^{n+2} Z_{ij} + Y_i \tag{27}$$

$$= \sum_j^{n+2} Z_{ji} + V_i, \text{ for } i = 1 \text{ to } n+2 \tag{28}$$

$$\tag{29}$$

These are $2(n+2)$ additional conditions in total, but $2(n+1)$ of them are already satisfied identically due to the aggregate balance conditions. This leaves only 2 new constraints and this brings the values to be determined further down to $4n+2$. There are further n constraints associated with inputs of new products into common products:

$$\bar{Z}_{n+1,j} = Z_{aj} + Z_{bj}, \text{ for } j = 1 \text{ to } n \tag{30}$$

and n constraints associated with output of common products into new products.

$$\bar{Z}_{i,n+1} = Z_{ia} + Z_{ib}, \text{ for } i = 1 \text{ to } n \tag{31}$$

and finally, one constraint of internal transfers between new products:

$$\bar{Z}_{n+1,n+1} = Z_{aa} + Z_{ab} + Z_{ba} + Z_{bb} \tag{32}$$

In total these are $2n+1$ constrains, which leave us with the $2n+1$ unconstrained elements of Z that must be somehow be specified using external data or assumptions. A useful way to think about those is as the elements Z_{ai}, Z_{ia}, Z_{aa} that describe the supply chain structure of the old (non-green) product. As noted, relevant information about this segment may be available in older versions of the EEIO database.

3.3 Solving the System

Given external values for Z_{ai}, Z_{ia} and Z_{aa} for the non-green product category and the above equations we can now solve for all other elements of the Z matrix in a straight-forward manner:

$$Z_{ab} = X_a - \sum_j^n Z_{aj} - Z_{aa} - Y_a \tag{33}$$

$$Z_{ba} = X_a - \sum_j^n Z_{ja} - Z_{aa} - V_a \tag{34}$$

$$Z_{bb} = \bar{Z}_{n+1,n+1} - Z_{aa} - Z_{ab} - Z_{ba} \tag{35}$$

$$Z_{ib} = \bar{Z}_{i,n+1} - Z_{ia}, \text{ for } i = 1 \text{ to } n \tag{36}$$

$$Z_{bi} = \bar{Z}_{n+1,i} - Z_{ai}, \text{ for } i = 1 \text{ to } n \tag{37}$$

$$Z_{ij} = \bar{Z}_{ij}, \text{ for } i, j = 1 \text{ to } n \tag{38}$$

$$\tag{39}$$

These expressions will in general solve the system while satisfying all constraints. One caveat is the possibility that imputed values for Y_a and/or V_a conflict with technology assumptions and this manifests as negative values for transactions, a clear signal that something is not consistent.

3.3.1 The Same Technology Assumption

The $2n + 1$ values we need to specify are what Wolsky calls the *distinguishing parameters*[10]. It is useful to frame these parameters in relation to the corresponding aggregated elements $\bar{Z}_{n+1,i}, \bar{Z}_{i,n+1}$ and $\bar{Z}_{n+1,n+1}$, in other words, the degree to which the non-green product market structure *differs* from the aggregate sector structure. Using the so-called *technology matrix* can facilitate providing additional input values for market structure (as the purpose of this matrix is to normalize the absolute figures of transfers between sectors that are capture by the $Z, \bar{Z}$ matrices into *rates*). In the language of EEIO the underlying "technology" of production is captured effectively by enumerating the *ratios of purchases and sales to total output*. This is expressed by introducing matrices $\bar{A}, A$ that factorize the production output volume from the transaction matrix elements. The technology matrices are in some sense similar to *chemical stoichiometry matrices*: they specify the proportions of different inputs that are needed to produce a unit of output. When we factorize Z in this way we get:

$$\bar{Z}_{ij} = \bar{A}_{ij}\bar{X}_j \tag{40}$$

$$Z_{ij} = A_{ij}X_j \tag{41}$$

$$\tag{42}$$

The aggregate technology matrix $\bar{A}$ is known (or it can be immediately derived). The disaggregated version A is obviously not known, but it facilitates implementing the "same technology" assumption we discussed above [17]. This assumption, can be used to *bootstrap* a consistent imputation of new data by postulating that the two sub-products are actually produced with exactly the same technology, and then optionally adjusting as appropriate. Under the same technology assumption we obtain the following specification:

[10]In his approach they are actually $2n + 3$ as he does not take into account the additional two balancing constraints

$$A_{aa} = w_a \bar{A}_{n+1,n+1} \tag{43}$$
$$A_{ab} = w_a \bar{A}_{n+1,n+1} \tag{44}$$
$$A_{ba} = w_b \bar{A}_{n+1,n+1} \tag{45}$$
$$A_{bb} = w_b \bar{A}_{n+1,n+1} \tag{46}$$
$$A_{ia} = \bar{A}_{i,n+1}, \text{ for } i = 1 \text{ to } n \tag{47}$$
$$A_{ib} = \bar{A}_{i,n+1}, \text{ for } i = 1 \text{ to } n \tag{48}$$
$$A_{ai} = w_a \bar{A}_{n+1,i}, \text{ for } i = 1 \text{ to } n \tag{49}$$
$$A_{bi} = w_b \bar{A}_{n+1,i}, \text{ for } i = 1 \text{ to } n \tag{50}$$
$$A_{ij} = \bar{A}_{ij}, \text{ for } i, j = 1 \text{ to } n \tag{51}$$
$$\tag{52}$$

In other words the old matrix $\bar{A}$ is expanded by only taking into account the production weights w_a, w_b.

3.4 Consistent Leontief and Ghosh Inverses

With the disaggregated system comprising matrices Z, Y, V, f now at our disposal, any of the available analyses of EEIO frameworks are now in principle available. The remaining preparatory step is to derive the standard Leontief and Ghosh matrices that enable analyzing upstream and downstream supply chains respectively. Concretely the matrices are defined as:

$$L_{ij} = (\delta_{ij} - Z_{ij}/X_j)^{-1} \tag{53}$$
$$G_{ij} = (\delta_{ij} - Z_{ji}/X_j)^{-1} \tag{54}$$

where L is the *total requirements matrix* or Leontief inverse matrix, and G is the *total sales matrix* or Ghosh matrix of the disaggregated system. Both of them are obtained from the transactions matrix and total output via simple linear algebra [12]).

3.5 Analysis of GPP policies impact

The classic proposal of so-called *carbon footprinting* is the most directly relevant application of the above in the GPP context, so we expand on this, but all three types of EEIO based attribution methodologies are worth having available [25], namely:

$$E_i^P = X_i f_i \tag{55}$$
$$E_i^U = Y_i \sum_j f_j L_{ji} \tag{56}$$
$$E_i^D = V_i \sum_j f_j G_{ij} \tag{57}$$

where producer responsibility E^P is obtained from calculating emissions from total output X, demand driven, consumer responsibility or upstream supply chain attribution E^U is linked to the final demand vector Y, while supply driven, or downstream supply chain attribution E^D is linked to the value added vector V.

From here we can proceed to calculate environmental impact from public procurement demand both in the aggregate, and when differentiating from green versus non-green products. Public procurement final demand sits alongside *domestic household consumption*, any other residual public sector consumption that is not documented via the procurement channels, and exports, which represent consumer, business and public sector demand in other countries. The procurement spending need not be explicitly identified and separated in the EEIO final demand data, i.e., the analysis can proceed on a marginal basis, calculating only the impact from the documented procurement spending. This is readily possible because the EEIO framework is essentially a linear structure[11]. If we split total final demand into two pieces $Y = P + O$ representing public procurement and all other final demand respectively we get:

[11] Reconciling total procurement spending in a country with the corresponding EEIO data would be an interesting exercise

$$\bar{E}_i = \bar{P}_i \sum_j \bar{f}_j \bar{L}_{ij} \tag{58}$$

$$E_i = P_i \sum_j f_j L_{ij} \tag{59}$$

$$\tag{60}$$

In conclusion, subject to a valid CPV-NACE mapping discussed in the previous section and the availability of a set of primary and secondary data that help consistently disaggregate an off-the-shelf available EEIO database we can attribute direct and upstream environmental impact to public procurement activities both in the aggregate and taking into account a sectoral disaggregation.

4 A worked out numerical example

We conclude with a simple numerical example to illustrate concretely the disaggregation method described in the previous section. We will reuse a minimal and stylized EEIO system that has been discussed before for exposition purposes [26, 27, 24]. We consider the challenge of green public procurement in a world with just two production sectors: Agriculture (Ag) and Manufacturing (Ma). These two sectors sell goods and services to each other and also to households (O) and the public sector (P) who purchase the final, finished products sold by each one of these two production sectors. The numbers we use below are entirely fictitious, they are merely a means to make the required calculations explicit. Following the classic IO framework, the production part of this economy is a 2×2 inter-industry sales or transactions matrix $\bar{Z}$:

$$\bar{Z} = \begin{bmatrix} 8 & 5 \\ 4 & 2 \end{bmatrix} \tag{61}$$

This matrix expresses the economic transactions (in monetary units, lets say Euro) between the two business sectors (equivalently products, as we assume there is a one-to-one map between sectors and products) for a given year. I.e., they express how much is spent and earned by each sector from the other sector. Concretely, the example indicates that the agriculture sector purchases 4€ worth of goods and services from the manufacturing sector and 8€ worth of goods and services from from itself. This self-trading data point expresses the fact that in EEIO databases the sectors are substantially aggregated. Thus, these diagonal elements capture the amounts of purchases and sales between companies *within the same sector*. This will be more prevalent for some sectors (e.g. it takes energy to produce energy).

4.1 Final Demand and Value Added

The total aggregated final demand (from both households and public sector) is $\bar{Y}$

$$\bar{Y} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} \tag{62}$$

The value added (e.g. salaries paid for labor) into the two product categories is $\bar{V}$

$$\bar{V}^T = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \tag{63}$$

We split demand into household and public sector consumption, assuming there is final demand of 2€ for agricultural products by households, 1€ for public procurement etc. This means final Demand by households H is:

$$O = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \tag{64}$$

and final demand from public procurement is P is:

$$P = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{65}$$

4.2 Total Input / Output

The total output $\bar{X}$ produced by the two sectors is

$$\bar{X} = \begin{bmatrix} 16 \\ 12 \end{bmatrix} \tag{66}$$

and is equal to the total input $\bar{X}^T$

$$\bar{X}^T = \begin{bmatrix} 16 & 12 \end{bmatrix} \tag{67}$$

All in all, up to this point we have the following collective tabular representation of this simple economy:

$$\begin{bmatrix} & Ag & Ma & H & P & \bar{X} \\ Ag & 8 & 5 & 2 & 1 & 16 \\ Ma & 4 & 2 & 3 & 3 & 12 \\ V & 4 & 5 & 0 & 0 & \\ \bar{X}^T & 16 & 12 & & & \end{bmatrix} \tag{68}$$

4.2.1 Total Output Disaggregation

We imagine now that the manufacturing sector can produce products by using either renewable energy or fossil fuels. We assume that our market research department established that 20% of the total manufactured products sold in the economy have been produced using factories powered with renewable energy, whereas the remaining 80% used fossil fuel power plants. This is the first piece of *primary data*. Thus the grouping matrix is:

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \tag{69}$$

whereas the weight matrix W is given by:

$$W^T = \begin{bmatrix} 1 & 0 \\ 0 & w_a \\ 0 & w_b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0.8 \\ 0 & 0.2 \end{bmatrix} \tag{70}$$

The disaggregated total output is obtained by matrix multiplication:

$$X = \begin{bmatrix} 16 \\ 9.6 \\ 2.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0.8 \\ 0 & 0.2 \end{bmatrix} \begin{bmatrix} 16 \\ 12 \end{bmatrix} \tag{71}$$

4.2.2 Environmental Impact Intensities

Let us further assume that an emissions inventory of this economy reported that production activities in the Ag sector resulted in the emission of 8 tons of carbon, whereas activities in the manufacturing sector resulted in the emission of 4 tons of carbon. So the total carbon budget of this economy is 12 tons of carbon.

Dividing element-wise the emissions vector E by total output X gives us the emission intensities per product:

$$\bar{f} = \bar{E}/\bar{X} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} / \begin{bmatrix} 16 \\ 12 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.33 \end{bmatrix} \tag{72}$$

Ag has thus a direct emissions intensity of 8 tons CO2/16€, or 0.5 tons CO2/€ and Ma has a direct intensity of 0.33 tons CO2/€. The aggregate direct intensity vector $\bar{f}^T = [0.5\, 0.33]$ describes the emissions intensity for the two sectors in this economy in aggregated form. We must now apportion the total direct emissions of the manufacturing sector to its two underlying production technologies. For this we need the emissions intensities of the respective manufacturing processes (which need to be consistent with the total sector emissions). Suppose the numbers are as follows (this constitutes the second piece of primary data):

$$f = \begin{bmatrix} 0.5 \\ 0.4 \\ 0.066 \end{bmatrix} \tag{73}$$

We thus have a situation where the average intensity of the manufacturing sector (0.33) has been decomposed into two constituent intensities (the green sector with an intensity of 0.066) and the non-green sector with an intensity of 0.4. Obviously the total emissions from production are preserved:

$$E^T = 0.5 * 16 + 0.33 * 12 = 0.5 * 16 + 0.4 * 9.6 + 0.066 * 2.4 = 12 \tag{74}$$

Finally, we need to establish how final demand and value added are distributed over the sub-products. In the absence of information we might assign them by proportion to total output:

$$Y = \begin{bmatrix} 3 \\ 4.8 \\ 1.2 \end{bmatrix} \tag{75}$$

which can be further split on household and public demand by the known proportions:

$$Y = O + P = \begin{bmatrix} 2 \\ 2.4 \\ 0.6 \end{bmatrix} + \begin{bmatrix} 1 \\ 2.4 \\ 0.6 \end{bmatrix} \tag{76}$$

and similarly for value added.

$$V^T = \begin{bmatrix} 4 \\ 4 \\ 1 \end{bmatrix} \tag{77}$$

4.2.3 Disaggregated Transactions Matrix

We finally need to disaggregate the (2 by 2) $\bar{Z}$ matrix into a (3 by 3) Z matrix:

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{bmatrix} \tag{78}$$

In this case, only one transaction value is directly known: the self-purchases of the Ag sector, or $Z_{11} = 8$. The intra-sector transactions between the two types of sub-products and their interactions with the agriculture sector are in-principle different. We have a set of equations that will determine them, but as discussed in the previous section, in the absence of any actual data inputs, we need to first impute information about the partial market structure of the non-green product: Z_{21}, Z_{12}, Z_{22}.

We do this here using same technology assumption for the non-green sector which reads:

$$A_{22} = w_a \bar{A}_{22} = 0.13 \tag{79}$$

$$A_{12} = \bar{A}_{12} = 0.42 \tag{80}$$

$$A_{21} = w_a \bar{A}_{21} = 0.2 \tag{81}$$

From these we obtain the transactions

$$Z_{22} = A_{22}X_2 = 1.28 \tag{82}$$
$$Z_{12} = A_{12}X_2 = 4 \tag{83}$$
$$Z_{21} = A_{21}X_1 = 3.2 \tag{84}$$

which enable us to compute the remaining elements:

$$Z_{23} = X_2 - Z_{21} - Z_{22} - Y_2 = 0.32 \tag{85}$$
$$Z_{32} = X_2 - Z_{12} - Z_{22} - V_2 = 0.32 \tag{86}$$
$$Z_{33} = \bar{Z}_{22} - Z_{22} - Z_{23} - Z_{32} = 0.08 \tag{87}$$
$$Z_{13} = \bar{Z}_{12} - Z_{12} = 1 \tag{88}$$
$$Z_{31} = \bar{Z}_{21} - Z_{21} = 0.8 \tag{89}$$
$$Z_{11} = \bar{Z}_{11} = 8 \tag{90}$$

or in summary:

$$Z = \begin{bmatrix} 8 & 4 & 1 \\ 3.2 & 1.28 & 0.32 \\ 0.8 & 0.32 & 0.08 \end{bmatrix} \tag{91}$$

which can be factorized into the technology matrix (notice that some of its elements are just what we assumed):

$$A = \begin{bmatrix} 0.5 & 0.42 & 0.42 \\ 0.2 & 0.13 & 0.13 \\ 0.05 & 0.03 & 0.03 \end{bmatrix} \tag{92}$$

which leads to a Leontief inverse matrix:

$$L = \begin{bmatrix} 2.67 & 1.33 & 1.33 \\ 0.64 & 1.48 & 0.48 \\ 0.16 & 0.12 & 1.12 \end{bmatrix} \tag{93}$$

4.3 Disaggregated Environmental Impact from GPP

We finally have the computational tools we need to establish the differentiated impact metrics. Let us first compute the carbon footprint of all final demand in the aggregated case:

$$\bar{E}_1 = \bar{Y}_1 \sum_j \bar{f}_j \bar{L}_{j1} \tag{94}$$
$$= \bar{Y}_1(\bar{f}_1\bar{L}_{11} + \bar{f}_2\bar{L}_{21}) = 4.8 \tag{95}$$
$$\bar{E}_2 = \bar{Y}_2 \sum_j \bar{f}_j \bar{L}_{j2} \tag{96}$$
$$= \bar{Y}_2(\bar{f}_1\bar{L}_{12} + \bar{f}_2\bar{L}_{22}) = 7.2 \tag{97}$$
$$\tag{98}$$

Notice that while the sum of emissions attributed to final demand continues to be 12, the attribution to products is quite different $(4.8, 7.2)$ versus $(8, 4)$.

In the same fashion we can compute disaggregated emissions:

$$E_1 = Y_1 \sum_j f_j L_{j1} \tag{99}$$
$$= Y_1(f_1 L_{11} + f_a L_{a1} + f_b L_{b1}) = 4.8 \tag{100}$$
$$E_a = Y_a \sum_j f_j L_{ja} \tag{101}$$
$$= Y_a(f_1 L_{1a} + f_a L_{aa} + f_b L_{ba}) = 6.1 \tag{102}$$
$$E_b = Y_b \sum_j f_j L_{jb} \tag{103}$$
$$= Y_b(f_1 L_{1b} + f_a L_{ab} + f_b L_{bb}) = 1.1 \tag{104}$$
$$\tag{105}$$

which show that the 7.2 carbon footprint of the manufacturing sector is split into 6.1 and 1.1 respectively for non-green/green products. The green product is thus less green when taking into account its supply chain.

Finally, the impact of the public procurement is obtained simply by replacing total demand with public procurement spending:

$$E_1 = P_1(f_1 L_{11} + f_a L_{a1} + f_b L_{b1}) = 1.6 \tag{106}$$
$$E_a = P_a(f_1 L_{1a} + f_a L_{aa} + f_b L_{ba}) = 3.0 \tag{107}$$
$$E_b = P_b(f_1 L_{1b} + f_a L_{ab} + f_b L_{bb}) = 0.6 \tag{108}$$
$$\tag{109}$$

References

[1] Jai Krishna Ranganathan and Miriam Thiemann. Green Public Procurement: Recommendations for the EU Public Procurement Directive. *European Environmental Bureau (EEB)*, 2025.

[2] OECD. Harnessing Public Procurement for the Green Transition. *OECD Public Governance Reviews*, 2024.

[3] P. Papadopoulos. WP17: Sustainable Public Procurement. Part I: Emissions Attribution. *Open Risk White Papers*, 2025. Online Link.

[4] Open Contracting Partnership. Open Contracting Data Standard (OCDS). Link.

[5] P. Papadopoulos. WP11: Sustainable Portfolio Management: Attribution and Allocation of Greenhouse Gas Emissions. *Open Risk White Papers*, 2021. Online Link.

[6] IPCC. Guidelines for National Greenhouse Gas Inventories. *Report*, 2006.

[7] W.K. Fong et al Greenhouse Gas Protocol. Global Protocol for Community-Scale Greenhouse Gas Inventories. *Report*, 2021.

[8] Greenhouse Gas Protocol. A Corporate Accounting and Reporting Standard, Revised Edition. *Report*, 2004.

[9] Greenhouse Gas Protocol. The GHG Protocol for Project Accounting. *Report*, 2005.

[10] Greenhouse Gas Protocol. Technical Guidance for Calculating Scope 3 Emissions. *Report*, 2013.

[11] Manfred Lenzen and Joy Murray. Conceptualising environmental responsibility. *Ecological Economics*, 70:261–270, 2010.

[12] Ronald E. Miller and Peter D. Blair. *Input-Output Analysis: Foundations and Extensions*. Cambridge, 2008.

[13] Bram Edens, Rutger Hoekstra, Daan Zult, Oscar Lemmers, Harry Wilting, and Ronghao Wu. A method to create carbon footprint estimates consistent with national accounts. *Economic Systems Research*, 27(4):440–457, 2015.

[14] Michael Z. Hauschild, Ralph K. Rosenbaum, Stig Irving Olsen eds. *Life Cycle Assessment, Theory and Practice.* 2018.

[15] Steubing, Bernhard and de Koning, Arjan and Merciai, Stefano and Tukker, Arnold. How do carbon footprints from LCA and EEIOA databases compare? A comparison of ecoinvent and EXIOBASE. *Journal of Industrial Ecology*, 26(4):1406–1422, 2022.

[16] John Ching-Han Fei. A Fundamental Theorem for the Aggregation Problem of Input-Output Analysis. *Econometrica*, 24(4):400–412, 1956.

[17] Alan M. Wolsky. Disaggregating Input-Output Models. *The Review of Economics and Statistics*, 66(2):283–291, 1984.

[18] Sören Lindner, Julien Legault and Dabo Guan. Disaggregating Input-Output Models with Incomplete Information. *Economic Systems Research*, 24(4):329–347, 2012.

[19] Dominique Maxime, Maxime Agez. Environmental footprint of procurements by Shared Services Canada. *CIRAIG Report*, 2020.

[20] M.A. Steenmeijer et al. The environmental impact of EUR 85 billion in annual procurement by all Dutch governments. *RIVM report 2021-0220*, 2021.

[21] CEN National Members. CWA 16138: Classification and catalogue systems used in electronic public and private procurement. *Report*, 2010.

[22] K. Attstrom, R.Krober, M.Jungclaus. Review of the functioning of the CPV codes/system. *Company Report to the European Commission*, 2010.

[23] A. Kuljanin, R.Krober, M.Jungclaus. Revision of CPV. *Company Report to the European Commission*, 2017.

[24] P. Papadopoulos. WP16: Follow the Money: Random Walks on Supply and Use Graphs. *Open Risk White Papers*, 2024. Online Link.

[25] Blanca Gallego and Manfred Lenzen. A Consistent Input-Output Formulation of Shared Producer and Consumer Responsibility. *Economic Systems Research, Vol. 17, No. 4, 365 –391*, 2005.

[26] J. Kitzes. An Introduction to Environmentally-Extended Input-Output Analysis. *Resources, 2, 489-503*, 2013.

[27] Jonas Bunsen and Matthias Finkbeiner. An Introductory Review of Input-Output Analysis in Sustainability Sciences Including Potential Implications of Aggregation. *Sustainability*, 15(1), 2023.