The GTAP Database as a Large Sparse Multi-Regional Input-Output Table

Rodrigues, João F. D.[*] Marques, Alexandra[†]
Domingos, Tiago[‡]

January 12, 2012

Abstract

The GTAP database is often used in Input-Output Analysis to generate a Multi-Regional Input-Output table (MRIOT) of the world, usually in dense format. In this paper we show how to generate a sparse MRIOT from GTAP and compare the computational requirements of the sparse and dense tables in terms of processing requirements and calculation of multipliers.

KEYWORDS: Multi-regional Input-Output table (MRIOT); GTAP database; sparse matrix; dense matrix.

1 Introduction

Multi-Region Input-Output tables (MRIOT) have been widely applied to study the environmental repercussions of human activities [Wiedmann, 2009]. This type of analysis requires MRIOTs that describe the trade relations between all sectors of all countries of interest in a given year. In a globalized world where production and consumption processes are spatially disconnected, this is a very important feature because it allows tracking environmental pressures through global supply chains. The main difficulty associated with these models is the lack of consistent and reliable source data.

---

[*] Technical University of Lisbon, Instituto Superior Técnico, IN+, DEM, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal. E-mail: joao.rodrigues@ist.utl.pt.

[†] Technical University of Lisbon, Instituto Superior Técnico, IN+, DEM, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal. E-mail: alexandra.marques@ist.utl.pt.

[‡] Technical University of Lisbon, Instituto Superior Técnico, IN+, DEM, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal. E-mail: tdomingos@ist.utl.pt.

At the moment there are several running and recently finished projects whose goal is to provide full global databases for MRIOT analysis, such as EXIOPOL [Tukker et al., 2009], WIOD (www.wiod.org), AISHA [Lenzen et al., 2010] and AIIOT [IDE, 2006]. The GTAP project periodically releases a global database, with country-level input-output (IO) tables, trade data and environmental extensions, designed to be used in Computable General Equilibrium models [Narayanan and Walmsley, 2008]. The GTAP database comprises a set of interconnected tables describing economic transactions between economic agents, in a number of prices and with differing degrees of aggregation. The GTAP database is very detailed in the agriculture sector and in the tax and subsidy structure. The fact that it provides already balanced and harmonized data that is periodically updated makes the GTAP database a very popular data source for MRIOT studies (for example, Peters and Hertwich [2008], Hertwich and Peters [2009], Davis and Caldeira [2010], Rodrigues et al. [2010], Peters et al. [2011b]), although that is not its purpose [Narayanan and Walmsley, 2008].

The problematic components in the conversion of the GTAP database to a global MRIOT are international commodity trade and international transportation, because of underdetermination in the GTAP tables. A recent paper by Peters et al. [2011a] shows how the database can be converted into a full dense MRIOT by allocating international commodity trade and international transportation using trade shares.

For example, regarding bilateral trade GTAP has information on the import (and export) of sector i from region r to region s but it is not known which sector j in the importing region actually receives the good or service. To solve this indeterminacy Peters et al. [2011a] allocate bilateral exports (to industries and final demand) according to the import shares of the importing region. To deal with international transportation data the authors present two options, an exogenous and an endogenous international transport. In the model with exogenous international transport, the provision of international transportation services is considered a final demand category of the supplying region (for example, Hertwich and Peters [2009], Davis and Caldeira [2010]). In the model with endogenous international transport, the provision of international transport is allocated to the importing region, with the product of several trade shares. First of all it is assumed that the supply of international transport is allocated evenly among suppliers in proportion to their contribution to the international transport pool, then the use of transport margins for each commodity is allocated in proportion to the use of imports by each sector [Peters et al., 2011a].

In the present paper we present a more parsimonious approach to the problem of converting the GTAP database to a global MRIOT table. We suggest a method that generates a MRIOT with the same properties as the endogenous model of Peters et al. [2011a] with no algebraic data manipulation to calculate trade share products and minimal processing and data storage requirements. Instead of attempting to build a dense IO matrix with detailed transactions between domestic firms of different regions, we build a sparse matrix [Golub and Van Loan, 1996] that describes the transactions between intermediate firms (domestic, import, export and transport), which correspond to quantities directly available in the GTAP database.

In Section 2 we discuss the motivation and the approach taken in this paper. In Section 3 we describe the structure of the GTAP database and of the corresponding sparse IO table. In Section 4 we compare the performance of the construction and analysis of sparse and dense IO tables derived from GTAP. Section 5 concludes.

2 Motivation

The construction of a MRIOT often involves the integration of data with different degrees of aggregation [Oosterhaven et al., 2008], where the most common situation is a table of internationally traded commodities (i.e., the exports of given sector from a given region to all sectors of another region), and a set of national tables indicating the imports of every sector by commodity type but not by source country (i.e., the imports of a given sector from a given country of all a given commodity type imported from all countries). This is the situation that occurs in the GTAP database, with some additional complications resulting from international tariffs and trade margins.

However, the MRIOT requires the specification of the trade that takes place between every two sectors from every two countries, an information that is not available from source data and must therefore be estimated. The common approach to this problem is to use the so-called trade share methods in which an aggregate quantity (the international trade in a commodity class between two regions) is multiplied by a fraction of imports (a trade share), in such a way that the aggregate quantity is proportionately distributed among disaggregate transactions. This is the approach followed by Peters et al. [2011a] for the construction of a MRIOT from GTAP.

In this paper we propose a different approach to this problem of incomplete information. Instead of increasing the number of nonzero entries in the MRIOT, we suggest the introduction of intermediate sectors, in such a way that the number of nonzero entries in the MRIOT corresponds to transactions that are available from the source data.

We believe that a simple example will clarify the general approach. Consider a closed economy with two sectors, $n = 2$, where $\mathbf{Z}$, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{v}$ are the matrix of intersectoral transactions and the vectors of total output, final demand and primary inputs. Now consider that intersectoral transactions are unknown while the sums in rows, $\mathbf{z}^R$, and in columns, $\mathbf{z}^C$, are known, and that the sum of all intersectoral transactions is z^T. Using the trade share method we produce a dense matrix:

$$\mathbf{Z} = \begin{bmatrix} z_1^R z_1^C/z^T & z_1^R z_2^C/z^T \\ z_2^R z_1^C/z^T & z_2^R z_2^C/z^T \end{bmatrix}.$$

However, we can also introduce an intermediate sector that receives all intersectoral outputs and delivers all intersectoral outputs, such that the entire intersectoral matrix is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & z_1^R \\ 0 & 0 & z_2^R \\ z_1^C & z_2^C & 0 \end{bmatrix},$$

and it is necessary to concatenate an entry z^T to $\mathbf{x}$ and zeros to $\mathbf{y}$ and $\mathbf{v}$. Using this formulation sector 1 exports z_1^R to the intermediate sector, from which it receives z_1^C. Since $z_1^R + z_2^R = z^T$ it is possible to discriminate the total imports of sector 1 z_1^C as $z_1^C z_1^R/z^T$ and $z_1^C z_2^R/z^T$, and notice that these terms are exactly the intersectoral inputs of 1 in the dense matrix.

The trade-share method implies algebraic manipulation while the intermediate-sector method implies a topological transformation of the IO system. However, the multipliers calculated using either of the above models will yield the same results.

The first of the above methods (trade-share method) yields a dense matrix, with n^2 nonzero entries and involves algebraic manipulation of the data. The second method (introduction of an intermediate sector) yields a sparse matrix, with n nonzero entries whose values can be obtained directly from the source data.

If there are only two sectors, the two methods are comparable, but as the number of sectors increases the computational requirements of the trade-share method quickly become unmanageable, while the intermediate-sector method allows for a much more compact data storage without any loss of information.

It must be emphasized that data storage here does not simply mean space in the hard disk but also space in active memory. Therefore, the range of MRIOT applications that are feasible in a personal computer are greatly extended by using the second method.

This simple example is a toy model only. The structure of the GTAP database and of MRIOTs in general is more complex. However, the main technique that will be applied throughout this paper is fully illustrated in the example above: intermediate sectors are introduced in the system in such a way that transactions that are known from source data appear as entries in the MRIOT.

In this paper we follow closely the work of Peters et al. [2011a], which provide a template for the construction of a dense MRIOT from the GTAP database. In that paper two models are presented, an exogenous model in which international trade data from GTAP is aggregated and an endogenous model in which the full GTAP data is used. We will use their endogenous model as a benchmark to compare the performance of the sparse MRIOT. However, it is trivial to aggregate the GTAP international trade data to obtain a sparse model that is equivalent to their exogenous model. We did not pursue that line of inquiry because the focus of the present work is the comparison of the dense and sparse formulations of MRIOTs.

3 Data and Methods

The GTAP database consists of a set of interconnected tables displaying transactions between several economic agents with differing degrees of agregation. Our approach to build a MRIOT from GTAP is to introduce intermediate firms, such that every entry in the GTAP database can be mapped directly to the MRIOT as a transaction between two such firms. We will not get into detail in the structure of the GTAP database – more information can be found in Brockmeier [1996], Narayanan and Walmsley [2008] and Peters et al. [2011a]. In the following paragraphs we will emphasize only the components that are relevant for the problem at hand.

The GTAP 7.1 database has $n_S = 57$ *domestic* industry sectors per region and $n_R = 112$ world regions, $n_F = 3$ *final demand* sectors (households, government and investment), $n_P = 5$ *primary inputs* (or endowment) sectors (several types of labour and capital) and several types of taxes and subsidies. We will use the term *firm* to refer to a sector that does not belong to final demand or endowments. For the purpose of IO analysis, all taxes and subsidies represent transactions between a firm and the government, and therefore we consider only a single sector of *net taxes*, which we classify together with endowments (note that a net tax can be negative, while all other transactions are assumed positive). We shall use symbols DF, Y, V to denote *domestic firm*, *final demand* and *endowments* (including taxes), respectively, and use $\rightarrow$ to represent the flow of a good or service. If a production

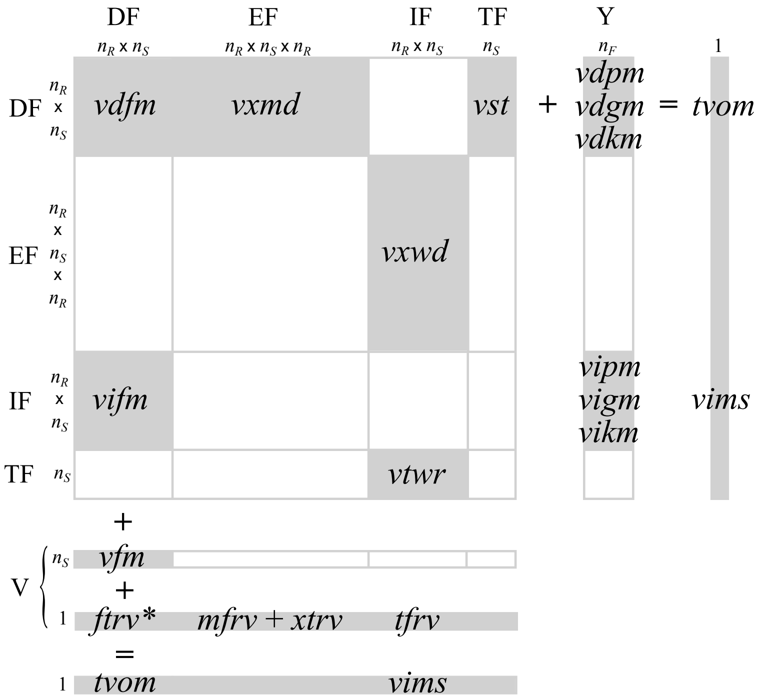

Figure 1: GTAP 7.1 tables and transaction classes in the sparse MRIOT.

chain has only domestic inputs and its product is consumed domestically, the full chain consists of only three classes of transactions: intermediate inputs, $DF \rightarrow DF$ (reported in GTAP table *vdfm*); primary inputs, $V \rightarrow DF$ (table vfm and $ftrv^*$); and final consumption, $DF \rightarrow Y$ (tables *vdpm*, *vdgm* and *vdkm*, where *vdkm* is the component of sales to gross capital formation in $vdfm$). The taxes paid by DF are provided in several tables. In Figure 1 the tax $ftrv^*$ is an aggregation of several GTAP tables, and is equal to tables: $ftrv - (fbep + isep + osep)$.

To account for international trade we shall consider additional intermediary firms of *exports*, EF, *imports*, IF, and *international transportation*, TF, such that each original GTAP table entry corresponds to a transaction between two intermediate firms and/or the other agents present in domestic transactions. If both final demand and intermediate inputs are allowed to be imported, additional transactions classes must be considered: exports, $DF \rightarrow EF$ (table *vxmd*); export duties/subsidies, $V \rightarrow EF$ (tables $mfrv$ and *xtrv*); provision of international transportation, $DF \rightarrow TF$, (table *vst*); imports, $EF \rightarrow IF$ (table *vxwd*); import duties/subsidies, $V \rightarrow IF$ (table $tfrv$); payment of international transport, $TF \rightarrow IF$ (table *vtwr*); sales of imports to final demand, $IF \rightarrow Y$ (tables *vipm*, *vigm* and *vikm*) and sales of

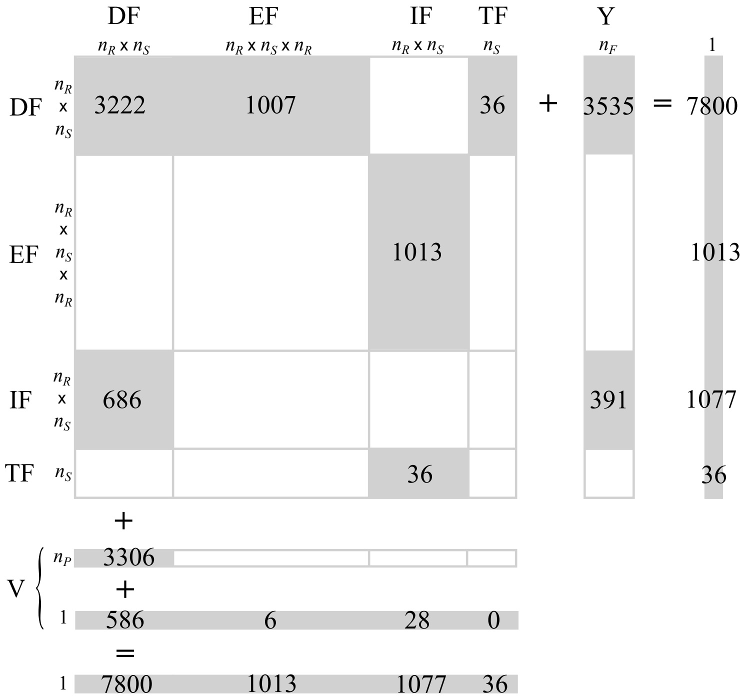

Figure 2: Global values of transaction classes in the sparse MRIOT (10^9 USD 2004).

imports to domestic firms, $IF \rightarrow DF$ (table $vifm$). The total output/input of domestic firms is provided by table $tvom$. Figure 1 presents the correspondence between the GTAP tables and transaction classes in the sparse MRIOT, while Figure 2 shows the global value of each transaction class.

Figure 3 illustrates the structure of each transaction class, considering a hypothetical aggregation of two regions and three sectors. Let n_D, n_E, n_I, n_T be the number of sectors in each firm class (domestic, export, import, and transport, respectively), such that the total number of firms is $n_Z = n_D + n_E + n_I + n_T$. The dimension of each firm class is $n_D = n_R n_S$, $n_E = n_R^2 n_S$, $n_I = n_R n_S$ and $n_T = n_S$. The dimension of the entire, sparse inter-industry matrix, n_Z^2, is larger than the dimension of the corresponding dense matrix, n_D^2 . However, the number of non-zero entries is substantially smaller.

Let n^* denote a number of non-zero values. At a macro level the interindustry matrix, Z, is sparse since only 6 out of 4×4 transaction classes are not empty. In the transaction classes $DF \rightarrow DF$ and $IF \rightarrow DF$, only the diagonal blocks, each $n_S n_S$, are not empty, corresponding to domestic transactions and totaling $n_{DD}^* = n_{ID}^* = n_R n_S^2$ non-zero values.

In the GTAP formulation, it is known how much of a given commodity

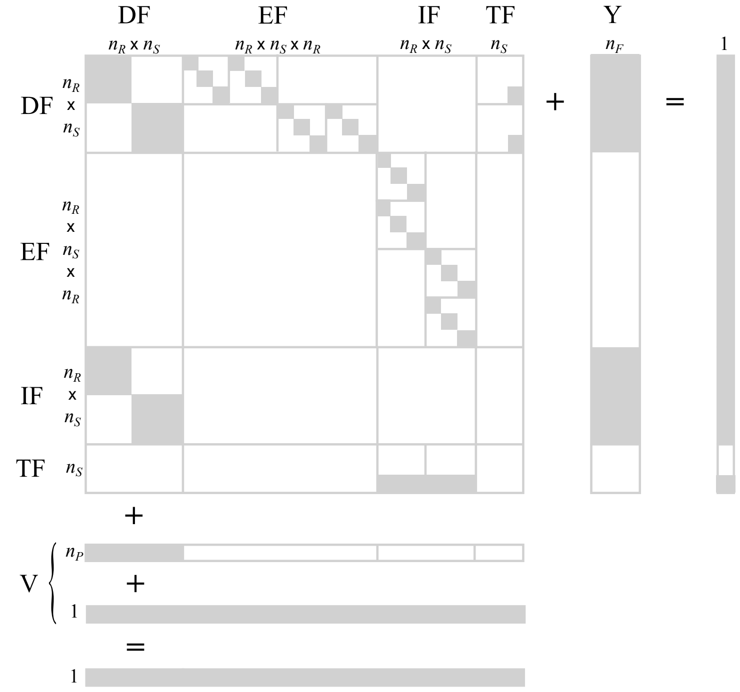

Figure 3: Structure of transaction classes in the sparse MRIOT.

is traded between any two regions, as well as the taxes and trade margins associated with each such transaction (tables $vxmd$, $vxwd$, $mfrv$, $xtrv$ and $tfrv$). However, it is not known how much of an imported product (tables $vifm$, $vipm$, $vigm$ and $vikm$) comes from any particular region. This can be conceptualized as there being an exporting sector from each region specialized in delivering commodities to any region (and paying export taxes), and an importing sector that bundles together the commodities received from all regions before delivering to either final demand or domestic firms (and paying import taxes). That is, in the transaction classes $DF \rightarrow EF$ and $EF \rightarrow IF$, there are n_R^2 blocks, and inside each only the diagonal line, of size n_S, is not empty, totaling $n_{DE}^* = n_{EI}^* = n_R^2 n_S$ non-zero values.

In the GTAP database, international trade is provided by three sectors (air, water and land transport), and the margin paid for the transaction of some commodity (table $vtwr$) is known but not the providing region. (For clarity, in Figure 3 only one international trade sector is represented.) However, it is known how much each region contributes to this global transport pool (table vst). Therefore, transaction class $DF \rightarrow TF$ contains n_R blocks, in each of which only three entries are not empty, and transaction class $TF \rightarrow IF$ contains n_R blocks, in which $3n_S$ entries are not empty. The total of non-zero values in each transaction class is therefore $n^*_{DE} = 3n_R$ and $n^*_{EI} = 3n_R n_S$.

The total of non-empty entries in the inter-industry matrix is $n^*_Z = 2n_R n_S^2 + 2n_R^2 n_S + 3n_R(1 + n_S)$. In the case of the GTAP 7.1 database this means that the number of non-zero elements is 5% of the number of entries in the corresponding dense matrix. In the case of final demand and primary inputs (not considering net taxes) only the blocks $DF \to Y$, $IF \to Y$ and $V \to DF$ are not empty. The vectors of total output and net taxes are dense.

The sparse MRIOT obeys the conventional identities, $\mathbf{Z1} + \mathbf{Y1} = \mathbf{x}$ and $\mathbf{Z}'\mathbf{1} + \mathbf{V1} = \mathbf{x}$, where all vectors are in column format and $\mathbf{1}$ is a unitary vector of appropriate size and $\mathbf{Z}$, $\mathbf{Y}$ and $\mathbf{V}$ are the matrices of inter-industry transactions, final demand and primary inputs, and $\mathbf{x}$ is the vector of total output.

A matrix is sparse or dense depending on the proportion of its entries which are zero. An $n \times m$ matrix with r nonzero entries is said to be in *sparse format* if only the r triplets containing the location and value of each nonzero entry are actually stored. There are different implementations for sparse format storage, for more information see Golub and Van Loan [1996]. Both the GTAP database and the sparse MRIOT presented here are stored in a sparse format.

All the processing and analysis were performed using the open-source GNU Octave software, and we made extensive use of its sparse algebra libraries, but other high-level languages such as MatLab or R have the equivalent functionalities. All computations were performed on a desktop computer with a 2.6 GHz dual-core CPU and 4 GB of RAM.

4 Results

We compared the computation time and data storage requirements of the construction of the sparse and dense MRIOTs. The sparse MRIO was compiled from GTAP tables following the method described in Section 3. The dense MRIOs are the exogenous and endogenous models described by Peters et al. [2011a]. In both cases, we started from a processed version of the GTAP database, which had been converted from the original .hrx (through .har and .gdx) to ASCII format [Rodrigues et al., 2010]. The resulting data structures were saved in compressed gzip format.

The dense MRIOs required circa 400 MB of disk space while both the sparse MRIO and the original GTAP database require circa 40 MB of disk space. Table 1 reports the time required to compile the MRIO using the different methods. The compilation of the dense MRIOs is substantially

	Compilation	Calculation					
		Direct	Iterative				
			1%	0.1%	0.01%	0.001%	0.0001%
Exogenous	1002.82	86.87	81.18 (8)	81.19 (13)	81.57 (17)	81.73 (22)	82.37 (26)
Endogenous	1863.39	109.13	83.79 (9)	83.80 (13)	83.96 (17)	84.55 (22)	84.81 (26)
Sparse	9.88	105.99	5.86 (13)	6.29 (20)	6.40 (27)	6.62 (33)	7.06 (40)

Table 1: Computation time (in seconds) of the different methods, for the compilation of the tables and the calculation of multipliers, using direct and iterative algorithms (number of iterations between brackets).

slower than the sparse one (two orders of magnitude) due to the large number of nested iterations required for the allocation of international transactions to direct transactions between domestic firms of different countries. The compilation time of the endogenous MRIO is double that of the exogenous one, due to the allocation of margins. If these iterations were performed in a low-level language such as C or Fortran the performance would surely be better, but it would always be worse than in the case of the sparse matrix. The compilation of the sparse matrix is very fast since no allocation is made and there is an injective relation between GTAP and sparse MRIO entries, i.e., every MRIO entry is either a GTAP entry or a sum thereof. This relation is not bijective, i.e., there is no one-to-one correspondence between entries in both datasets, since some of the MRIO entries are aggregations of GTAP entries (essentially taxes and subsidies).

We compared the time required to calculate the carbon intensity and the aggregate carbon emissions embodied in the final demand of all GTAP regions. A carbon intensity is equivalent to a price multiplier in a cost-push IO model [Oosterhaven, 2006], and is computed as:

$$\left(\mathbf{I}-\hat{\mathbf{x}}^{-1}\mathbf{Z}'\right)\mathbf{m}^{U}=\mathbf{m}^{L}, \tag{4.1}$$

where vectors are in column format, $'$ is transpose, $\hat{}$ is diagonal matrix, $\mathbf{m}^L$ and $\mathbf{m}^U$ are the vectors of direct and upstream embodied emissions [Rodrigues et al., 2010] and $\mathbf{I}$, $\mathbf{Z}$ and $\mathbf{x}$ are, respectively, the identity matrix, the matrix of inter-industry transactions and the vector of total output (Equation 4.1 is often presented in row vector format). We did not explicitly compute the Leontief inverse, for two reasons. First, our purpose is to compute the multiplier, and there are optimized methods to do so, i.e., to solve Eq. 4.1 directly (we used the default algorithms of Octave). Second, the inverse of a sparse matrix is usually not sparse, as happens in this case.

We can see in Table 1 that the computation time of the multipliers using a direct solver is roughly the same in the three models, but for different reasons. In the dense MRIOs the computational bottleneck is the upload into active memory of the table (circa 80 s), while the calculation of the multipliers itself is fast. In the case of the sparse MRIO the reverse is true.

Besides solving Eq. 4.1 directly, we used the following iterative expression to compute intensities:

$$\mathbf{m}_{i+1}^{U} = \mathbf{m}_{\mathbf{L}} + \hat{\mathbf{x}}^{-1}\mathbf{Z}'\mathbf{m}_{i}^{U}, \tag{4.2}$$

with $\mathbf{m}_0^U = \mathbf{m}_L$. We used the following stopping. Let $e^D = (\mathbf{m}^L)'\mathbf{x}$ and $e_i^U = (\mathbf{m}_i^U)'\mathbf{y}$ be total direct emissions and the total upstream emissions embodied in final demand, where $\mathbf{y}$ is the vector of total final demand. The iteration proceeded until:

$$1 - \frac{e_i^U}{e^D} < \delta,$$

where the accuracy, δ, is defined as the amount of direct emissions that remained unaccounted for in the embodied emissions of final demand.

The iterative expression, Eq. 4.2, was derived by rearranging Eq. 4.1 to yield:

$$\mathbf{m}^U = \mathbf{m}^L + \hat{\mathbf{x}}^{-1}\mathbf{Z}'\mathbf{m}^U.$$

The iterative expression is obtained simply by determining the vector of upstream embodied emissions in the left hand side (the $i+1$-th iteration) as a function of the vector in the right hand side (the i-th iteration).

Table 1 shows the the computation time of the multipliers using the iterative expression, Eq. 4.2, for five levels of accuracy. The number of iterations is displayed between brackets.

The iterative calculation of the multipliers does not offer any advantage in the case of the dense models, but in the case of the sparse matrix the benefit is substantial. An accuracy of $\delta = 0.0001\%$ and higher can be obtained in less than 10% of the computation time required using the direct solver.

It is also interesting to note that in the dense models the number of iterations required to attain a certain accuracy is lower than in the sparse model. This could be expected since in the sparse matrix an international transaction requires several steps while in the dense matrix only one takes place.

GTAP Region number and name		Direct	Exogenous	Endogenous	Sparse
1	Australia	315.27	296.91	305.47	305.47
2	New Zealand	28.34	33.66	34.30	34.30
3	Rest of Oceania	17.22	17.10	17.75	17.75
4	China	4071.13	3156.11	3147.11	3147.11
5	Hong Kong	54.70	101.18	97.23	97.23
6	Japan	924.98	1200.09	1214.09	1214.09
7	South Korea	344.35	371.76	335.40	335.40
8	Taiwan	220.70	165.03	167.41	167.41
9	Rest of East Asia	75.80	57.20	57.27	57.27
10	Cambodia	2.81	3.52	3.71	3.71
11	Indonesia	295.57	255.83	261.55	261.55
12	Lao People's Dem. Rep.	1.40	1.91	2.00	2.00
13	Malasya	125.32	76.35	68.91	68.91
14	Philippines	67.38	72.22	72.68	72.68
15	Singapore	38.20	73.02	58.33	58.33
16	Thailand	192.72	143.71	144.05	144.05
17	Vietname	72.93	69.15	67.97	67.97
18	Rest of Southeast Asia	7.44	8.81	9.08	9.08
19	Bangladesh	28.78	39.68	41.11	41.11
20	India	919.76	857.82	860.79	860.79
21	Pakistan	111.19	124.66	126.67	126.67
22	Sri Lanka	10.86	14.90	15.63	15.63
23	Rest of South Asia	8.35	13.36	13.97	13.97
24	Canada	460.01	425.97	424.99	424.99
25	United States of America	4879.14	5450.74	5511.71	5511.71
26	Mexico	327.08	347.13	353.65	353.65
27	Rest of North America	3.15	4.77	4.95	4.95
28	Argentina	118.20	87.71	88.41	88.41
29	Bolivia	8.96	8.32	8.52	8.52
30	Brazil	234.81	215.44	215.53	215.53
31	Chile	54.98	46.70	44.07	44.07
32	Colombia	45.19	47.25	48.14	48.14
33	Ecuador	17.31	21.45	21.09	21.09
34	Paraguay	2.87	4.59	4.55	4.55
35	Peru	25.09	29.06	30.06	30.06
36	Uruguay	4.02	6.38	6.30	6.30
37	Venezuela	123.52	87.29	88.30	88.30
38	Rest of South America	1.86	2.28	2.37	2.37
39	Costa Rica	4.14	5.87	6.39	6.39
40	Guatemala	8.47	12.55	13.49	13.49
41	Nicaragua	3.51	4.33	4.56	4.56
42	Panama	4.87	7.61	7.86	7.86
43	Rest of Central America	11.00	15.35	16.51	16.51
44	Caribbean	142.85	137.44	139.65	139.65
45	Austria	52.27	84.68	82.86	82.86
46	Belgium	72.39	131.31	124.15	124.15
47	Cyprus	7.05	9.18	9.24	9.24

Continued on next page

GTAP Region number and name		Direct	Exogenous	Endogenous	Sparse
48	Czech Republic	99.41	82.18	81.15	81.15
49	Denmark	44.27	64.41	62.00	62.00
50	Estonia	15.03	14.13	13.55	13.55
51	Finland	57.67	69.01	69.19	69.19
52	France	255.58	410.26	410.46	410.46
53	Germany	599.25	802.95	804.46	804.46
54	Greece	74.78	91.55	94.89	94.89
55	Hungary	42.71	52.02	52.20	52.20
56	Ireland	33.97	44.84	46.59	46.59
57	Italy	332.60	465.56	476.05	476.05
58	Latvia	6.45	12.30	11.84	11.84
59	Lithuania	9.42	15.06	14.51	14.51
60	Luxembourg	9.73	13.11	11.25	11.25
61	Malta	2.73	3.14	3.43	3.43
62	Netherlands	165.81	186.56	172.01	172.01
63	Poland	240.70	216.23	212.64	212.64
64	Portugal	50.19	64.56	64.34	64.34
65	Slovakia	24.67	25.56	25.91	25.91
66	Slovenia	12.62	13.75	13.93	13.93
67	Spain	266.76	321.35	324.69	324.69
68	Sweden	37.41	71.56	69.97	69.97
69	United Kingdom	438.29	653.99	657.36	657.36
70	Switzerland	26.69	70.04	72.40	72.40
71	Norway	52.45	56.98	46.51	46.51
72	Rest of EFTA	4.62	5.76	5.64	5.64
73	Albania	4.24	5.63	5.77	5.77
74	Bulgaria	41.83	31.72	31.29	31.29
75	Belarus	50.59	44.87	43.54	43.54
76	Croatia	15.20	19.56	20.30	20.30
77	Romania	76.53	69.83	69.01	69.01
78	Russian Federation	1332.95	1025.38	1016.77	1016.77
79	Ukraine	217.62	137.92	126.61	126.61
80	Rest of Eastern Europe	5.89	8.34	8.15	8.15
81	Rest of Europe	70.96	66.41	68.06	68.06
82	Kazakhstan	161.61	134.56	134.85	134.85
83	Kyrgyzstan	5.18	5.58	5.71	5.71
84	Rest of former Soviet Union	132.88	95.18	94.27	94.27
85	Armenia	3.38	4.28	4.29	4.29
86	Azerbaijan	24.18	26.81	26.86	26.86
87	Georgia	2.43	4.70	4.67	4.67
88	Iran, Islamic Rep. of	299.80	300.51	301.86	301.86
89	Turkey	163.34	186.78	192.71	192.71
90	Rest of West Asia	909.16	731.14	707.86	707.86
91	Egypt	120.29	101.43	101.71	101.71
92	Morocco	31.86	36.40	38.01	38.01
93	Tunisia	18.38	17.88	18.61	18.61
94	Rest of North Africa	127.64	110.01	110.65	110.65

Continued on next page

GTAP Region number and name		Direct	Exogenous	Endogenous	Sparse
95	Nigeria	39.92	37.09	38.16	38.16
96	Senegal	4.15	5.41	5.91	5.91
97	Rest of West Africa	19.85	32.37	34.51	34.51
98	Rest of Central Africa	7.80	11.04	11.54	11.54
99	Rest of South Central Africa	9.09	13.53	14.40	14.40
100	Ethiopia	3.70	6.78	6.74	6.74
101	Madagascar	1.36	1.92	2.03	2.03
102	Malawi	0.55	1.44	1.57	1.57
103	Mauritius	1.83	3.56	3.80	3.80
104	Mozambique	1.60	3.27	3.40	3.40
105	Tanzania, United Rep. of	3.06	6.48	6.51	6.51
106	Uganda	2.26	3.31	3.51	3.51
107	Zambia	1.77	3.01	3.09	3.09
108	Zimbabwe	8.78	6.83	6.89	6.89
109	Rest of Eastern Africa	21.04	33.25	34.27	34.27
110	Botswana	3.76	6.24	6.38	6.38
111	South Africa	329.12	209.92	213.38	213.38
112	Rest of South African CU	3.44	6.12	6.33	6.33
		21730.77	21730.77	21730.77	21730.77

Table 2: Direct carbon emissions and carbon emissions embodied in the final demand of GTAP regions (Mt CO_2).

Table 2 shows the comparison of the total embodied carbon in each region, compared to direct emissions, using the three models (with the direct calculation method). The values shown do not include household emissions, which are identical in all models, and values may be different from those reported in Peters et al. [2011a] because we used the unprocessed GTAP emissions data.

As expected, we find that the results of the endogenous dense MRIOT and the sparse MRIOT models are identical, apart from numerical rounding errors: the relative difference between the two methods is always less than 10^{-4}%. This is valid for the aggregate values show in Table 2 and for the disaggregate multiplier values for every sector of every region. The trade share allocations of the endogenous dense MRIOT and the intermediate firms of the sparse MRIOT play the same mathematical role, which is to distribute a given aggregate flow homogeneously among a certain number of disaggregate sectors.

In contrast, the difference between the exogenous dense MRIOT and the sparse MRIOT models can be as large as 25% and has a median value of 2%. By term of comparison, the relative difference between direct emissions and the sparse MRIOT model can be as large as 81% with a median of 25%. As already noted by Peters et al. [2011a], the allocation of the provision of international transport to final consumption instead of its provision to a global pool of international transport (the difference between the exogenous and the sparse model) is much larger than the difference between the endogenous and sparse models, but much smaller than the difference between direct emissions and the results of the sparse model.

5 Conclusions

The most appropriate computational tool for a given task depends on the scale of the problem considered. The demands posed by the processing of a highly aggregated single-region closed IO table or those of a detailed multi-regional IO table covering the whole world are vastly different. The latter case involves a considerable amount of data and substantial computational requirements.

The GTAP database is often used to build a world MRIOT. In this paper we have shown that this can be done with minimal processing, producing a light and fast sparse MRIOT. The gains over an equivalent dense model are of an order of magnitude in terms of data storage and computation time, using the iterative implementation, and of more than three orders of magnitude in terms of processing time.

It is important to emphasize that data storage refers both to space in the hard disk and to active memory. Therefore, the use of the sparse MRIOT greatly expands the range of possibilities offered to researchers that do not have access to supercomputers.

We also note that the advantage of the sparse over the dense format are not specific to the current size of the GTAP database. Therefore, the sparse format allows for a substantial increase in the size of the system (for example by integrating the GTAP with sub-national regional data or process-oriented life-cycle data). The dense format, on the other hand, is already very close to the computational limit posed by the RAM and cache specifications of modern personal computers (a few GB).

The preparation of a MRIOT is much more time-demanding than the final computation, and most of that time is spent debugging code. To debug the code, however, requires performing the computation multiple times, which leads to a multiplier effect: by saving computation time, the sparse MRIOT also saves programming time.

The GTAP database is already provided in sparse format, and so the conversion to a sparse MRIOT is particularly straightforward. However, we believe that the use of a sparse format in the construction and analysis of a large

MRIOT is convenient, whichever the data source, because in such models the problem of under-determined transactions always arises. The consideration of intermediate firms is more parsimonious than the mathematically equivalent use of trade share allocations because it avoids a substantial amount of data processing, which is always error-prone, and it is also conceptually clearer.

In conclusion, we believe that the sparse format should be preferred over the dense format both due to the computational advantages and the conceptual clarity that it offers in the construction and analysis of multi-regional input-output tables.

Acknowledgments

We would like to thank the financial support of FCT through grant PTDC/-AMB/64762/2006 and FCT and the MIT Portugal Program through scholarship SFRH/BD/42491/2007 (to AM).

References

M. Brockmeier. A graphical exposition of the GTAP model. *GTAP Technical Paper*, 1996.

S.J. Davis and K. Caldeira. Consumption-based accounting of CO_2 emissions. *Proceedings of the National Academy of Sciences*, 12(107):5687–5692, 2010.

G. H. Golub and C. F. Van Loan. *Matrix Computations.* John Hopkins University Press, Baltimore, USA, 1996. 3rd edition.

E.G. Hertwich and G.P. Peters. Carbon footprint of nations: a global, trade-linked analysis. *Environmental Science and Technology*, 16(43):6414–6420, 2009.

IDE. Asian international input-output tables 2000 volume 1. explanatory notes. Technical report, Institute of Developing Economies, Japan External Trade Organization, Japan, 2006.

M. Lenzen, K. Kanemoto, A. Geschke, D. Moran, P. Munoz, J. Ugon, R. Wood, and T. Yu. A global multi-region input-output time series at high country and sector detail. *18th International Input-Output Conference of the International Input-Output Association*, 2010.

B. Narayanan and T.L Walmsley. Global trade, assitance and production: the GTAP 7 data base. Technical report, Center for global trade analysis – Purdue University, West Lafayette, IN, 2008.

J. Oosterhaven. On the plausibility of the supply-driven input-output model. *Journal of Regional Science*, 2(28):203–217, 2006.

J. Oosterhaven, D. Stelder, and S. Inomata. Estimating international interindustry linkages: Non-survey simulations of the asian-pacific economy. *Economic Systems Research*, 20 (4):395–414, 2008.

G.P. Peters and E.G. Hertwich. CO_2 embodied in international trade with implications for global climate policy. *Environmental Science and Technology*, 5(42):1401–1407, 2008.

G.P. Peters, R. Andrew, and J. Lennox. Constructing an environmentally extended multi-regional input-output table using the GTAP database. *Economic Systems Research*, 2(23):131–152, 2011a.

G.P. Peters, J.C. Minx, C.L. Weber, and O. Edenhofer. Growth in emission transfers via international trade from 1990 to 2008. *Proceedings of the National Academy of Sciences*, 21(108):8903–8908, 2011b.

J. Rodrigues, A. Marques, and T. Domingos. *Carbon Responsibility and Embodied Emissions: Theory and Measurement.* Routledge, London, UK, 2010.

A. Tukker, E. Poliakov, R. Heijungs, T. Hawkins, F. Neuwahl, J.M. Rueda-Cantuche, S. Giljum, S. Moll, J. Oosterhaven, and M. Bouwmeester. Towards a global multi-regional environmentally extended input-output database. *Ecological Economics*, 69(7):1928–1937, 2009.

T. Wiedmann. A review of recent multi-region input-output models used for consumption-based emission and resource accounting. *Ecological Economics*, 69(2):211–222, 2009.