**Disaggregating input-output models with incomplete information**

SÖREN LINDNERa, JULIEN LEGAULTb, DABO GUANc\*[[1]](#footnote-1)

*a) Department of Land Economy, University of Cambridge*

*b) Department of Engineering, University of Cambridge*

*c) School of Earth and Environment, University of Leeds*

Disaggregating a sector within the Leontief input-output (IO) framework is not a straightforward task since there is more than one possibility for the unknown technical coefficients of the disaggregated IO table, and access to more information than embodied in the aggregated IO table is thus required. This paper presents a methodology for disaggregating sectors into an arbitrary number of new sectors when the only available information about the newly formed sectors is their output weights. A random walk algorithm is used to explore the convex polytope containing the admissible combinations for the unknown technical coefficients of the disaggregated IO table. The developed methodology is illustrated by disaggregating the electricity production sector of China’s 2007 IO table, and by looking at the CO2 emission intensity factors of all the sectors of the economy.

*Keywords*: Disaggregation; Input-output analysis; Electricity sector

# introduction

Information made available in national IO (input-output) tables on sectors and inter-industry flows is in most cases not detailed enough because accurate data collection of the exact production and output of an industry is rather difficult: it relies on comprehensive surveying of sales and purchase patterns of firms and companies. Due to time and data constraints, lack of institutional resources or insufficient cooperation between the surveying body and companies, similar sectors are often aggregated, resulting in merging all individual outputs into one aggregated output.

Whereas aggregation of some types of industries may have only a minor effect on the overall economy displayed in an IO table, the aggregation of certain types of sectors can have important consequences (see Wolsky, 1984; Fei, 1956; Su et al., 2010, Weber, 2009). Lenzen (2011) gives the example of the rice and wheat sectors being combined into one grain growing sector. Both sectors individually have different outputs (in monetary terms), and also different input requirements for water use. Thus, having both sectors aggregated into one may lead to under/overestimation of water use intensity for each individual sector. Another relevant example is the electricity sector, where CO2 emissions associated with a unit of output from coal are very different compared with a unit of output from renewable energy sources, but all generation units are generally combined in one sector in the input-output tables.

In the literature, the problem described above has been termed the “aggregation bias” problem (Morito, 1970; Kymn, 1990), and has been extensively discussed (see Miller and Blair 2009 for a summary). Gallego and Lenzen (2009) as well as Foran et al. (2005) have covered the aggregation mismatch of environmental satellite accounts in input-output tables, a problem that frequently occurs in environmental IO analysis, and proposed coping strategies.

Another way for IO practitioners to cope with aggregation bias is to disaggregate the sensitive sectors. However, disaggregating a sector into several new sectors within the Leontief IO framework is not a straightforward task, because there is a range of possibilities for unknown technical coefficients of the disaggregated IO table, and not only one possibility. Access to more information about the sector than embodied in the aggregated form is thus required, such as the knowledge of the exact input of each new sector stemming from the disaggregation into the other sectors of the economy (common sectors), and the input proportions of these common sectors into the new sectors.

When confidentiality prohibits surveying of industries and no other detailed data is available about the make-up of the sector that needs to be disaggregated, a simple estimate based on the weight output ratios of the new sectors (which are generally known or well estimated) can be used. However, the main drawback of this estimate is that all the inputs of the new sectors into the common sectors are in proportion to the weight factors, which may be fairly different from reality, especially for economies having strong regional disparity such as China or India. For instance, if most of the production sites of an industry are located in regions where the local energy mix has a higher percentage of fossil fuels than the national average, then this industry is probably consuming more energy from fossil fuels than the initial estimate will prescribe. The issue of spatial aggregation and its effect on emissions embodied in trade has been discussed by Su and Ang (2010).

Because the initial estimate is not the sole possibility for the disaggregated IO table, but only *one* of many possibilities, there will also be more than one possibility if the disaggregated IO table is later used for economic life-cycle analysis, where indirect and direct effects of sectors are investigated (see Marriott, 2007). From a practical point of view, the existence of multiple possibilities for the disaggregated IO table, and the lack of information that allows finding the possibility that actually corresponds to the real economy, raises various questions. For instance, what is the full range of possibilities for the disaggregated IO table? How can this range be systematically explored? How will changes in the disaggregated IO table affect the results of the life-cycle analysis? The aim of this paper is to provide tools that will help IO practitioners answer these questions. A general methodology for disaggregating sectors of IO tables is developed, which follows work done by Wolsky (1984) who presented the methodology for disaggregating one sector into two sectors. Here, Wolsky’s ideas are extended in two ways.

First, the disaggregation is generalized to an arbitrary number of new sectors *n*, and an additional constraint related to the fact that the final demand of the newly formed sectors cannot be negative is added. Second, a random walk algorithm is used to explore the full range of admissible values for the distinguishing parameters, i.e. the parameters that describe the level of departure of the disaggregated IO table from the initial estimate. To illustrate the method, the Chinese economy is used as a case study. The electricity production and distribution sector of China’s 2007 Input-Output table is disaggregated into three components: 1) hydro-electricity and others (including nuclear, solar, wind and biomass), 2) subcritical coal and 3) other fossil fuels (supercritical coal, ultra-supercritical coal (USC) and gas). An aggregated 12 × 12 version of the IO table is used for sake of conciseness. CO2 satellite factors for each of the newly formed electricity production sectors are then introduced (in grams of emitted CO2 released per kWh of electricity produced), allowing the analysis of the emission intensity factors for each industry sector (in grams of CO2 per RMB of final demand).

It should be noted that Lenzen (2011) recently showed the advantage of using a disaggregated IO table for environmental analysis, even when the disaggregation is based on limited information. This was done by comparing the relative standard error (RSE) associated with the environmental multiplier matrix of an aggregated system with its disaggregated counterpart. When performing the disaggregation, Lenzen assumed that no information regarding the individual total output or individual final demand of the newly formed sectors was fully certain, except of course that their sum must match the aggregated sector values. In the current study, it is assumed that the values of the total outputs of each of the newly formed sectors are available and are certain quantities, which may sometimes be the case, or may be reasonable to assume. This additional information adds complexity to the problem and requires an extended procedure to generate the constrained range of possibilities for the disaggregated IO tables. Therefore, whereas Lenzen’s study (2011) was primarily concerned with the comparison of the error related to an aggregated IO table versus a disaggregated IO table, the present study focuses solely on the exploration of the admissible solutions for the disaggregated IO table when partial, but certain information is available about the newly formed sectors. In this sense the study presented is also different from Gillen and Guccione (1990) and Su et al. (2010), who disaggregate sectors in IO tables with use of price data. Specificially Gillen and Guccione (1990) adress Wolsky’s method by presenting an alternative disaggregation method by using input prices, output prices for a period other than the base one for the n element vectors. If final demand and gross output of the aggregated sector is known then finding the the new intersectoral inputs and outputs is possible, and even exact if the additional data is taken from a year close to the base one. Like Lenzen (2011), Su et al. (2010) note that input-output data and environmental data are often not in the same aggregated format. In China, sectoral CO2 emissions are often derived from energy consumption data,and this dataset is published by the National Bureau of Statistics (NBS) in a 44 sector format, whereas Chinese IO tables are in a more disaggregated format. Su et al (2010) show, using the example of analysing the CO2 emissions embodied in trade of China for an economy with various sector size, how with use of additional information on electricity consumption by sectors and electricity price energy data is disaggregted to the size of the IO tables and be superior to the approach where energy consumption data is uniformly distributed to each sector according to weight factors. Focus of this paper is on exploring the full range of possible solutions for disaggagregation, and one case where disaggregation is performed with only very basic knowledge (ie.: only output weights but no further information on input or output prices) is presented. Thus, it is not the aim to present a disaggregation of an economic sector that is necessarily as accurate as possible, but merely to provide an analytical tool with which a full range of possibilities for disaggregation can be explored. In the conclusion other approaches to disaggregation which make use of a range of additional information are briefly discussed.

# THEORETICAL BACKGROUND

## Input-output models

### Leontief framework

Consider an economy with *N*+1 sectors where each sector *i* produces a unique good. The total output of good *i* from the *i*th sector is noted *xi* and the amount of good *i* that sector *j* consumes from sector *i* is noted *zij*. The total output *xi* corresponds to the sum of the intermediate consumption by the economy and the final demand *fi*

 , for *i* = 1 to *N* + 1. (1)

In the input-output Leontief framework, it is assumed that the industry flow from sector *i* to sector *j* depends linearly on the total output of sector *j*. If sector *j* needs *aij* units of good *i* to produce 1 unit of good *j*, Eq. (1) can be rewritten as

 , for *i* = 1 to *N* + 1, (2)

which leads to the following matrix representation

 , (3)

where **A** is the technical coefficient matrix of the economy. Inverting this system lead to

 , (4)

where **I** is the identity matrix of size *N*+1 × *N*+1 and **L** the Leontief inverse matrix. The *ij*th coefficient in the inverse Leontief matrix **L** represents the total requirement of sector *i*’s production to meet the final demand of sector *j*. In life-cycle analysis, these coefficients are used to determine the direct and indirect requirements on all sectors of the economy associated with the final demand of each sector (Miller and Blair, 2009).

## Disaggregation

### Problem statement

Let the technical coefficient matrix **A**\* describe the same economy as **A**, with the only difference that the last sector of the economy (sector *N*+1) has been disaggregated into *n* distinct sub-sectors. Matrix **A**\* is thus of size *N*+*n* × *N*+*n*. The total output of sector *i* in the disaggregated economy is noted *xi*\* and the final demand *fi*\*. The *N* sectors that were not disaggregated (*xi*\* = *xi* and *fi*\* = *fi* for *i* = 1 to *N*) are referred to as the “common sectors” while the sub-sectors originating from the disaggregated sector are referred to as the “new sectors”. The number of technical coefficients associated with the aggregated sector in matrix **A** is 2*N*+1. These coefficients correspond to the *N* coefficients associated with the input to common sectorsfrom the disaggregated sector, to the *N* coefficientsassociated with input to the disaggregated sector from the common sectors and to the intra-industry input/output coefficient of the disaggregated sector. Because the common sectors in matrix **A**\* are unchanged,

 , for *i*,*j* = 1 to *N*. (5)

In matrix **A**\*, the remaining 2*Nn* + *n*2 technical coefficients associated with the new sectors cannot be constrained straightforwardly like Eq. (5). These coefficients are the *Nn* coefficients associated with the input of common sectors into the new sectors, the *Nn* coefficients associated with the input of the new sectors into the common sectors and the *n*2 coefficients associated with input of the new sectors into themselves. The core challenge of the disaggregation consists in attributing a value to these remaining 2*Nn* + *n*2 coefficients, knowing that they will be related to the 2*N*+1 coefficients associated with the aggregated sector in matrix **A** by a set of constraints. The following section describes these constraints.

### Constraints

Let *wk* be the weight ratio of the total output of the *k*th new sector to the total output of the disaggregated sector (*wk*­ = *xN*+*k\**/*xN*+1). Since the total output produced by the disaggregated sector must be conserved,Σ*wk* = 1. When a sector is disaggregated, the total outputs of the new sectors are almost always known and hence the weights *wk* are also known. If the weights *wk* are known, the conservation of the amount of goods consumed by the common sectors from the disaggregated sector, by the disaggregated sector from the common sectors and by the disaggregated sector from itself leads to the following 2*N* + 1 constraints

 , for *i* = 1 to *N*, (6)

 , for *i* = 1 to *N*, (7)

 . (8)

The constraints of Eqs. (5) to (8) were given by Wolsky (1984). An aspect that was not explicitly considered by Wolsky is the constraints associated with the final demand of the new sectors. In fact, as long as the choice for the technical coefficients of the new sectors respects Eqs. (7) and (8), the total final demand of the disaggregated sector will *necessarily* be conserved. However, under certain conditions, the final demand of an individual new sector may go below zero if its range is not explicitly constrained. Gillen and Guccione (199) implicitly state this constraint in the definition given for intermediate inputs (U). For instance, take the case of a two-sector disaggregation where the technical coefficients of the second new sector are all set to zero. This means the total output *xN*+2\* of the second new sector will go straight to the final demand, and thus the final demand of the first new sector will be equal to *fN*+1\* – *xN*+2\*, which may be negative (depending on *fN*+1 and *wN*+2). To avoid this outcome (the final demand of each new sector must remain positive), the following *n* “final demand” constraints can be explicitly added to the problem:

 , for *i* = 1 to *n*. (9)

Dividing Eq. (9) by *xN*+1 leads to

 , for *i* = 1 to *n*, (10)

where *bi* is weight ratio of the final demand of the *i*th new sector to the total final output of the disaggregated sector, i.e. *bi* = *fN*+*i*\*/*xN*+1. These final demand ratios are introduced as additional variables to the problem and similar to the technical coefficients, they must remain positive. The *n* constraints described in Eq. (10) add an important but necessary complexity to the problem: they couple the unknown technical coefficients of all the bottom part of the **A**\* matrix together (i.e. rows *N*+1 to *N*+*n*). This means that the technical coefficients associated with the consumption of a given sector from the new sectors can be modified independently of the same coefficients for the other sectors, but only as long as the final demand can absorb the variations without becoming negative.

In total, there are 2*Nn* + *n*2 + *n* unknowns (2*Nn* + *n*2 unknown technical coefficients and *n* unknown final demand ratios) and 2*N* + *n* + 1 constraints. Therefore, there will be 2*Nn* + *n*2 – 2*N* – 1 free parameters. These free parameters are the so-called the “distinguishing parameters”. They describe the space of admissible solutions for the 2*Nn* + *n*2 + *n* unknowns, subject to the constraints of Eqs. (5) to (8) and (10). In section 2.24, it will be shown how they can be derived. For simplicity, the following notation is henceforth adopted: *Nu* is the number of unknowns, *Nc* the number of constraints and *Nd* the number of distinguishing parameters (*Nd* = *Nu* – *Nc*).

### Initial estimate

Following Wolfsky (1984), an initial estimate can be made for the unknown technical coefficients and the final demand ratios by assuming that the new sectors have identical technologies and that they supply the other sectors proportionally to their output weights *w*. This estimate corresponds to the initial estimate mentioned in the Introduction. It is described by the following set of equations:

 , for *i* = 1 to *N*, (11)

 , for *k* = 1 to *n* and *i* = 1 to *N*, (12)

 , for *k* = 1 to *n*, (13)

 , for *k* = 1 to *n*. (14)

In this paper, attention is focused on the technical coefficients associated with the consumption of the common sectors and the new sectors from the new sectors (Eqs. (12) to (14)). Therefore, it is assumed that the coefficients associated with the consumption of the new sectors from the common sectors (Eq. (11)) will stay fixed on the initial estimate. However, the tools presented in this paper could also be applied to deal with the deviation of the latter coefficients from the initial estimate.

Having fixed the unknown coefficients associated with the consumption of the new sectors from the common sectors, the number of unknowns *Nu*, of constraints *Nc* and of distinguishing parameters *Nd* are now *Nn* + *n*2 + *n*, *N* + *n* + 1 and *Nn* + *n*2 – *N* – 1, respectively. The *Nu* unknowns can be regrouped and arranged in the column vector **yu** as follow

 , (15)

where T is the transpose operator. The vector **y**containing the initial coefficients is noted **y0** and is given by

 , (16)

where **w** = [*w*1,*w*2,...,*wn*] is the vector of the new sector weights.

### Distinguishing parameters

The equality constraints in Eqs. (6) to (8) and (10) can be expressed in the following matrix form:

 . (17)

where **C** is the constraint matrix of size *Nc* × *Nu* and **q** the vector of constraint constants of size *Nc*:

 . (18)

The constraint matrix **C** has the following structure:

 , (19)

where **In** is the identity matrix of size *n* and where the matrices **C11**, **C22, C31** and **C32** are of size *N* × *Nn*, 1 x *n*2, *n* × *Nn* and *n* × *n*2, respectively. These matrices are given by:

 , , , , (20)

where **1n** is a line vector of size *n* containing ones. Deriving the *Nd* distinguishing parameters consists in finding the vectors that describe the subspace orthogonal to the subspace described by the *Nc* constraints in matrix **C**. This can be done using the Gram-Schmidt algorithm. First, a vector of size *Nu* with random entries is created using a random number generator. An orthogonal projection of this vector is then made on each of the line vectors of matrix **C** and the orthogonal component is kept each time. The resulting vector is divided by its norm which gives the first distinguishing vector of the distinguishing basis. A second random vector is generated and is projected on the line vectors of matrix **C**, but also on the first distinguishing vector. The remaining orthogonal component is normalized which gives the second distinguishing vector. This procedure is repeated *Nd* times to build the distinguishing matrix **D** of size *Nu* × *Nd*. The columns of matrix **D** correspond to the orthonormal distinguishing vectors. These vectors are orthogonal to one another and to all the constraint vectors in matrix **C**. It should be noted that before generating the distinguishing vectors, the line vectors of the constraint matrix **C** must also be made orthogonal using this procedure. The *Nu* unknowns can now be expressed in function of the initial estimate **y0** and of the distinguishing matrix **D**

 , (21)

where **d** is the vector containing the *Nd* unknown distinguishing parameters *di*

 . (22)

The fact that the technical coefficients and the final demand ratios must be positive means that the distinguishing parameters are bounded by the following inequality constraints

 . (23)

The admissible values for the distinguishing parameters prescribed by the inequality constraints of Eq. (23) are contained within a convex polytope of dimension *Nd* delimited by *Nu* facets of dimension *Nd* – 1. In other words, this polytope describes the set of all admissible values for the unknown coefficients of the disaggregated IO table. To alleviate the terminology, the polytope is simply called the “solution space” of the disaggregated IO table. It should be noted that for a disaggregation with two new sectors (*n*=2), closed-form expressions were given by Wolfsky (1984) for the distinguishing parameters and their associated bounds.

## Exploring the solution space

The simplest technique to generate points uniformly in the polytope described by Eq. (23) and thus to explore the solution space for the unknown coefficients of the disaggregated IO table is to use rejection sampling. One way to implement rejection sampling is to find the smallest hyper-sphere that circumscribes the polytope, to generate points uniformly in that sphere and to reject points that fall outside the polytope. However, this procedure will be very costly for high dimensional problems (*Nd* > 10) due to the high rejection rate (curse of dimensionality). A technique that is frequently employed to circumvent this problem is the use of Markov chain random walks. In this paper, the random walk algorithm proposed by Kannan and Narayanan (2009) was chosen for that purpose. This algorithm was implemented using MATLAB.

For each sample generated in the polytope with the random walk algorithm, the technical coefficient matrix **a** can be reconstructed and the inverse Leontief matrix of the disaggregated economy can be computed. The computation of Leontief inverse can be made efficient by means of matrix partitioning (Wolsky, 1984). Hence, for each generated sample, only a matrix of size *n* × *n* needs to be inverted instead of a matrix of size (*N* + *n*) × (*N* + *n*).

# Case study: disaggregating the energy sector of China

## Preliminary considerations

The disaggregation procedure described in the previous section is illustrated by using an aggregated version of the 2007 Chinese national IO table (National Bureau of Statistics of China, 2008). In its original form, the IO table was of size 135×135, but in order to be able to present the numerical results in a concise manner, the table was aggregated up to a 12×12 format by using a concordance matrix (see Peters et al., 2006). The final table contains the agricultural sector, the three primary energy sectors (coal mining and processing, crude petroleum products, gas production and distribution), eight other sectors and the electricity sector. This table and its sector description is given at table 1.

 As mentioned in the Introduction, the electricity sector is disaggregated into three components: 1) hydro-electricity and others, 2) subcritical coal and 3) other fossil fuels. The disaggregated IO table is thus of size 14×14, and there are 42 unknown technical coefficients in that table. Even though this case study is not used for policy analysis and recommendation, but only for illustration of the developed methodology, it is not an arbitrary choice because the electricity sector of China is generally of high importance for environmental analysis. China has just surpassed the US in being the number one emitter of total emissions in the world, and the electricity sector of China alone contributes to 70% of these emissions. Because of its newly built fleet of coal-fired power plants, China will, in the nearby future, continue to be largely dependent on coal for generating electricity. As stated in various papers (Urban et al., 2009; He et al., 2008; Guan et al., 2008), the electricity sector is seen as having key potential for carbon reduction of China’s economy. Therefore, in order to harmonize Chinese IO tables for effective environmental IO-analysis, it is useful to disaggregate the electricity production and distribution sector into its main generation components.

In order to perform the analysis of embodied CO2 emissions per RMB of final demand for each sector, the monetary units of the electricity sector first needs be converted in kWh. This can be done by using the mean price *p* of electricity in China, which was taken from the Chinese electricity yearbook and was on average 0.8 RMB/kWh in 2007 (CEY, 2008). The disaggregated monetary Leontief matrix can then be multiplied by the CO2 satellite account **e** to compute the vector of CO2 emission intensity factors **ε**:

 , (24)

where the CO2 satellite account **e** is a row vector of size *N* + *n* whose first *N* components are zero and whose *n* last components are the satellite emission factors *ei* of the newly formed electricity production sectors, divided by mean the energy price *p*, i.e. **e** = [**0**,*e*1/*p*,*e*2/*p*,...,*en*/*p*]. The satellite emission factor *ei* quantifies the embodied emissions of CO2 per kWh of produced electricity from the new sector *i*. Table 2 presents the numerical values used in the present case study for these factors. The latter were taken from (Nsakala and Marion, 2001) and are given in grams of emitted CO2 per kWh of produced electricity. The emission intensity factors *εi* will thus be in grams of emitted CO2 per RMB of final demand. Note that the combined sector “hydro-electricity and others” has a factor of 30 grams of emitted CO2 per kWh, which is a weighted average of all generation units combined in this sector (the satellite emission factor for the “other fossil fuels” sector is also a weighted average). In fact, most of the CO2 emitted from the “hydro-electricity and others” sector comes from Biomass power plants.

In table 2, the installed capacities (GW) of each generation type for the year 2007 are also given. Here, it is assumed that the output weights *w* of the energy generation sectors *in monetary terms* correspond directly to these capacity weights. These weights are then used to form the initial estimate of the disaggregated table as described in section 2.2.3. Using the weights of the output capacity as the weights of the sector outputs in monetary terms is a rough guess that assumes uniform sale prices of electricity across all generation types in the six regional grids of China, and no private contracting between industries and electricity companies. It also assumes that the energy production over the year was made in proportion to that capacity, i.e. that each electricity generation sector produced a share of the total produced energy that is proportional to its capacity proportion. Finally, it assumes that the transmission and distribution sector is merged with the energy generation sectors, and thus it disregards the potential effects of this intermediate sector. In reality, this would only hold true if each generation unit had its own unique network, which is obviously not the case.

In their respective studies, Limmeechokchai and Suksuntornsiri (2007), Allan et al. (2005) and Shresthe and Marpaung (2005) also disregarded the potential impact of the transmission and distribution sector. On the other hand, Cruz (2002) and Allan et al. (2006) conducted studies in which the disaggregation of the electricity sector was made so that that all the output of the power generation sectors was first sold to the transmission and distribution sector, and then all the other sectors purchased electricity from this sector. Here, this approach is not used because the main objective of the case study is to illustrate the disaggregation methodology, and because very limited information was available to the authors concerning the transmission and distribution sector.

 Table 1: Input-Output table of China (2007), in billion RMB

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Ag** | **CmP** | **Pp** | **Fm** | **Ppc** | **Ch** | **Msp** | **M+e** | **Gp+d** | **Co** | **T+w** | **Ep+d** | **ID** | **FD** | **TO** |
| **Ag** | 687.7 | 7.0 | 0.8 | 2223.1 | 0.0 | 167.6 | 0.7 | 66.4 | 0.0 | 25.9 | 255.0 | 0.0 | 3434.4 | 1420.5 | 4854.9 |
| **CmP** | 2.7 | 97.0 | 5.7 | 37.1 | 112.0 | 193.5 | 122.7 | 22.7 | 7.1 | 5.7 | 25.5 | 330.2 | 961.9 | 41.4 | 1003.3 |
| **Pp** | 0.6 | 1.3 | 114.8 | 11.0 | 1189.4 | 442.2 | 933.4 | 29.3 | 55.7 | 83.5 | 17.5 | 36.8 | 2915.4 | 62.3 | 2977.8 |
| **Fm** | 482.2 | 15.7 | 25.0 | 3813.9 | 15.8 | 326.7 | 98.6 | 370.1 | 3.3 | 171.3 | 1368.1 | 27.5 | 6718.2 | 4675.6 | 11393.9 |
| **Ppc** | 39.4 | 13.6 | 89.2 | 46.2 | 121.4 | 463.0 | 298.4 | 83.7 | 3.4 | 126.7 | 771.3 | 127.5 | 2183.9 | 145.5 | 2329.4 |
| **Ch** | 379.8 | 27.1 | 122.8 | 885.2 | 48.0 | 3176.6 | 250.9 | 1098.6 | 7.4 | 1579.0 | 758.9 | 15.5 | 8349.7 | 1189.9 | 9539.6 |
| **Msp** | 14.6 | 69.3 | 86.6 | 136.6 | 10.3 | 228.8 | 2972.3 | 2684.5 | 4.7 | 1208.8 | 109.4 | 17.3 | 7543.1 | 1085.9 | 8629.0 |
| **M+e** | 58.6 | 98.0 | 197.2 | 307.8 | 50.1 | 339.4 | 683.5 | 6359.0 | 8.4 | 531.9 | 1331.4 | 295.0 | 10260.1 | 8754.1 | 19014.2 |
| **Gp+d** | 1.1 | 1.7 | 9.2 | 17.6 | 4.9 | 29.8 | 17.8 | 17.7 | 9.5 | 3.0 | 40.1 | 9.3 | 161.6 | 64.9 | 226.5 |
| **Co** | 1.1 | 1.3 | 1.4 | 2.6 | 1.2 | 2.7 | 2.1 | 3.5 | 0.2 | 59.8 | 123.1 | 1.0 | 200.0 | 6018.7 | 6218.7 |
| **T+w** | 309.7 | 129.5 | 189.0 | 917.1 | 130.9 | 787.8 | 570.3 | 1366.1 | 27.1 | 942.5 | 3873.2 | 278.2 | 9521.5 | 10119.7 | 19641.1 |
| **Ep+d** | 45.8 | 60.2 | 174.7 | 171.0 | 48.3 | 436.4 | 367.9 | 214.1 | 25.0 | 82.7 | 276.1 | 1129.4 | 3031.7 | 241.8 | 3273.5 |

**Abbreviations**: Ag = Agriculture, CmP = Coal mining and processing, Pp = petroleum processing and natural gas products, Fm = food manufacturing and tobacco products, Ppc = petroleum processing and coking, Ch = chemicals, Msp = Metals smelting and pressing, M+e = Machinery and equipment, Gp+d = gas production and distribution, Co = construction, T+w = transport and warehousing, Ep+d = Electricity production and distribution, ID = intermediate demand, FD = final demand, TO = total output

Table 2: Satellite emission intensities and national production capacity (China, 2007)

|  |  |  |  |
| --- | --- | --- | --- |
| **Sector** | **Emission intensity (grams of CO2 per kWh)** | **Production capacity (GW)** | **Percentage of total production capacity** |
| Hydro-electricity and others (nuclear, wind, biomass, etc.) | 30 | 160.8 | 24.1% |
| Subcritical coal | 1100 | 433.6 | 65.8% |
| Other fossil fuels (supercritical coal, USC coal and gas) | 830 | 74.3 | 11.1% |

## Numerical results

### IO table coefficients: initial estimate vs. random walk samples

As mentioned in section 2.3, the solution space for the disaggregated IO table is explored using the random walk algorithm proposed by Kannan and Narayanan (2009). Table 3 compares the technical coefficients associated to the consumption of the electricity sector (bottom part of the technical coefficient matrix) for five cases: the aggregated table coefficients, the disaggregated table coefficients of the initial estimate and the disaggregated table coefficients of three samples obtained with the random walk algorithm. A last column showing the ratio of the final demand divided by the total output of the electricity sector is also given. As expected, the sum of each common sector column for the disaggregated cases matches the aggregated coefficients, confirming that the constraints for the input of the new sectors into the common sectors are respected (see Eq. (7)). The conservation of the final demand is also respected (see Eq. (10)). For the input of the new sectors into themselves, the weighted sum of the disaggregated column sums matches the aggregated coefficient, i.e 0.345 (see Eq. (8)). The inverse Leontief matrix (total requirements matrix) associated with cases 2 and 5 are shown in the appendix. One can note the difference of the individual input requirements from the new sectors between these two cases.

Table 3: Technical coefficients of the electricity sector

|  |
| --- |
| Case 1: aggregated table coefficients |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Ep+d |  |  | FD/*xN*+1 |
| Ep+d | 0.009 | 0.06 | 0.059 | 0.015 | 0.021 | 0.046 | 0.043 | 0.011 | 0.110 | 0.013 | 0.014 | 0.345 |  |  | 0.074 |
| Case 2: disaggregated table coefficients, initial estimate |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Hy+O | SubC | O-FF | FD/*xN*+1 |
| Hy+O | 0.002 | 0.014 | 0.014 | 0.004 | 0.005 | 0.011 | 0.010 | 0.003 | 0.027 | 0.003 | 0.003 | 0.083 | 0.083 | 0.083 | 0.018 |
| SubC | 0.006 | 0.039 | 0.038 | 0.010 | 0.013 | 0.030 | 0.028 | 0.007 | 0.071 | 0.009 | 0.009 | 0.223 | 0.223 | 0.223 | 0.048 |
| O-FF | 0.001 | 0.007 | 0.007 | 0.002 | 0.002 | 0.005 | 0.005 | 0.001 | 0.012 | 0.001 | 0.002 | 0.038 | 0.038 | 0.038 | 0.008 |
| Sum | 0.009 | 0.060 | 0.059 | 0.015 | 0.021 | 0.046 | 0.043 | 0.011 | 0.110 | 0.013 | 0.014 | 0.345 | 0.345 | 0.345 | 0.074 |
| Case 3: disaggregated table coefficients, sample 1 (random walk algorithm) |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Hy+O | SubC | O-FF | FD/*xN*+1 |
| Hy | 0.003 | 0.043 | 0.016 | 0.010 | 0.001 | 0.013 | 0.003 | 0.003 | 0.020 | 0.010 | 0.002 | 0.022 | 0.010 | 0.333 | 0.026 |
| SubC | 0.001 | 0.016 | 0.036 | 0.004 | 0.011 | 0.022 | 0.038 | 0.008 | 0.070 | 0.001 | 0.009 | 0.021 | 0.356 | 0.344 | 0.038 |
| O-FF | 0.005 | 0.001 | 0.007 | 0.001 | 0.009 | 0.011 | 0.001 | 0.000 | 0.020 | 0.003 | 0.002 | 0.017 | 0.026 | 0.011 | 0.009 |
| Sum | 0.009 | 0.060 | 0.059 | 0.015 | 0.021 | 0.046 | 0.043 | 0.011 | 0.110 | 0.013 | 0.014 | 0.060 | 0.392 | 0.688 | 0.074 |
| Case 4: disaggregated table coefficients, sample 2 (random walk algorithm) |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Hy+O | SubC | O-FF | FD/*xN*+1 |
| Hy | 0.006 | 0.020 | 0.023 | 0.010 | 0.005 | 0.001 | 0.002 | 0.000 | 0.030 | 0.005 | 0.011 | 0.041 | 0.038 | 0.032 | 0.044 |
| SubC | 0.002 | 0.019 | 0.033 | 0.003 | 0.006 | 0.031 | 0.039 | 0.009 | 0.072 | 0.005 | 0.003 | 0.112 | 0.132 | 1.570 | 0.027 |
| O-FF | 0.001 | 0.021 | 0.003 | 0.002 | 0.010 | 0.014 | 0.002 | 0.002 | 0.008 | 0.004 | 0.000 | 0.029 | 0.013 | 0.041 | 0.003 |
| Sum | 0.009 | 0.060 | 0.059 | 0.015 | 0.021 | 0.046 | 0.043 | 0.011 | 0.110 | 0.013 | 0.014 | 0.182 | 0.183 | 1.643 | 0.074 |
| Case 5: disaggregated table coefficients, sample 3 (random walk algorithm) |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Hy+O | SubC | O-FF | FD/*xN*+1 |
| Hy | 0.000 | 0.004 | 0.032 | 0.003 | 0.002 | 0.002 | 0.009 | 0.000 | 0.024 | 0.001 | 0.002 | 0.328 | 0.002 | 0.466 | 0.016 |
| SubC | 0.007 | 0.025 | 0.022 | 0.007 | 0.008 | 0.039 | 0.032 | 0.010 | 0.013 | 0.010 | 0.010 | 0.207 | 0.000 | 1.220 | 0.056 |
| O-FF | 0.002 | 0.032 | 0.005 | 0.005 | 0.011 | 0.005 | 0.002 | 0.001 | 0.073 | 0.001 | 0.002 | 0.008 | 0.021 | 0.101 | 0.002 |
| Sum | 0.009 | 0.060 | 0.059 | 0.015 | 0.021 | 0.046 | 0.043 | 0.011 | 0.110 | 0.013 | 0.014 | 0.543 | 0.024 | 1.787 | 0.074 |

**Abbreviations**: Hy+O = Hydro + others, SubC = Subcritial Coal, O-FF = Other Fossil Fuels

### Inverse Leontief coefficients

In table 3, only three samples obtained with the random walk algorithm were given. With the ensemble of samples generated by the random walk, it is however possible to construct the full distribution of each coefficient of the inverse Leontief matrix, i.e. their probability density function (PDF). Figure 1a and 1b shows the PDFs of the inverse Leontief coefficients associated with the requirement of “hydro-electricity and others” and “subcritical coal” from the gas production and distribution sector, i.e. the entries (12,9) and (13,9) of the disaggregated inverse Leontief matrix. The initial estimate value is also shown for comparison. Using 106 samples, convergence is achieved for the PDFs (increasing the number of samples beyond this point makes very little difference on the global shape of the PDFs).

Because the random walk algorithm maps the solution space uniformly (see Kannan and Narayan, 2009), i.e. that each point in the solution space polytope is assigned an equal weight when the number of samples becomes very large, the PDFs shown in Fig. 1 are widely spread around the initial estimate value. Assigning an equal weight to all solutions (uniform distribution) in fact corresponds to the maximum entropy distribution when only the bounds of the uncertain variables (here the technical coefficients) are known. According to the principle of maximum entropy (Jaynes, 1957), this is thus the distribution that is the most non-committal given that the bounds are the only available information.

Yet, even though the wide range observed in Fig. 1 can provide useful information to IO practitioners, such as the maximum and minimum admissible values for the coefficients of the inverse Leontief matrix, it is arguable that solutions which are closer to the initial estimate in the solution space should be given more weight, because unless the transaction patterns of the economy are very heterogeneous, the actual disaggregated IO table should not be too far from the initial estimate. This additional information should be included into the problem, which will necessarily narrow the PDFs of the inverse Leontief coefficients around the initial estimate. However, the question of how it should be done is not straightforward, because the additional weight given to the solutions closer to the initial estimate has to be linked (in some way) to the degree of heterogeneity of the transaction patterns of the economy. This task is thus deferred for future work.

A last point that needs to be mentioned concerning the PDFs shown in Fig. 1 is that these PDFs are *marginal* densities and therefore, the correlations that exist between the coefficients of the inverse Leontief matrix will not appear explicitly. For instance, if the requirement of hydro-electricity from the gas production and distribution sector is high, then the requirement of fossil fuels from that sector is likely to be low and vice-versa. To study these correlations more specifically, the correlation matrix of the coefficients of the inverse Leontief matrix can be used.

### CO2 emissions intensities

With the knowledge of the inverse Leontief matrix, the emission intensities *ε* can now be computed according to Eq. 24. Figure 2 compares the emissions intensities of the common sectors (in grams of emitted CO2 per RMB of final demand) yielded by the initial estimate with the emission intensities yielded by the three samples shown in table 3. The bar chart shows that even though the total amount of CO2 embedded in the economy stays fixed (because the output weights of the electricity production sectors are fixed), the individual sector intensities can vary. Therefore, when the total output weights of the new sectors are fixed, the choice of disaggregated IO table does not affect the overall emissions of an economy, but it does matter for the intensity of the individual sectors. Depending on the choice of the unknown technical coefficients of the disaggregated IO table, the emission intensity factors of some sectors will be higher, while emission factors of other sectors will necessarily be lower. These findings are consistent with the conclusion in Su et al. (2010) that sector aggregation of sectors could affect not only their own sector results but also other sectors’ results.

Figure 3 shows the full PDF of the emission intensity factor of the gas production and distribution sector. Similar to the PDF of the inverse Leontief coefficients (see previous section), this distribution was constructed by attributing an equal weight to the 106 samples yielded by the random walk algorithm. Interestingly, it is seen that the emission intensity factor of the gas production and distribution sector can be more than twice that of the initial estimate. Even though this scenario is very unlikely (the associated probability is very small), this important difference shows how sensitive the emission intensity factors can be depending on the choice of unknown coefficients of the disaggregated IO table.

(b)

(a)

Fig. 1. Probability density function (PDF) of the inverse Leontief coefficients associated with the electricity requirements of the gas production and distribution sector. (a) Electricity requirement from hydro-electricity and others. (b) Electricity requirement from subcritical coal.



Fig. 2. Emissions intensity factors of common sectors: initial estimate vs. samples 1 to 3.



Fig. 3. Probability density function (PDF) of the emission intensity factor for the gas production and distribution sector.

# Conclusion

In this paper, a methodology was presented to explore the range of admissible solutions for the disaggregation of a sector of the IO table. This range can provide useful information to IO practitioners such as the maximum and minimum admissible values of the coefficients of the inverse Leontief matrix, or their full probability density function when all the admissible combinations for the unknown coefficients of the disaggregated IO table are assigned an equal probability. This range can also be compared with an initial estimate of the disaggregated IO table based solely on the output weights of the newly formed sectors. Hence, when no information is available about the newly formed sectors besides their output weights, the IO practitioner can make a better judgement of the potential variability of the inverse Leontief coefficients.

To illustrate the developed methodology, the electricity of China’s 2007 input-output table was disaggregated into three components: 1) hydro-electricity and others (nuclear, wind, biomass, etc.), 2) subcritical coal and 3) other fossil fuels (supercritical coal, USC coal and gas). The inverse Leontief matrix was multiplied with an environmental satellite account to calculate the emission intensity factors of each sector of the economy (in grams of emitted CO2 per 1 RMB of output). It was shown that although the choice of the unknown coefficients of the disaggregated IO table does not affect the overall emissions of an economy, it affects the emission intensities of the individual sectors, and can lead to emission intensity factors that are twice the value of the initial estimate. This shows that the choice of information used for disaggregation can have a strong effect on in emission intensities of individual sectors and that use of specific and reliable additional information on the new sectors ought to be considered whenever possible. For disaggregating the electricity sector in China additional information can be effectively used to make the disaggregation as realistic as possible. For example, output price data (electricity prices) and commodity input prices can generally be obtained from the NBS so that a disaggregation according to the methodological approach by Gillen and Guccione (1990) can be attempted. Also, CO2 emission data for China can be derived by using energy consumption data and the IPCC reference approach for converting energy data int emisssions. In this case Su et al. (2010) have shown that price data for electricity and specific sub-sector electricity consumption data are helpful elements to make a disaggregation more exact.

There are more refinements in future studies that could be done for the specific case of the electricity sector. For one, the number of newly formed sectors should be increased in order to refine the emission intensity analysis. A disaggregation level of 4-6 according to the full range of electricity generation options present in China, should be optimal. For studies aiming to forecast the fuel- and power generation mix in China and its effect on CO2 emissions this level of disaggregation may become relevant because it allows to split the electricity production sector into new efficient technologies like coal fired power stations with integrated gasificiation combustion cycle (IGCC), or natural gas power plants. Then, the initial estimate for the disaggregated IO table should be refined because in its current form, this estimate assumes a uniform spread of industries in China. Therefore, it does not consider the fact that some industries are regionally clustered and that China has six independent electricity transmission and distribution grids, each having a specific electricity generation mix that deviate from the national average. In reality, industries are likely to consume a specific electricity mix based on the magnitude of their relative presence in one of the six electricity grids. A choice for initial estimate that considers the above mentioned factors is likely to provide a more accurate picture for the individual electricity consumption mix of industries, and thus their related emissions intensities.

Acknowledgments

The authors thank Xin Li and two reviewers for their valuable comments

References

Allan, G. J., McGregor, P. G., Swales, J. K., and Turner, K. R. The ‘hypothetical extraction’ of electricity generation activities fromScotland: an input–output analysis. Paper presented at EPSRC SuperGen Marine Publicworkshop, 2 December 2005, University of Strathclyde, Glasgow

Allan, G. J., McGregor, P. G., Swales, J. K. and Turner K. R. 2006. Impact of alternative electricity generation technologies on the Scottish economy: an illustrative input–output analysis. Journal of Power and Energy, Special Issue.

Chinese Energy Statistics Yearbook (CEY, 2008). Chinese National Bureau of Statistics (2008).

Cruz, L., 2002. Energy-Environment-Economy Interactions: An Input-Output Approach Applied to the Portuguese Case. Paper for the 7th Biennial Conference of the International Society for Ecological Economics, “Environment and Development: Globalisation & the Challenges for Local & International Governance”, Sousse (Tunisia), 6–9 March 2002.

Fei, John Ching-Han," A Fundamental theorem for the Aggregation Problem of Input-Output Analysis," Econometrica 24 (1956), 400-412.

Gallego, B. and Lenzen, M. 2009. Estimating generalized regional input-output systems: A case study of Australia. In: The dynamics of regions and networks in industrial ecosystems.

Gillen, W.J., Guccione, A., 1990. Disaggregating input-output models: An alternative to Wolsky’s method. Economic Systems Research 2 (1), 39-42.

Guan, D., Hubacek, K., Weber, C. L., Peters, G. P., Reiner, D.M. 2008. The drivers of Chinese CO2 emissions from 1980 to 2030. Global Environmental Change 18 (2008) 625-634.

Foran, B., Lenzen, B., Dey, C., Bilek, M. 2005. Integrating sustainable chain management with triple bottom line accounting. Ecological Economics 52 (2005) 143– 157.

He, J., Deng, J., Su, M., 2010. CO2 emission from China’s energy sector and strategy for its control. Energy 11 Volume 35, 4494 - 4498.

Jaynes E.T., Information theory and statistical mechanics, Physical Review Letters, 1957, vol. 106 (4), pp. 620-630, vol. 108 (2), pp. 171-190

Kannan R., Narayanan H., Random walk on polytopes and an affine interior point method for linear programming, *Proceedings of the ACM Symposium on Theory of Computing*, 2009

Kymn, K. O., 1990. Aggregation in Input-Output Models: a Comprehensive Review, 1946 – 1971. Economic Systems Research, 2, 65 -93.

Lenzen, M. 2011. Aggregation versus disaggregation in Input-Output analysis of the environment. Economic Systems Research, 2011, Vol. 23(1), March, pp. 73–89.

Limmeechokchaia, B. and Suksuntornsirib, P. 2005. Assessment of cleaner electricity generation technologies for net CO2 mitigation in Thailand. Renewable and Sustainable Energy Reviews 11 (2007) 315–330.

Marriot, J. 2007. An Electricity-focused Economic Input-output Model: Life-cycle Assessment and Policy Implications of Future Electricity Generation Scenarios. PhD Thesis dissertation, Carneggie Mellon University.

Morimoto, Y., 1970. On Aggregation Problems in Input-Output Analysis. Review of Economic and Statistics, 37, 369 -383.

National Bureau of Statistics of China. 2008. Chinese Provincial and National Input-Output tables.

Nsakala N., Marion J. 2001. Controlling Power Plant CO2 emissions. A long range view. Alstom Power Technology Centre. Working paper.

Peters, G., Weber, C., Liu, J. 2006. Construction of Chinese Energy and Emissions Inventory. NTNU, working paper.

Shresthaa, R. M. and Marpaung C. O. P. 2005. Integrated resource planning in the power sector and economy-wide changes in environmental emissions. Energy Policy 34 (2006) 3801–3811.

Su, B.,Huang, H.C., Ang, B.W., Zhou. P. 2010**.** Input–output analysis of CO2 emissions embodied in trade: The effects of sector aggregation. Energy Economics 32 (2010) 166–175.

Su, B., Ang, B.W., 2010. Input-output analysis of CO2 emissions embodied in trade: The effects of spatial aggregation. Ecological Economics 70 (1), 10-18.

Urban, F., Benders, R.M.J., Moll, H.C., 2009. Renewable and low carbon energies as mitigation options of climate change for China. Climatic Change (2009).

Weber, C.L., 2009. Measuring structural change and energy use: Decomposition of the US economy from 1997 to 2002. Energy Policy 37 (4), 1561-1570.

Wolsky, A. 1984. Disaggregating Input-Output Models. The Review of Economics and Statistics, Vol. 66, No. 2 (May, 1984), pp. 283-291.

Appendix

Table A-1: Inverse Leontief, initial estimate

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Hy | FF | others |
| Ag | 1.220 | 0.035 | 0.016 | 0.376 | 0.020 | 0.067 | 0.026 | 0.036 | 0.026 | 0.052 | 0.061 | 0.025 | 0.025 | 0.025 |
| CmP | 0.014 | 1.133 | 0.024 | 0.026 | 0.078 | 0.062 | 0.053 | 0.025 | 0.073 | 0.037 | 0.017 | 0.187 | 0.187 | 0.187 |
| Pp | 0.033 | 0.058 | 1.089 | 0.050 | 0.602 | 0.152 | 0.242 | 0.081 | 0.322 | 0.132 | 0.055 | 0.087 | 0.087 | 0.087 |
| Fm | 0.213 | 0.079 | 0.046 | 1.609 | 0.058 | 0.132 | 0.080 | 0.096 | 0.079 | 0.131 | 0.164 | 0.074 | 0.074 | 0.074 |
| Ppc | 0.033 | 0.050 | 0.056 | 0.044 | 1.098 | 0.110 | 0.093 | 0.048 | 0.061 | 0.087 | 0.071 | 0.092 | 0.092 | 0.092 |
| Ch | 0.190 | 0.112 | 0.106 | 0.270 | 0.114 | 1.576 | 0.143 | 0.193 | 0.128 | 0.482 | 0.131 | 0.086 | 0.086 | 0.086 |
| Msp | 0.038 | 0.187 | 0.094 | 0.081 | 0.086 | 0.118 | 1.615 | 0.366 | 0.107 | 0.396 | 0.065 | 0.110 | 0.110 | 0.110 |
| M+e | 0.075 | 0.253 | 0.165 | 0.145 | 0.162 | 0.180 | 0.284 | 1.610 | 0.187 | 0.282 | 0.177 | 0.302 | 0.302 | 0.302 |
| Gp+d | 0.002 | 0.004 | 0.005 | 0.005 | 0.006 | 0.008 | 0.006 | 0.004 | 1.047 | 0.005 | 0.004 | 0.007 | 0.007 | 0.007 |
| Co | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.003 | 1.013 | 0.009 | 0.003 | 0.003 | 0.003 |
| T+w | 0.161 | 0.267 | 0.147 | 0.260 | 0.192 | 0.254 | 0.236 | 0.230 | 0.272 | 0.351 | 1.322 | 0.267 | 0.267 | 0.267 |
| Hy | 0.012 | 0.035 | 0.030 | 0.020 | 0.029 | 0.037 | 0.038 | 0.020 | 0.058 | 0.028 | 0.014 | 1.140 | 0.140 | 0.140 |
| FF | 0.032 | 0.093 | 0.080 | 0.054 | 0.077 | 0.100 | 0.102 | 0.055 | 0.156 | 0.075 | 0.038 | 0.377 | 1.377 | 0.377 |
| Others | 0.005 | 0.016 | 0.014 | 0.009 | 0.013 | 0.017 | 0.018 | 0.009 | 0.027 | 0.013 | 0.007 | 0.065 | 0.065 | 1.065 |

Table A-2: Inverse Leontief, sample 3 (random walk algorithm)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Ag | CmP | Pp | Fm | Ppc | Ch | Msp | M+e | Gp+d | Co | T+w | Hy | FF | others |
| Ag | 1.220 | 0.036 | 0.016 | 0.376 | 0.020 | 0.067 | 0.026 | 0.036 | 0.029 | 0.052 | 0.061 | 0.031 | 0.018 | 0.059 |
| CmP | 0.014 | 1.140 | 0.025 | 0.028 | 0.082 | 0.062 | 0.052 | 0.025 | 0.093 | 0.037 | 0.018 | 0.230 | 0.133 | 0.437 |
| Pp | 0.033 | 0.061 | 1.090 | 0.051 | 0.604 | 0.152 | 0.242 | 0.080 | 0.331 | 0.132 | 0.055 | 0.107 | 0.062 | 0.203 |
| Fm | 0.214 | 0.082 | 0.047 | 1.610 | 0.059 | 0.132 | 0.079 | 0.096 | 0.087 | 0.131 | 0.164 | 0.090 | 0.052 | 0.172 |
| Ppc | 0.033 | 0.053 | 0.057 | 0.045 | 1.100 | 0.110 | 0.093 | 0.047 | 0.071 | 0.087 | 0.071 | 0.113 | 0.066 | 0.215 |
| Ch | 0.190 | 0.116 | 0.107 | 0.271 | 0.116 | 1.576 | 0.143 | 0.193 | 0.137 | 0.482 | 0.131 | 0.106 | 0.062 | 0.202 |
| Msp | 0.038 | 0.192 | 0.095 | 0.082 | 0.088 | 0.117 | 1.614 | 0.365 | 0.119 | 0.396 | 0.065 | 0.135 | 0.078 | 0.257 |
| M+e | 0.076 | 0.265 | 0.167 | 0.148 | 0.168 | 0.180 | 0.283 | 1.609 | 0.219 | 0.282 | 0.177 | 0.372 | 0.215 | 0.707 |
| Gp+d | 0.002 | 0.005 | 0.005 | 0.005 | 0.006 | 0.008 | 0.006 | 0.004 | 1.048 | 0.005 | 0.004 | 0.009 | 0.005 | 0.016 |
| Co | 0.002 | 0.004 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.003 | 1.013 | 0.009 | 0.003 | 0.002 | 0.006 |
| T+w | 0.161 | 0.278 | 0.150 | 0.262 | 0.197 | 0.254 | 0.234 | 0.229 | 0.301 | 0.351 | 1.323 | 0.328 | 0.190 | 0.623 |
| Hy | 0.010 | 0.047 | 0.063 | 0.025 | 0.052 | 0.029 | 0.045 | 0.017 | 0.126 | 0.027 | 0.015 | 1.525 | 0.032 | 0.846 |
| FF | 0.034 | 0.114 | 0.063 | 0.058 | 0.076 | 0.105 | 0.094 | 0.055 | 0.186 | 0.074 | 0.039 | 0.384 | 1.062 | 1.668 |
| Others | 0.007 | 0.047 | 0.011 | 0.016 | 0.024 | 0.018 | 0.012 | 0.008 | 0.099 | 0.012 | 0.008 | 0.036 | 0.033 | 1.184 |

1. Corresponding Author: dg346@cam.ac.uk [↑](#footnote-ref-1)