Going global to local: achieving agri-food sustainability from a spatially explicit input-output analysis perspective
Sun, Z.

Citation

Sun, Z. (2021, June 1). *Going global to local: achieving agri-food sustainability from a spatially explicit input-output analysis perspective*. Retrieved from https://hdl.handle.net/1887/3180744

Universiteit Leiden

2.

Chapter 2.
Going global to local:
connecting top-down accounting and
local impacts, a methodological review of
spatially explicit input-output approaches

2 Going global to local: connecting top-down accounting and local impacts, a methodological review of spatially explicit input-output approaches [1]

Abstract

Environmentally Extended Input-Output databases (EEIOs) provide an effective tool for assessing environmental impacts around the world. These databases have yielded many scientific and policy relevant insights, especially through the national accounting of impacts embodied in trade. However, most approaches average out the spatial variation in different factors, usually at the level of the nation, but sometimes at the subnational level. It is a natural next step to connect trade with local environmental impacts and local consumption. Due to investments in earth observation many new datasets are now available, offering a huge potential for coupling environmental datasets with economic models such as Multi-Region Input-Output (MRIO) models. A key tool for linking these scales are Spatially explicit Input-Output (SIO) models, which provide both demand and supply perspectives by linking producers and consumers. Here we define an SIO model as a model having a resolution greater than the underlying input-output transaction matrix. Given the increasing interest in this approach, we present a timely review of the methods used, insights gained, and limitations of various approaches for integrating spatial data in input-output modelling. We highlight the evolution of these approaches, and review the methodological approaches used in SIO models so far. We investigate the temporal and spatial resolution of such approaches and analyze the general advantages and limitations of the modelling framework. Finally, we make suggestions for the future development of SIO models.

2.1 Introduction

Environmentally-Extended Input-Output (EEIO) models have been widely applied and have been used to link production and consumption while accounting for the direct and indirect relationships between different economic activities [1,51,52]. Prominent consumption based studies include analyses of air emissions [53–55], waste generation [56], water use [57], land use [58], and biodiversity loss [59] around the world [60]. Part of the popularity of EEIO databases (EEIOs) is due to the steady increase in the level of environmental impacts embodied in trade [1,52]. Additionally, since these models connect producer and consumer through supply chains [61] (which are often complex), it is possible to investigate policy interventions from production-based [62–64], consumption-based [51,65], income-based [66,67], and other, in betweenness-based perspectives [68].

Currently, the vast majority of EEIO applications are based on results at the national level. This is acceptable for well-mixed, global environmental stressors such as greenhouse gases, and for broader investigations on a national level, but it limits the usefulness of models for stressors which have highly local impacts, and for nexus investigations which examine the interaction and interdependence of several resources. Particular examples of these types of stressor include water use, land use, biodiversity, water pollution, and local air pollution (such as SO_2, NO_x, $PM_{2.5}$, PM_{10}). Given this, there has been a recent trend to link EEIOs with global environmental maps and databases by disaggregating modelled or directly measured production activity by sector [69–73].

Maps and databases of environmental impacts or stressors are typically generated from observations by monitoring stations [74] and satellite remote sensing measurements [75]. They can

---

also be modelled by using spatially explicit simulations that often use direct observations as model boundary conditions [71,76]. Monitoring stations collect environmental information *in situ*, with common examples including air quality ($PM_{2.5}$, PM_{10}, O_3, SO_2, CO, NO_2) (for example see: http://aqicn.org), soil quality [77], and water quality [78]. Remote sensing aims to measure environmental impacts from a distance, including land use [30,76], water [79], air quality [75], and biodiversity [80]. Remote sensing may include ground-, sky-, or space-based observation. The spatial distribution of environmental impacts derived from these methods can be very accurate and are often available over time. For example, fixed monitoring stations record in real time, while satellite imagery products for air quality and land use are generally updated annually [29,75]. Most efforts are carried out by governmental, public research institutions, or some private research groups. However, there is always a balance between large-scale, continuous modelling, and the scientific resources available to make such global assessments. In response, researchers have also used spatially explicit simulation models to estimate the spatial distribution of environmental impacts at higher temporal or spatial resolutions. Such examples include: the global dynamic vegetation model LPJmL (Lund-Potsdam-Jena managed Land) [81]; the global freshwater model WaterGAP [82], the dynamical atmospheric Sulphur transport model DEHM [83]; a variant of this model DEHM-POP to depict transport of persistent organic pollutants (POPs) [83]; and a historic model of environmental impacts (HYDE) to simulate land use and land cover change over time [84]. These examples depict environmental issues caused by local production, but they do not connect local environmental impacts with consumption from other regions. Additionally, they do not contain enough sectoral information to provide information on policy impacts.

What we have described so far is the estimation of environmental impacts from a supply perspective, that is, in the production of materials for good and services. These data can be connected to the consumer in several different ways. We focus on input-output models in this review, but there are other methodologies for assessing consumption-side impacts using process-based methods to identify spatially explicit environmental impacts [85]. For example, Hoekstra and Mekonnen et al. estimated the spatial distribution of global water footprints (blue, green, and grey water footprints) using a grid-based dynamic water balance model, calculating virtual water flow from water embodied in the direct consumption of agricultural and industrial commodities [86,87]. This approach describes the embodied impacts through direct consumption of commodities between two regions, but they omit the complex supply-chain relationships between different sectors and different regions.

With the increasing availability of spatially explicit environmental data disaggregated by sector, there have been several efforts to take EEIO analysis to the local level. As yet these approaches are disparate and spread across the literature. Given the likely similarities of approaches, the diversity of environmental and resource assessments, and the possible utility of such approaches, we provide a critical review of the approaches so far, limitations, and future opportunities.

We start by highlighting recent efforts in SIO modelling and we categorize these analyses by the form of disaggregation used in the input-output model. We then provide an overview of spatial data sources and their resolution. The need to balance spatial resolution with policy recommendations is addressed. We discuss the potential for uncertainty analysis in SIO investigations and the possibility for further incorporation with other environmental models. Finally, we highlight the major obstacles going forward in developing, utilizing, and extending SIO models.

2.2 An expanding field

Although the term *spatially explicit* implies a variety of meanings, there is no uniform definition. Here we define it as involving a result where the spatial information available from a study is at a spatial scale greater than the available IO (input-output) data itself. Studies such as Wang et al.[88], Ridoutt et al. [89] , Wilting et al. [90], and Verones et al.[91], average results to the national level, and would fall under the category of input-output modelling not SIO since spatial information is not available. Multi-region input-output tables at the regional level, such as those using Chinese provincial data also do not conform to the definition above since the IO table is already at the spatial scale of the region. These regional MRIOs have been previously been reviewed in Ploszaj et al [92] . Ploszaj et al found 42 articles using subnational input-output papers based on data from 15 countries between 1980 and 2013. We omit these studies from the review herein.

Interest in incorporating spatially explicit information into EEIO is a relatively new development, with fast growth since 2014 (see Figure 2.1, and the Supporting Information for how the literature was selected). However, James et al. provided the first methodological approach in 1985, by integrating an input-output model with an air pollution dispersion model [93]. It focused on a small region (the Hunter Region, Australia) and traced the spatial diffusion of sulphur oxide and fluoride emissions from location-specific production sites given by the regional authority as part of the New South Wales Clean Air Act. These emissions were calculated based on the output from specific production sites and emission coefficients from a regional input-output table. They assumed the same emission coefficient for the specific site as the corresponding sectors in the input-output table [93]. This early approach showed that the spatial distribution of emissions varied with the level of regional economic development.

It subsequently took 20 years for the next publication of an SIO study, probably due to the limited availability of spatial datasets with adequate sectoral resolution along with the limitations of computing power. From 2010, there has been a large increase in the number of spatially explicit studies. These studies use a variety of different methods and approaches, and analyze a number of different environmental pressures. We are currently in a period of exponentially increasing citation counts for SIO papers, especially from 2013 onwards. This is likely to continue with increasing dataset availability and the fast development of SIO approaches.
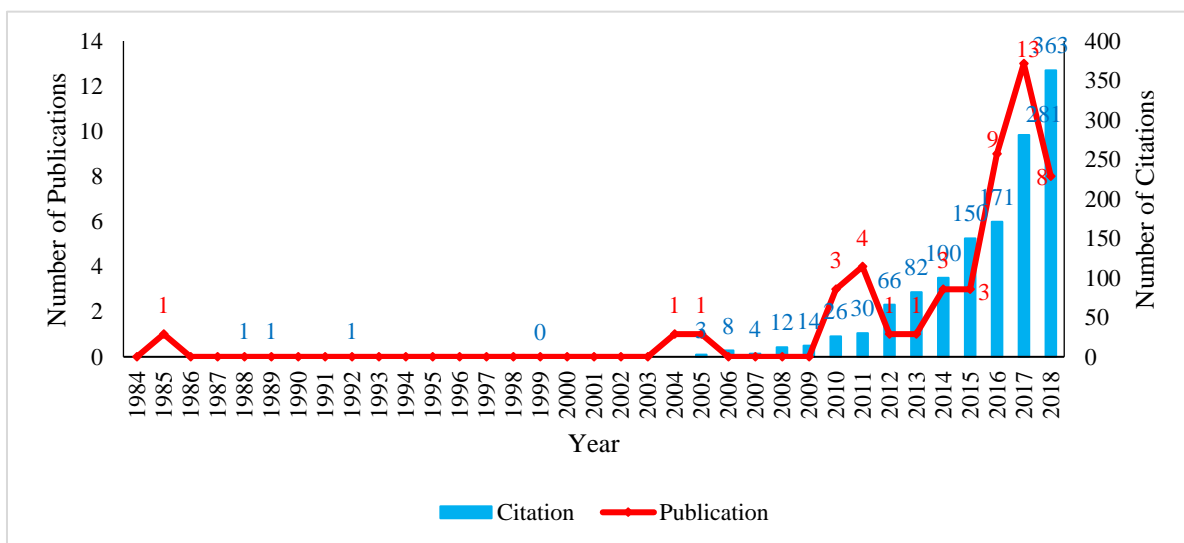

Figure 2.1 Number of published papers and their citations using SIO approaches

2.3 Methodological and spatial categories

Methodological categories: In total, we identify ten distinct methodological approaches for linking spatially explicit data to input-output databases (please see Supporting Information for qualitative and mathematical descriptions of each). We then classify these further based on the structure of Environmentally Extended Input-Output (EEIO) databases. The canonical structure for an EEIO database is a matrix for environmental extensions by sector and region, a final demand matrix by sector and region, and a transaction matrix where sectors purchase the output from other sectors to produce goods for final demand. We further classify the 10 methods to 3 categories based on the matrices to which the spatial disaggregation is applied (the studies falling into each category are shown in Table 2.1). The categories are enumerated as:

Category 1:Disaggregation in environmental extensions. Here, environmental extensions are disaggregated by mapping the environmental impacts between the production sectors in the input output model and impacts identified spatially from the spatial databases. The result is a spatially explicit mapping of consumption-based footprints. In this category, analysis mainly focuses on hotspot assessment, which can be driven by a specific country, region, or sector. Examples of spatial databases that can be used in this way include: WaterGAP [70] for fresh water use and consumption; emissions data from EDGAR [69] giving greenhouse gas (CH4, CO2, N2O), air pollution (BC, CO, NH3, NOx, PM10, PM2.5, SO2), and toxic pollutants (Mercury); the IUCN red list which provides details on threatened species [72,73]; and, Aqueduct Global Maps which maps water stress [94]. The connection between the spatial database and IO tables is usually made by assuming proportionality between impacts. That is, the demand for products in an IO table is assumed to be proportionally distributed to the production information in the spatially explicit database. Some studies have then aggregated these impacts or stressors to the national or regional level to build a set of new environmental extensions for the IO model first, and then used the same spatial information to allocate consumption-based impacts into grid cells [95,96]. Others use the existing source data for the environmental extensions provided by the IO datasets to calculate consumption-based impacts, and use spatial information to disaggregate to a finer scale [97].

Some studies then perform additional analyses to model the diffusion of a pollutant based on the consumption footprint. Typically, these studies use an input-output model to calculate the volume of emissions from a region and then apply a physical model to simulate the spatial diffusion of emissions (which may also include an atmospheric chemistry model) [98–103]. For example, the approach (Method 7 in the supporting information) has been used to investigate the human health impact in China driven by the overseas consumption of Chinese products (this also required the use of a health impact model) [104].

Category 2: Disaggregation in final demand. Here the final demand matrix is disaggregated using regional statistics. These statistics can be derived from household or enterprise surveys or from local purchasing data such as electricity bills [105–107]. This disaggregation is then used as a stimulus vector whereby the disaggregated matrix is used instead of the total national final demand (and the traditional Leontief analysis is performed). Note that this approach still uses the original national (or multi-regional) input-output transaction matrix and environmental extensions to trace upstream environmental impacts. Instead of finding spatially explicit hotspots in production driven by overseas and domestic consumption (as in Category 1), this approach is used to show the environmental footprints of consumption across regions, for example the differences in the environmental impacts of consumers in different subnational areas [108,109]. For example, Method 6 (see supporting information) was used to show spatially explicit consumption footprints in the EU[105,110]. First, the product classifications used in consumer expenditure surveys (CES) were mapped to the sectors available in the MRIO

EXIOBASE using concordance matricies at the country level. Then this combined CES-MRIO model was used to calculate subnational environmental impacts according to household consumption across the EU.

Category 3: Disaggregation in the transaction matrix. Here the transaction matrix is disaggregated, which by definition requires the spatial disaggregation of final demand and environmental extensions also [111,112]. As previously stated, this category requires the construction of an entirely new IO table in all IO elements, which does not conform to our definition above, and has been reviewed elsewhere [92]. As we will discuss later in this paper, data limitations preclude doing this for every analysis. Often we have limited information on the structure of value chains linking spatially separated production and production consumptions. Therefore, we focus mainly on Categories 1 and 2 and reflect on the further development of Category 3 studies in the discussion.

Table 2.1 Categories of SIO linked with the methods and data sources applied

Category	Example spatial database or model used	Methods	References
Disaggregation in environmental extensions	The WaterGAP model	Methods 1 & 2: Identifying hotspots	50,113
	EDGAR emissions data		95,96
	Extent-of-occurrence, from IUCN red list and BirdLife datasets		114
	Aqueduct Global Maps for water stress	Method 3: Integrating a process-based model with an input-output model	97
	IFA hazardous substance database	Method 4: Integrating an MRIO model with production location information	115
	Survey data from enterprises (SABI, Sistema de análisis de balances ibéricos: base de datos)		116
	Location of volcanic eruptions and ash volume (Auckland Volcanic Field, from Geology of the Auckland urban area)	Method 5: Quantitative risk assessment of economic output reduction due to final-demand perturbations.	117
	GEOS-Chem chemical transport model	Method 7 & 8 : Integrating an MRIO model with an air pollution dispersion model.	98–104
	Pollutant dispersion models (Smeared Concentration Approximation (SCA))		93
	Spatially explicit econometric model (spatial Regional Econometric Input–output Model (REIM))	Method 9: Integrating an econometric model with an MRIO model	118

	GIS methods and approaches	Method 10: Integrating an MRIO model with, for example, spatial interpolation	119–122
Disaggregation in final demand	Local statistical data	Method 6: Integrating an MRIO model with demand-side subnational information.	123–125
	Consumer Expenditures Survey (CES) data		107,108,134–142,126–133
	Enterprise survey data (Italian company information and business intelligence (AIDA))		106
	Zip code tabulation (U.S. zip code tabulation areas (ZCTAs))		107,143,144
	Gridded population		107
	Purchasing power index		107
Disaggregation in the transaction matrix	Disaggregation all matrices via Non-survey methods (such as location quotients (LQs), gravity models, behavior-based models, neural networks), survey methods, or hybrid methods		145–149

Spatial categories: Summarizing Categories 1 and 2, the inclusion or development of spatially explicit data in the literature arises from: 1) synthesis and incorporation of spatial environmental extensions data with EEIO (23 of the 48 articles evaluated), and 2) tracing environmental footprints using subnational stimulus vector in final demand (25 of the 48 evaluated, see supporting information). We found no papers belonging to Category 3 – papers coming closest to this represented sub-national MRIO tables at e.g. provincial level, which as discussed above we do not see as disaggregated transaction matrices at a high spatial detail but rather special cases of regular multi-country IO tables. The spatial resolution varies to a large extent both between and within categories.

Studies falling in **Category 1** used a total of 20 different spatial databases or models to trace spatial hotspots driven by consumption. These databases originate from different sources; some databases are based on remote sensing observations. For example, Moran et al. apply IUCN red list and BirdLife data, whose species-specific habitat loss is estimated from remote sensing and traces species threat driven by consumption [114]. Others use data from WaterGAP [50,113] and EDGAR [95,96], or point measurements of impacts, such as the location of volcanic eruptions [117], power plants [93], and spatial maps of national enterprises [106,116]. Based on these databases and models, researchers have explored issues including water consumption [50,113,131], nitrogen and phosphorus loading [115], biodiversity loss [96], volcanic eruption risk [117], energy consumption [129], greenhouses gas emissions (CO_2, CH_4) [95,122,130], and other air pollution emissions (NO_x, SO_2, PM_{10}, $PM_{2.5}$) [96,102]. The highest spatial resolution used was $0.5^\circ \times 0.5^\circ$ for water consumption and $0.1^\circ \times 0.1^\circ$ for greenhouse gas emission and air pollution. All studies are at the temporal resolution of a year and range from 1970 to 2008.

Studies falling in **Category 2** used 42 different databases to identify environmental impact footprints at local consumption level. These databases are mainly from Consumer Expenditures Survey (CES) (26 European countries, plus US, Austria, and Canada) according to the

international COICOP (Classification of Individual Consumption by Purpose) division. For example, Ivanova, et al., build EU27 subnational household environmental footprints (carbon, land, water, organic materials, non-organic materials) based on CES databases of every country (except Croatia, Netherlands, and Sweden) [105]. Additionally, other databases source from national statistics (such as, Australian Land Use Mapping Program, the National Pollutant Inventory, and the Australian Business Register; U.S. zip code tabulation areas (ZCTAs)) [131,143], commercial enterprise information (e.g., Italian company information and business intelligence (AIDA))) [106], *in situ* questionnaire survey data [138,142], and regional purchasing power and gridded population data [107]. These researches explores environmental impacts footprints, including greenhouse gas emission [105,126], water and land [105,108], which map to the consumption-based side of IO analysis. The highest spatial resolution mapped carbon footprints into 250 m based on GHS-POP gridded population model [107], and US household carbon footprints at zip code tabulation areas (ZCTAs). Most studies are at the temporal resolution of a year and range from 1990 to 2015, and even some researches project to 2050 [144].

Studies falling into **Category 3** would give spatially explicit information on production processes and related extensions, spatially explicit information on consumption patterns, with transaction matrices matching this spatial and sectoral disaggregation. The disaggregation of transaction matrices is a clear bottleneck. All existing work uses non-survey approaches or suffers from other crucial limitations, as illustrated by work on subnational MRIO tables (e.g. Australia [150], China [112], Japan [151], Indonesia [146], Spain [152] and Germany [149]), the Industrial Ecology Virtual Laboratory (IELab) [89,111,146] and the Transparent Supply Chains for Sustainable Economies (Trase.earth) project [109,153,154]. The IELab [89,111,150,155] includes a lot of detailed regional data (especially for Australia), from which customized input-output models can be developed based on a specific research question, but it cannot provide a input-output database including all sectors and regions due to a lack of computing power (one approach could result in over 5 petabytes of data, and impractical computation times)[111]. Trase.earth [109,153,154] focuses on constructing trade flows at finer scale (see Method 11 described in the Supporting Information), disaggregating producers on a finer scale, but consumers still at national level. None of these examples use directly-measured data or survey-based data for estimating spatially explicit transaction matrices. This is natural due to the expense of ad-hoc surveys for interregional trade data [111] . Claims such as 'city X consumes Y beef from pixel Z' by necessity requires estimates of highly estimated and modelled flows. We have found no work yet that solves this problem and hence do not discuss this category further in this section.

We classify studies into 5 further spatial categories: global, regional, national, subnational regional, and city, depending on the resolution of the final result. Table 2.2 shows a breakdown of the spatial scale and environmental impacts for different studies in the literature. We then discuss the major highlights from studies at each of these scales and their policy relevance.

Table 2.2 Studies analyzing social or environmental impacts at different spatial scales

Spatial scale	Category 1	Category 2
Global	Air pollution (SO$_2$, NO$_x$, and PM$_{10}$, PM$_{2.5}$, BC, CO)[96,98,102,104]	
	Greenhouse gases (CO$_2$, CH$_4$)[95]	Greenhouse gases (CO$_2$)[107]
	Biodiversity [114]	
	Water [50,97,113]	

	Grey water [115]	
Macro Regional		Carbon (EU27; 19 cities around the Mediterranean)[105,135]
National		Water (UK, Australia)[131]
	Grey water (Spanish)[116]	
	Air pollution (SO_2, NO_x, and $PM_{2.5}$, China) [99,101,103]	
	Atmospheric Mercury (China) [100]	
	Carbon (Japan)[122]	Carbon (Norway, USA. UK. Australia, Estonia, China, Germany)[124,125,139,143,126–128,131,132,134,137,138]
		Natural disasters (earthquakes, floods, landslides)(Italy)[106]
Subnational regional		Ecological footprint (15 cities, Canada)[133]
		Carbon (15 cities, Canada; San Francisco Bay Area in USA, 20 cities in Finland; 24 cities in China; Helsinki Metropolitan Area in Finland)[133,136,140,142,144]
City	Air pollution (Hunter region, Australia)[93]	
	COD (chemical oxygen demand) (Changzhou City, China)[119,120]	
	Volcanic eruptions (Auckland region, New Zealand)[117]	
		Economic loss driven by earthquake (Beijing, China)[123]
	Employment, population (Chicago, USA)[118]	
	Flood (South-Holland, Netherlands)[121]	
		Energy (Sydney, Australia)[129]
		CO_2 (Sydney, Australia; Boston, USA)[130,141]

Global studies: Globalization has served to disconnect commodity consumption with production-related impacts [113]. High-income countries to some extent have improved their local environmental footprints and impacts by outsourcing through the global supply chain [96]. However, these impacts will impact high-income countries as well. For example, along the US West Coast, 3-10% of annual average surface sulfate and 0.5-1.5% of ozone, both of which are

deleterious to health, arise from the atmospheric transport of Chinese pollution driven by exports [98]. In another example, Zhang et al., build a global spatial distribution estimate of premature mortality driven by $PM_{2.5}$ in 2007 at 100 km × 100 km resolution, and find that some 411,100 deaths (12% of total premature mortality) are caused by pollutant transport from one location to a more distant location, and 762,400 deaths (22% of total premature mortality) are linked indirectly through the supply chain [104].

A common analysis is the tracing of embodied environmental impacts flowing through global supply chains at the national level. Spatially explicit approaches allow for a greater resolution in assessing local environmental impacts. We illustrate this using two examples: CO_2 emissions, and biodiversity threats. Firstly, at the national level, Davis et al. identified national flows of CO_2 emissions, and identified American imports as having the largest embodied CO_2 flows in 2004 (0.7 Gt net import) [156]. By using the EDGAR database, further work made these flows spatially explicit, showing that US footprints of CO_2, SO_2, NO_x, PM_{10} in 2008 are highly concentrated, with 90% of the footprints located in only 1.6%, 3.1%, 3.6%, 9.9% of the land area, respectively 2008 [95,96]. Similarly, at the national level, Lenzen et al., find that American consumption drives the largest number of biodiversity threats (2424 total threat records, and 995 from net imports) [59]. Further spatially explicit analysis showed that 23.6% of species threats were concentrated on just 5% of global land area, and 60.7% of species threats were concentrated on 5% of the global marine area [114]. The identification of these hotspots may help facilitate global policy responses.

Macro-regional studies: A good example of macro-regional scale applications is the use of European Union data to investigate the spatial variation of environmental impacts driven by consumption. Ivanova et al. introduced a method for calculating carbon footprints, driven by household consumption in 177 regions of the EU27 [105]. They used this approach (described in Method 5 in the supplementary information) to calculate land, water, organic materials, and non-organic materials footprints under different kinds of consumption categories—shelter, food, clothing, mobility, manufactured products, and services [105].

National studies: High-income countries and lower-income countries have different perspectives for environmental impact research. For the UK, a high-income country with relatively large consumption-based impacts, researchers have focused on carbon footprints [126,127] and water footprints [108] at the scale of the local authority. Some researchers go further, attempting to find even higher resolution spatial distributions, for example of the grey footprint of Spain [116] and the carbon footprint in Estonia [128]. Since some atmospheric pollutants have highly local health effects, but can also be transported within the nation, there has been work to model the diffusion of pollutants such as $PM_{2.5}$ and mercury, which also incorporate dynamics of international and interprovincial trade [99–101].

Subnational regional studies: These studies are very useful for interregional management, especially for a large country. For example, by linking census data with input-output models, researchers were able to show that the lowest Canadian per capita carbon footprint was found in metropolitan areas, since they often share goods and services [133]. Another study investigated the opportunities for high-income and low-income consumers to reduce their carbon footprints across California, USA [144], finding that lifestyle modes have a large impact on overall carbon levels.

City studies: Compared with other scales, cities tend to have highly-local measurements of environmental impacts, for example via *in-situ* air or water pollution measurement devices. Researchers have used input-output models and local city data to estimate direct economic losses to cities from natural hazards, including earthquakes [123], volcanic eruptions [117] and floods

[121]. Analysis of the energy requirements of cities at suburban scales has also been made [129,130]. City-based input-output tables are useful for this sort of analysis, and can provide specific support for local decision makers.

Most SIO investigations are on global and national levels. This makes a certain amount of sense since global input-output models (WIOD, Eora, GTAP, EXIOBASE) and national input-output tables (official statistical publications) have been available for some time now, as have large-scale, spatially explicit, global models of environmental stressors and impacts

2.4 Options for enhancing spatial and sectoral resolution

Theoretically, an MRIO framework could provide arbitrary spatial and sectoral resolution if the data and resources are available to those constructing the models. In one sense, at the extreme, a full MRIO model could include all interactions of economic activities for very fine spatial units, for example, 1 m × 1 m. This sort of model would fall under Category 3 above and is the ultimate ideal in developing SIO models (this would involve trillions of data points) because it can reveal all sectoral and spatial heterogeneity.

This approach overcomes the spatial homogeneity assumption, which is an intrinsic shortcoming of input-output models (i.e. that each sector and region has specific environmental impacts across all products produced by that sector and across regions). However, data and computation limitations preclude such an approach for the foreseeable future. SIO models in Category 1 and 2 attempt to gain insights that such an approach might yield without the significant data and computation challenges. There is a large potential for developing these models further. In the following we expand on the opportunities for developments, first from a sectoral perspective, and then from a spatial resolution perspective. We then present some avenues for the development of approaches for Category 3.

Enhancing Category 1, the sectoral and spatial resolution of environmental extensions.

Sectoral resolution: The individual sectors included in analyses are important for further environmental and policy insights beyond the total environmental impact. However, it is difficult to create spatial maps for each sector, especially in the form of grid-cell data. In general, there are a greater variety of spatial data available for primary sectors such as crops [157] and livestock [158]. Primary sectors account for most of the land use, water use, and other environmental impacts resulting from production, so these sectors receive more research attention. Also, the function of land for primary sectors is often unique, so it is generally easier for remote-sensing to identify. In contrast, identifying the distribution of secondary and tertiary sectors is much more challenging. For example, a particular building could be used as a residence, restaurant, school, a company, or several other uses. Many land classification schemes do not include factories, refineries, restaurants etc. It is this underlying inability to specify land use that causes much of the problem. Some environmentally important industries may still be possible to spatially identify, for example, transport and stationary power plants [159]. The phenomenon is especially evident in spatial distribution of carbon emission from EDGAR databases, which have detailed carbon emission for transport sectors, but much more coarse for manufacturing and service sectors [69]. Given this issue, the environmental impacts of most of industries are proportionally allocated into sectors based on their output. For example, for lack of sector-specific data Moran et al. mapped all sectors in an input-output table into 11-13 spatial maps of air pollution and greenhouse gas emission [95,96]. Similarly, for water, in Lutter et al. and Holland et al., WaterGAP consumption data was combined with an MRIO model (EXIOBASE for Lutter et al, and GTAP for Holland et al.), which was relatively straightforward to link in the case of agricultural and electricity sectors, but not directly possible in the manufacturing

sectors [50,113]. These difficulties in pinpointing secondary and tertiary sectors are a focus of ongoing research for water and energy modelers [71].

One way to solve this sectoral information problem is to construct a map with a detailed land use classification based on current high-resolution map data, for example, Google Earth or OpenStreet map, which can identify location of secondary and tertiary sectors precisely. It would then be possible to link the sectoral map with spatially explicit environmental models to create more accurate spatial distribution of environmental impacts for more sectors in input-output models. Multi-use buildings will remain a challenge for the foreseeable period; for example, one building may include resident households, restaurants, banks and other service. The phenomenon is particularly prominent in metropolis areas with high-density population and complex industrial structures. While some regions in some datasets have information on building-by-building use, the data is currently too patchy and limited for full integration into input-output models.

Spatial resolution: Compared to subnational statistical data, grid-cell data is not limited to administrative boundaries, and it has the possibility of depicting spatial variation more accurately. But spatial variation still depends on the area of the grid cell. The coarsest resolution used in the 48 papers reviewed this study was 2° lon × 2.5° lat (about 60,500 km^2 at the equator) [98], slightly larger than smaller countries such as Netherlands, Switzerland, Slovakia, and Belgium. While increasing the spatial resolution of databases may be important for SIO models, there is little the input-output practitioner can realistically do about this given that these models often result from large research campaigns, for example NASA Earth Observations (NEO)[160]. For ease of viewing we have presented a non-exhaustive selection of some common spatial databases that have the potential to be used in combination with input-output models in Table 2.3.

A key issue arises when looking at the spatial resolution of Category 1 studies. All the studies in our review use a proportionality assumption in assigning regions for production (which fulfills international demand). That is, all regions of production are treated the same whether products are used domestically or exported. This means that regions which do not have good access to markets and are likely producing goods for local consumption are 'counted' as part of the footprint of overseas consumers. Studies have suggested that regions with good transportation services and access to ports are more likely to be regions which export commodities [161,162]. It may be possible to use this fact to apply a first-order correction to which regions may be producing domestically or exporting goods. Regions where road density is highest could be used as the first-priority for export, with the remaining area as the first-priority for domestic production and consumption. Similarly, we can allocate environmental impacts in the same way. In some cases, the subnational trade data is directly available (for example, at the municipal level in Brazil) [163], but is difficult to implement globally due to data limitations.

Table 2.3. Potential spatial information sources to improve SIO models

Environmental impacts	Databases	Sectoral resolution	Temporal resolution	Spatial resolution
Land use and land cover	European Space Agency Climate Change Initiative [29,164]	Cropland for crop sectors; grassland for livestock;	Each year, from 1992 to 2015	300 m × 300 m, global

	Database		Year	Resolution
	MODIS land cover [30,76]	forestland for forest products; urban area for manufacturing and service sectors	Each year, data from 2001 to 2012	5'× 5' minute, global
	USGS Global Cropland Area Database (GCAD)[165]	Detailed cropland classification, including wheat, rice, maize, barley, soybean, cotton, orchards, sugarcane, cassava.	2010	1 km × 1 km, global
			2015	30 m × 30 m , global
			Annually 2003 to 2014	250 m × 250 m, Africa
			Annually 2000 to 2015	250 m × 250 m , Australia
			Annually 2001 to 2013	250 m × 250 m , USA
Water	Aqueduct Global Maps [166]	No specific mapping relationship with input-output databases, but can be combined with other spatial information, for example, crop distribution, power plants distribution, to create mapping relationship with input output databases	2010	Shape file by water basin, Global
	12 Global hydrological models (HDTM, Macro-PDM, MPI-HM, GWAVA, VIC, LaD, WaterGAP, PCR-GLOBWB, LPJmL, WASMODM, H08, ISBA-TRIP), details see [167]	Details for agricultural sectors and electricity sectors; difficult to combine with other manufacturing sectors	Varying, from hours to month	Varying, from 0.5° × 0.5° to 2° × 2°, Global
Air pollution, GHG	Emissions Database for Global Atmospheric Research (EDGAR) [69]	Varying from 7 to 28 sectors related to energy consumption.	Annually, 1970 to 2012	0.1° × 0.1°, Global
Pesticides	USGS, Grids of Agricultural Pesticide Use in the Conterminous United States [168]	All detailed crop sectors for input-output tables in US.	1992	1 km × 1 km

Biodiversity	Global Mammal Assessment [169]	Details see [59]	Annually, 2000 to 2050	1 km × 1 km, Global
	IUCN Red List [72]		Annually, 2009 to present	Shape file by hydrological basins for freshwater basins; by taxonomic groups (species) for territorial and marine animals
	BirdLife [73]		Annually, 2007 to present	Shape file by taxonomic groups (species)
Agriculture	Global Gridded Crop Model Intercomparison (GGCMI) [170]	All detailed crop sectors for input output databases.	Annually, 1979 to 2010	0.5° × 0.5°, global
	Spatial Production Allocation Model (SPAM) [171]		2005	5' × 5', global
Soil organic carbon	Food and agricultural organization (FAO) [172]	No specific mapping relationship with input-output databases, but can be combined with other spatial information, for example, crop distribution to create mapping relationship with input-output databases	2017	30" × 30", global
Electricity	US, environmental protection agency, Emissions & Generation Resource Integrated Database (eGRID) (USEPA, 2018)	Power generation sectors in input-output databases	Annually, 1996 to 2016	Point locations, global
	Global Energy Observation (http://globalenergyobservatory.org/)		Annually, 1950 to present	
	Platts (https://www.platts.com/)		Quarterly, from 1998 to present 2017	

Going beyond the data sets provided in the Table 2.3, the increasing number of monitoring stations provided by local authorities, such as those for air and water quality may also provide further data available for analysis [74]

Enhancing Category 2, the sectoral and spatial resolution of final demand.

Sectoral resolution: typically, commodity classification is more detailed than the products or sectors given in input-output models. However, the classification of consumer expenditure surveys is based on direct household consumption in mind, rather than economic sectors like those included in input-output models and so a conversion has to be applied. In addition, most categories within a consumer expenditure survey are food commodities. For example, out of the 183 commodities in the Norwegian database, 66 are food-related, but there are only 26 food related products in 200 products in high sectoral MRIO databases, EXIOBASE [110]. Additionally, surveys cannot distinguish domestically made or imported products consumed by households. Other import parts, such as government consumption expenditure, and gross fixed capital information still lack of research. Blockchain with IoT devices would be a good way to trace these final consumptions in the future [173].

Spatial resolution: From the final demand perspective, the spatial distributions of environmental footprints are generally performed at the local authority level – as described by Method 6 in the supporting information. This is mainly due to the lack of spatial distribution of consumption at any other resolution. Some scholars, for example, Moran, et al., applied global gridded population and local per-capita purchasing power databases to spatialize consumption-based environmental impacts [107]. Beyond that, Big Data methodologies have been suggested by various researchers as the possibility of collecting detailed human activities consumption with geolocation at a very high spatial (as well as sectoral) resolution [174].

Pathways for moving towards Category 3 SIO's.

In an ideal situation, efforts to enhance sectoral and spatial resolution ultimately leads to a Category 3 SIO database. That is, spatially explicit information on production processes and related extensions, spatially explicit information on consumption patterns, with intermediate transaction matrices that match this spatial and sectoral detail. Compiling the intermediate transaction matrix is extremely challenging when compared to compiling spatially explicit extensions or final demand. Some national statistical institutes may have detailed, sectoral, statistical data consistent with international standards [175]. But even in these cases it is extremely challenging to build spatially explicit input-output databases, since it requires a large amount of *in-situ* surveys. As discussed, studies that provide such transaction information generally do so using non-survey methods, leading to highly estimated transaction information.

The most common non-survey method for constructing intraregional input-output models is to compile subnational input-output models and then estimate interregional trade flows separately [145]. However, this approach requires the common assumption that regional production technologies and preferences of customers are similar to the national level [145]. Clearly this introduces uncertainties at the subnational level. Furthermore, interregional trade flows are usually estimated using a gravity model that assumes trade is only related to economic size and geographical distance of the producing and consuming regions [112,176] (other non-survey models, such as entropy and information models, neuronal network models, and behavior-based models can also be used [148]).

From above analysis, we find that most studies concentrate on the global or national level since national input-output models are readily available, and GMRIO models (e.g. EXIOBASE, WIOD, EORA, GTAP) have become increasingly available in recent years. Few studies focus at city level, due to a lack of official data. New technologies based on Big Data approaches and

blockchain may offer ways forward in the future. Blockchain is a shared, distributed ledger that protects records from deletion, tampering or revision. Some researchers have used distributed ledgers combined with IoT devices to trace food supply chains, from plantation to processing and to retailers [177]. Similarly, if any commodity is labeled with a unique code, it can be traced using advanced database approaches. Once a complete network of supply chains is constructed, it may be possible to use these data to build a transaction matrix for input-output models. The technology may reduce the cost of collecting transaction data, improve the efficiency and reliability of databases [173], and provide real-time information.

Balancing resolution and policy needs: Ideally, finer spatial and sectoral scales will reveal more spatial heterogeneity in environmental impacts and will be of increasing relevance to policy makers. However, if there are no (reliable) data or reasonable assumptions for downscaling some regions, and attempting to do so might introduce unquantifiable uncertainty. Additionally, there may be cases where the policy need does not require higher-resolution in the first place. From the papers reviewed herein, the resolution of the final result is almost always dependent on external spatial information beyond input-output models. For this reason, papers which fall under Category 2: the disaggregation of final demand, are focused on American, European, and Australian regions since they have more complete local consumption statistics and a high availability of household surveys. It's best if the spatial scale chosen relates to the policy relevance of the environmental impact findings. For example, water pollution is regarded as local environmental impact, but a river will run through many regions and countries, so local and regional water balances need to be considered as well as the linkages to trade through input-output tables [50]. For example, Lutter et al. [50] and Wang et al. [97] study fresh water at the spatial resolution of the water basin, which may be more helpful to inform general, sector-based policies for water extraction and pollution within a region. Conversely, greenhouse gas (GHG) emissions are well-mixed and is an impact suitable for analysis at the national scale. However, identifying the spatial distribution of GHG driven by consumption helps connect consumers with the impacts of their consumption [95]. This is the case for Kanemoto et al. where they develop a hotspot analysis of carbon footprints at a global resolution of $0.1° \times 0.1°$ [95].

A different approach is needed for other types of air pollution such as particulate matter, which is very much a local issue and most often driven by point source emission [96]. In addition, aerosols which remain in the atmosphere for several days, can easily diffuse to other regions [102]. Therefore, locating pollution sources and exploring the spatial distribution of emission diffusion embodied in trade is a more appropriate scale to help consumers participate in abating targeted air pollution. This also requires additional modelling of emission diffusion and a temporal resolution greater than the yearly average as commonly used in studies.

Increasing the spatial resolution of input-output models may also put pressure on increasing the temporal resolution. One of the major drivers of making an IO database spatially explicit is to examine the local impacts of resource availability or pollutant emissions, which can sometimes vary more temporally than spatially. Since input-output databases are annual aggregations of activity, this elides some of the seasonal complexities. For example, the availability for water used in the cooling of thermal power plants vary more through the year than across the nation [178]. Some level of temporal resolution may be possible simply by using time-explicit final demand vectors, however these data will first have to be collected by national or regional bodies. For example, emission transport models need time series data, (hourly, daily, weekly or monthly), since aerosols diffuse to other locations on the order of several days. Temporal issues may also impact uncertainties, a topic to which we turn next.

2.5 Addressing uncertainties

Underlying sources of uncertainty: General uncertainties for input-output models arise from the source statistical data, sector aggregation, and data allocation approaches [65,179–184]. For EEIO models, further issues with source data and assumptions about the density of environmental impacts also contribute to uncertainties [183,184]. SIO analyses add two further, related uncertainties: 1) Uncertainty in spatial databases themselves, and 2) uncertainty from spatial and sectoral aggregation.

With respect to uncertainty in spatial databases, this can vary depending on the type of source. For remote sensing, sensor quality, image generation, and processing techniques will drive uncertainty [185]. In local statistical data, uncertainty will be driven mostly by statistical methods [186]. Finally, for modelling approaches, input data, assumptions, and model methodology will all drive uncertainties. Furthermore, the spatial resolution in environmental impacts will, on its own, result in some uncertainty. For example, the resolution of WaterGAP is $0.5° \times 0.5°$ (about 50 km $\times$ 50 km at the equator), implying that water consumption is the same within a 50 km $\times$ 50 km region.

Uncertainty will also be introduced when aggregating spatial databases into regions matched with input-output databases, especially at the border between regions. Spatial databases often have to break down spatial information into different sectors, using assumptions which will further drive uncertainty. Often, we can resolve the spatial distribution of primary sectors (e.g. food crops and livestock) and some secondary sectors (large power plants, for example). But as mentioned above, most manufacturing and service sectors remain difficult to locate.

As we will see below, it is often hard to obtain a firm grasp of where the largest uncertainties may arise. In some cases, researchers have found it is likely that more uncertainty arises from additional pollutant modelling and not the input-output models themselves. For example, emission transport and health impact models have been found to have more uncertainty than the underlying input-output model [98,104].

Approaches for estimating uncertainty: Uncertainty analysis for EEIO modelling is already challenging given the diversity of data [187] and the model structure [188]. Approaches have been developed to estimate uncertainties [187,189–191] and perform sensitivity analyses [183,192,193], but there is still a lot more work to do to fully understand uncertainties. Given these existing difficulties, spatially explicit uncertainties add another layer of complexity. Given the variation of possible uncertainties, approaches such as Monte-Carlo simulations can be computationally prohibitive [100,104,107].

Still, some researchers have attempted to clarify uncertainties by narrowing down the number of uncertainties for sensitivity analysis. For example, Lin et al. ran over 10,000 Monte-Carlo simulations [98] for each type of air pollution in their study. Zhang et al. estimate overall uncertainty in SIO models by aggregating 4 sources of uncertainty, including uncertainty from air pollution (via the spatial database), uncertainty in the MRIO model, uncertainty from chemical transport model—GEOS-Chem model using Normalized Root Mean Square Deviation (NRMSD) method, and uncertainty from health impact model [104]. Lenzen et al., simulate standard errors of household factor multipliers, embodied factor multipliers and household expenditure, and then integrate all these parts of standard error into a total standard error estimate of the entire SIO model using Monte-Carlo simulations. In another example, Moran et al., employ a Monte-Carlo approach to build up range of alternative global Lorenz curves for carbon emissions [107]. These methods inherit approaches used in the uncertainty analysis of traditional input-output models [194].

These examples are all based on conventional Monte-Carlo simulations, extracting a large number of samples with assumed distributions, usually normal or log-normal. These simulations require an assumption that the extracted data are independent. Rodrigues et al. use a Bayesian approach to compare the uncertainties of independent sampling such as this [195], and find that this approach underestimates the uncertainty of results [195]. Future uncertainty analysis could expand this concept to include spatial data, since spatial data are often developed by incorporating the same underlying databases as those used in input-output models, resulting in non-independent errors.

2.6 Integration with other environmental models

Future options for SIO models may include integration with other environmental assessment models including technology-rich Integrated Assessment Models (IAMs), such as IMAGE [196], GCAM [197], AIM/CGE [198], MESSAGE [199], REMIND [200]. Generally, IAMs use macroeconomic models to downscale the world spatially into 10-30 aggregated regions [201], after which they are coupled to earth system models or environmental data using spatially explicit models. The environmental impacts are then downscaled to that resolution (Figure 2.2). IAMs have already been integrated with other spatially explicit land use models such as the CLUE-s model [202,203], the Global Land-use Model [204,205], and the Land Use Land Cover Change (LULCC) model [206], and these could be all be combined with input-output approaches. IAMs have also been used to make water demand spatially explicit [207,208]. For example, the LPJmL land model has been used to examine carbon balances [209], the dynamic GLOBIO model for evaluating biodiversity impacts [196], the integration of the GLOFRIS model for estimating impacts of flood risks [196], and the GISMO model for human development [196]. However, IAMs lack physical linkages between capital stock and material flows and they cannot trace the entire supply chain, this means IAMs are generally not able to assess environmental impacts other than on a production-basis. However, input-output models are an effective tool to assess impacts including those impacts embodied in the trade [210].

Generally, input-output models are constructed from historical data and used for historical analysis. Given their structure there are no built-in dynamic mechanisms. Conversely, technology-rich IAMs are used to project different scenarios for industrial structure, final demand, and spatial distribution of environmental impacts, which are the components of input-output models. It would be possible to establish soft links between IAMs and SIOs [210]. IAMs could be used to project components of input-output models needed to provide input-data to perform scenario-based consumption focused accounting [210–212].
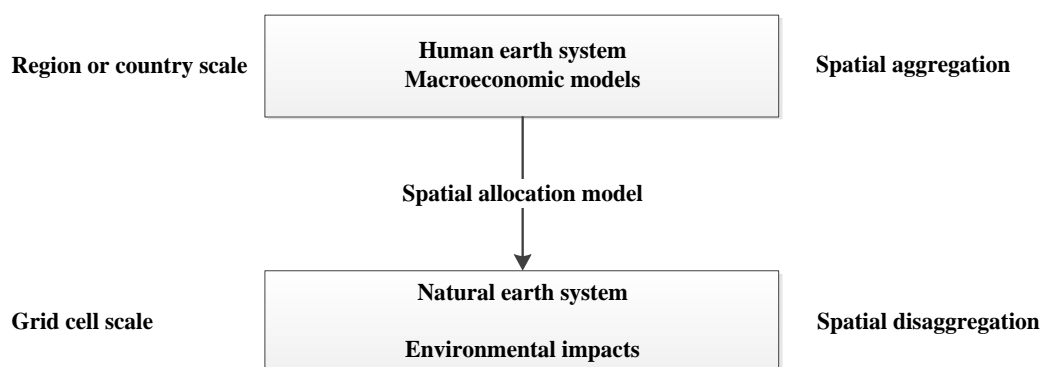

Figure 2.2 Schematic of general structure that integrates macroeconomic models with spatially explicit modes.

2.7 Outlook

Spatially explicit approaches inherit the advantages of EEIO, linking environmental impacts from production to consumption [65], while revealing the spatial variation of local environmental impacts. The recent growth in SIO is not necessarily surprising given the extensive role of globalization in outsourcing production and the associated environmental impacts of goods and services worldwide. Such SIO approaches can allow for a better understanding of the distribution of impacts from consumption, and provide data for targeted consumption-based mitigation measures.

We have critically reviewed recent SIO analyses and provided an overview of their methodologies and strengths. These analyses can be broadly separated into three approaches: 1) spatial disaggregation in environmental extensions, 2) spatial disaggregation in final demand, and 3) construction of a new input-output table with spatially disaggregated transaction matrices. We describe the considerations and issues that are raised when performing these analyses, and have presented an overview of specific findings. We have outlined the main challenges and limitations in present SIO modelling, including: the availability of spatially explicit data of different spatial and sectoral resolutions, the balancing of spatial resolution with research goals and policy advice, and the difficulty in assessing uncertainties. We also discuss the possibility of incorporating SIO modelling with integrated assessment models.

We expect that future efforts will focus on several key areas: as further spatial databases become available with greater sectoral resolution – especially in secondary and tertiary sectors – more options for deeper analysis and linkage with other environmental models will become possible; we see the opportunity of temporal analysis for certain resources, such as water, becoming increasingly tractable; and, studies that combine both demand-side (Category 1) and consumption-side (Category 2) disaggregation will become possible. A major hurdle in building accurate Category 3 input-output models at a high level of spatial detail is the lack of information about intermediate transactions and the structure of the value chains at this level of detail. Issues with uncertainties will likely remain problematic for some time, given the difficulties in assessing input-output model uncertainty even without spatial disaggregation. However, this is a problem which is not specific to input-output modelling, and is faced by many other large-scale environmental model approaches such as IAMs.

2.8 Acknowledgements

2.9 Supporting Information Available

This information is available free of charge via the Internet at http://pubs.acs.org