# The estimation of international inter-industry flows and embodied greenhouse gas emissions

João Rodrigues[a]*, Tiago Domingos[a]**
[a] Environment and Energy Section - DEM, Instituto Superior Técnico
Av. Rovisco Pais 1, 1049-001 Lisboa, PORTUGAL
* Tel: +351-21-841 91 63, Fax: +351-21-841 73 85, e-mail: joao.rodrigues@ist.utl.pt
** Tel: +351-21-841 91 63, Fax: +351-21-841 73 85, e-mail: tdomingos@ist.utl.pt

## Abstract

In input-output (I-O) data international inter-industry flows are not usually available, but have to be estimated from more aggregate quantities. In the present paper we present a new estimation method, based on the maximum entropy principle (MEP) and computed numerically using a genetic optimization algorithm. Unlike conventional methods, the MEP allows for the incorporation of arbitrary information and the quantification of the estimation error. As a case-study we compute upstream and downstream embodied greenhouse gas emissions and intensity, using different methods. In the case study, we consider a multi-regional I-O model with 80 world regions and 3 sectors, using data from the GTAP and UNFCCC databases. We conclude that the results of the MEP are not well approximated by conventional methods and that the estimation error is small.

**Keywords:** Greenhouse gas (GHG) emissions; input-output (I-O) analysis; international inter-industry flows; maximum entropy (MEP) principle; genetic optimization algorithm; upstream and downstream; GHG intensity; GHG embodied emissions.

## 1 Introduction

According to the Intergovernmental Panel for Climate Change (IPCC, 2001), "[h]uman activities have increased the atmospheric concentrations of greenhouse

gases (GHG) and aerosols since the pre-industrial era"; and "[t]here is […] evidence that most of the warming observed over the last 50 years is attributable to human activities."

Climate change policy, as contemplated by the Kyoto Protocol, considers the use of direct GHG emissions as environmental indicator. However, several authors have proposed the use of indicators that account for indirect emissions (Lenzen et al., 2007; Rodrigues et al., 2006; Tukker and Janssen, 2006, Wiedmann et al., 2007).

The quantification of indirect emissions is made using single-region (e.g., Munksgaard and Pedersen, 2001) or multi-region (e.g., Ahmad and Wyckoff, 2003; Lenzen et al., 2004; Peters and Hertwich, 2006) input-output (I-O) models (Miller and Blair, 1985), usually focusing only on upstream indirect effects.

The quantification of indirect GHG emissions embodied in international trade is problematic because statistical offices do not provide disaggregate information on these flows, as they do for domestic flows.

However, the magnitude of international trade is non-negligible. The global value of international trade in 2001 was $7.14 \times 10^{12}$ USD while domestic trade was $29.47 \times 10^{12}$ USD (GTAP database). Therefore international trade accounted for almost 1/5 of global trade, and for small open economies, this fraction can be substantially higher.

Therefore, in order to account for the indirect effects of international trade with I-O analysis (as opposed to CGE models, Kainuma et al., 2000; linear programming, Duchin, 2006; or simulation models, Lutz et al., 2007), it is necessary to estimate disaggregate international flows from aggregate information (or to consider a single sector per region, thus disregarding information on domestic trade, Proops et al.,

1999). Aggregate information can be, for example, in the form of bilateral trade or industry specific imports (Dimaranan, 2007).

Two main methods for the estimation of international inter-industry exist (Wiedmann et al., 2007), requiring different amounts of information: (*i*) to consider that imports have the same upstream GHG intensity as domestic products (Munksgaard and Pedersen, 2001); (*ii*) to apply "trade shares" from aggregate bilateral trade data to industry-specific imports by industry (Ahmad and Wyckoff, 2003; Lenzen et al., 2004).

These basic methods are complemented with ad hoc assumptions such as neglecting higher order effects, lumping marginal trade partners and assuming a standard technology for the rest of the world (Hertwich and Peters, 2006).

In the present paper we present a new method for the estimation of international inter-industry flows, based on the maximum entropy principle (MEP), that allows the incorporation of arbitrary information in the estimation process and the quantification of the estimation error. Therefore, this method can be used to validate the use of simpler methods and of auxiliary ad hoc assumptions.

The MEP is a Bayesian method (Lee, 1989), proposed by Jaynes (1957; 1983) to estimate an unknown quantity. The MEP assumes that the unknown quantity is a random variable whose probability distribution is the least informative one, constrained by all available information. In this context, the unknown quantity is an international inter-industry flow, and the available information is a set of aggregate constraints simultaneously involving different international inter-industry flows. We show that the probability distribution of an arbitrary estimated international flow is a truncated exponential distribution, whose expected value is determined by a non-

linear set of algebraic equations. This solution cannot be solved analytically, but it can be solved with a numerical genetic optimization algorithm (Holland, 1975; Schmitt, 2001).

As a case-study for this method, we use as source data the United Nations' Key GHG Data (UNFCCC, 2005) database for direct GHG emissions and the GTAP database (Dimaranan, 2006) for trade data. We consider 80 world regions and 3 sectors per region. We compare the upstream and downstream GHG intensity and embodied emissions of products (Rodrigues and Domingos, 2007), computed using different methods.

The structure of the paper is as follows. Section 2 presents the theory; Section 3 presents materials and methods; Section 4 presents the results; and Section 5 concludes.

## 2 Theory

In the present Section we first define the input-output (I-O) economic (Miller and Blair, 1985) and environmental (Rodrigues and Domingos, 2007) model. Then we review existing methods for the estimation of disaggregate international flows (Wiedmann et al., 2007). We conclude with the application of the maximum entropy principle (Jaynes, 1957; 1983) to the estimation of disaggregate international flows.

Throughout the paper the following notation is used.

$N$ denotes a total (integer); $\sigma$ denotes a set. The following subscripts appear commonly: $C$ denotes composite; $R$ denotes regions; $I$ denotes industries; $K$ denotes constraints; $T$ denotes number of trade flows.

Subscripts *i*,*j*,*m* and *n* denote sector or region; (*ij*) denotes a flow from *i* to *j*; sector *i*, $i > 0$, denotes an industry.

Scalar quantities are denoted by *italic*: the units of monetary transactions (several symbols) are $10^9$ USD2001; the units of *e* (GHG *emissions*) are$10^9$ $KgCO_2$ equivalent; the units of $\mu$ (GHG *intensity*) are KgCO2/USD2001.

Vectors (lowercase) and matrices (uppercase) are denoted by **bold**; **‘** denotes matrix transpose; ^ denotes diagonal matrix; # denotes entry-wise multiplication; ÷ denotes entry-wise division; ~ denotes modification and * denotes target. All vectors are column-vectors by default.

Superscripts *L*, *U* and *D* denote respectively *local*, *upstream* and *downstream*; superscripts *P* and *M* denote respectively *product* (good or service) and *margin*; Superscripts *I* and *F* denote *intermediate* and *final*.

**Q**, **S** and **T** denote the inter-industry flow matrices of total flows (domestic + international), of domestic flows and of international trade flows. **A**, **x**, **y** and **v** denote the Leontieff matrix, total output column vector, final expenditure column vector and added value row vector.

A vector with superscript **r** denotes the row sums of a matrix; a vector with superscript **c** denotes the column sums of a matrix (e.g., $\mathbf{t}^{\mathbf{r}}$ and $\mathbf{t}^{\mathbf{c}}$ are row and column sums of **T**) .

**k** and **λ** denote, respectively, a *constraint* and a *Lagrange multiplier* vectors of length $N_K$ and $N_T$; **G** denotes an *aggregator* matrix of size ($N_K$.$N_T$); $\pi$ denotes potential; $\tau$ and $\theta$ denote a *random numbers* and its *upper bounds*. $\phi$, $\gamma$, $\chi$ and $\Lambda$ are auxiliary variables; $f(\cdot;\cdot)$ and $h(\cdot;\cdot)$ are functions and *H* is the Hamiltonian.

$E[\cdot]$ denotes expectation and $Var[\cdot]$ denotes variance; $p$ denotes probability and $\omega$ denotes a realization; $\varepsilon$ denotes error.

## 2.1 I-O analysis and GHG responsibility

Following a simplification of SNA 1993 nomenclature (UN, 1994), consider that the world is divided in $N_R$ *regions* and that each region possesses $N_I$ *industries*. Consider that each industry produces a single *product* (i.e., a good or service). Consider that industries engage in domestic and international trade with other industries; receive flows from added value and deliver flows to final expenditure (UN, 1999).

Let $g_{ij}$ denote the economic flow from industry $i$ to industry $j$. Let $v_j$ denote the flow of *added value* (surplus, wages, interests, taxes) of industry $j$ and let $y_i$ denote the flow of *final expenditure* (investment, private and public consumption) of industry $i$. Let $x_i$ be the *total input or output* of sector $i$. The I-O identity for product (or industry) $i$ is:

$$x_i = \sum_{j=1}^{N_R N_I} Q_{ij} + y_i \text{ and } x_i = \sum_{j=1}^{N_R N_I} Q_{ji} + v_i .$$

Or in matrix notation $\mathbf{x = Q.1 + y}$ and $\mathbf{x = Q'.1 + v}$ ; where $\mathbf{1}$ is a vector of 1's. The Leontieff matrix is $\mathbf{A = Q.\hat{x}}$ m where ^ denotes diagonal matrix, and the Ghosh matrix is $\mathbf{A'}$ , where ' denotes transpose ((Leontief, 1970; Ghosh, 1958; Miller and Blair, 1985; UN, 1999).

Let $e_i^L$ denote the local GHG emissions of industry $i$.

The *upstream* (or *downstream*) *GHG intensity* of an industry is the amount of direct and indirect GHG emissions required to generate one unit of output (or input) of that industry. Intensities are computed as:

$$\boldsymbol{\mu}^{\mathbf{U}\prime} = \boldsymbol{\mu}^{\mathbf{L}\prime}\left(\mathbf{I} - \mathbf{A}\right)^{-1} \text{ and } \boldsymbol{\mu}^{\mathbf{D}} = \left(\mathbf{I} - \mathbf{A}'\right)^{-1}\boldsymbol{\mu}^{\mathbf{L}} \ , \qquad (1)$$

where ' denotes transpose (Rodrigues et al., 2006; Rodrigues and Domingos, 2007). Once these vectors are known we can compute the *upstream* and *downstream embodied GHG emissions* in *final expenditure* and *added value* as $e_i^U = \mu_i^U y_i$ and $e_j^D = \mu_j^D v_j$ respectively. In matrix notation:

$$\mathbf{e^U} = \boldsymbol{\mu}^{\mathbf{U}} \# \mathbf{y} \text{ and } \mathbf{e^D} = \boldsymbol{\mu}^{\mathbf{D}} \# \mathbf{v}, \quad (2)$$

where # denotes entry-wise multiplication.

The quantities presented in Eq. (2) can be aggregated on a regional basis to define $E_k^U$ and $E_k^D$, the producer and consumer GHG responsibility of region *k* (Rodrigues and Domingos, 2007) as:

$E_k^U = \sum_{i \in \sigma_k} e_i^U$ and $E_k^D = \sum_{i \in \sigma_k} e_i^D$, where $\sigma_k$ is the set of sectors that compose region *k*.

Throughout the paper we systematically use the RAS algorithm (UN, 1999). The RAS algorithm iteratively transforms an arbitrary matrix **Q** into matrix **Q***, which verifies row and column sums $\mathbf{q^r}$ and $\mathbf{q^c}$, through rescaling of rows and columns. The algorithm is as follows.

Let $\mathbf{a_0}$ and $\mathbf{b_0}$ be vectors with all entries 1. Now let $\mathbf{a_i} = \mathbf{g^r} \div (\mathbf{Q}\mathbf{b_{i-1}})$ and $\mathbf{b_i} = \mathbf{g^c} \div (\mathbf{Q'}\mathbf{a_i})$, for any $i > 0$, where ÷ represents Hadamard (entry-wise) division. After a finite number of iterations, $\hat{\mathbf{a}}_\mathbf{i}\mathbf{Q}\hat{\mathbf{b}}_\mathbf{i} \approx \mathbf{Q}^*$.

(More sophisticated methods to balance I-O tables, incorporating arbitrary information, are presented in Gilchrist and St. Louis, 2005, and Robinson et al., 2001).

Let the matrix of world inter-industry transactions, **Q**, be **Q** = **S** + **T**, where **S** is the known matrix of domestic trade, and **T** is the unknown matrix of international trade.

In the remainder of this Section we shall look at the different methods to construct matrix **T** such that Eqs. (1-2) can be computed.

## 2.2 Estimation of international trade

For the present Subsection let **T** be a square matrix of length $N_I N_R$, where $N_I$ is the number of industries and $N_R$ is the number of regions. That is, entry $((m-1)N_R + i, (n-1)N_R + j)$ in **T** is the flow from industry $i$ in region $m$ to industry $j$ in region $n$.

Now consider **T** is subject to a set of constraints. That is, let $N_T = (N_I N_R)^2$; let **G** be an $(N_K.N_T)$-aggregation matrix, whose entries are either 0 or 1; let **t** be an $N_T$-vector whose $((i-1).(N_I N_R) + j)$ entry equals the the $(i,j)$ entry of matrix **T**; and let **k** be an $N_K$-vector of known real non-negative constraints. Vectors **k** and **t** are related as:

$$\mathbf{k} = \mathbf{G}\mathbf{t} .$$

Example: If constraint $k_5$ represents the flows from industry 5 of all regions other than 1 to industry 1 of region 1 then:

$$k_5 = \sum_{m=2}^{N_R} T_{(m-1)N_R+5,1} ,$$

and row 5 in matrix **G** reads 1 in entry $((m-1)N_R + 5).N_R N_I + 1)$, for $m = 2,..., N_R$, and 0 in all other entries.

Furthermore, let $\mathbf{t}^{\mathbf{r}}$ be the export vector, a column-vector whose entries are the row sums of **T**, and let $\mathbf{t}^{\mathbf{c}}$ be the import vector, a row vector whose entries are the column sums of **T**.

$\mathbf{t}^r$ and $\mathbf{t}^c$ are constraints, just like $\mathbf{k}$, but they play a special role for several reasons. First, they are more easily available since they are provided by national statistical offices. Second, for computation purposes, the I-O table must be balanced, and thus it is acceptable that some of the constraints in $\mathbf{k}$ are not verified but $\mathbf{t}^r$ and $\mathbf{t}^c$ must always hold ($\mathbf{t}^r + \mathbf{s}^r + \mathbf{y} = \mathbf{t}^c + \mathbf{s}^c + \mathbf{v} = \mathbf{x}$).

We now briefly review the two main conventional methods to estimate $\mathbf{T}$: the "homogeneous intensity" and the "trade share" method. A more comprehensive review can be found in Wiedmann et al. (2007).

**Homogeneous intensity**

Consider that only the total of imports and exports of a single region are known (e.g., as provided by a national statistical office). In this circumstance, homogeneous intensity is a common assumption (Munksgaard and Pedersen, 2001), by which imports (or exports) are assumed to have the same upstream (or downstream) GHG intensity as domestic products.

Following this approach matrix $\mathbf{T}$ is the diagonalized matrix $\hat{\mathbf{t}}^c$, if upstream intensity is computed, or the diagonal matrix $\hat{\mathbf{t}}^r$, if downstream intensity is computed.

**Trade share**

This is the standard method in multi-regional I-O models (Ahmad and Wyckoff, 2003; Lenzen et al., 2004) that besides total imports and exports also uses aggregate bilateral trade data (i.e., total flow from region *m* to region *n*) and imports by industry (i.e., sum of flows from industry *i* of all regions to industry *j* of region *n*). Using this method the estimate for the flow from industries *i* to *j* from regions *m* to *n*:

$$T_{(m-1)N_I+i,(n-1)N_I+j} = \frac{T_{ij}^n T_{mn}^R}{t_n^R},$$

where $T_{ij}^{n}$ is the sum of imports from sector *i* of all regions to sector *j* of region *n*, $T_{mn}^{R}$ is the imports from region *m* to region *n* and $t_{n}^{R}$ is the total imports of region *n*.

The RAS method should afterwards be applied to balance matrix **T**, thus also incorporating data from total imports and exports.

This is a multiplicative method since information from a fraction of (more aggregate) constraints is multiplied by another (more disaggregate) constraint.

With multiplicative methods it is not easy to evaluate how much information is obtained from each constraint and it is not easy to incorporate different constraint sets.

For example, it is not clear how export data by industry (i.e., sum of exports from sector *i* of region *m* to sector *j* of all regions) can be incorporated in the estimation.

**Simplifications**

Authors often use complementary ad hoc assumptions, in order to simplify calculations. We highlight three.

The "small region" assumption consists in not considering exports from the region whose upstream intensity is being computed (Hertwich and Peters, 2006), assuming that the indirect effects of these exports are diluted in the rest of the world.

The "rest of the world" assumption transforms a many-region I-O model into a few-region one by lumping together all the marginal trading partners of the target region, and using technology from a representative region to describe that aggregate region (Hertwich and Peters, 2006).

The "homogeneous region" assumption consists in considering a single sector per region (Imura et al., 2000), for marginal trading partners for which aggregate bilateral trade data is available.

All simplifying assumptions are potentially useful, if the error they lead to is small compared to the computation gain. With current estimation methods a quantification of this error is not possible.

**Maximum entropy principle**

The new method proposed is this paper is the based on the maximum entropy principle (MEP), and its theory is presented in the following Subsection.

The MEP method is able to incorporate arbitrary information in the estimation process in a consistent manner; and allows the quantification of the estimation error.

## 2.3 Maximum entropy principle

The maximum entropy principle (MEP), proposed by Jaynes (1957; 1983) is a method from Bayesian statistics (Lewis, 1989) that addresses the problem of estimating the probability distribution of a random variable when partial information is available.

An example: consider that a person is playing with a 6-face die (numbered 1 to 6) and she knows that in a recent run, the average turnout has been 3. The next throw is more likely to deliver a 1 or a 6? (Answer a few paragraphs below.)

This method is appropriate to estimate international flows because the structure of the problem is similar. We can treat an arbitrary unknown international flow as a realization of a random variable $\tau_i$ of unknown distribution. Let $N_T$ be the total of such unknowns, let $N_K$ be total of constraints and let $g_{mi}$ be 1 if $i$ is constrained by $m$, and 0 otherwise.

If each international flow is treated as a real random variable, the constraint (being a non-random number) applies to the expected value of the sum of international flows. That is, let $E[\cdot]$ and $k_m$ denote the value of constraint $m$. We can write:

$$k_m = E\left[\sum_{i=1}^{N_T} G_{mi}\tau_i\right], \tag{3}$$

and each particular random variable *i* is subject to $g_i^c = \sum_{m=1}^{N_T} G_{mi}$ such constraints. Likewise, the number of elements of constraint *m* is $z_m^r = \sum_{i=1}^{N_K} Z_{mi}$. The range of $\tau_i$ is $(0,\theta_i)$ where $\theta_i$ be the minimum constraint that affects *i*:

$$\theta_i = \min\{k_m \mid G_{mi} = 1\}_{m=1}^{N_K}. \tag{4}$$

The maximum entropy principle (Jaynes, 1957; 1983) states that $\tau_i$ possesses the most likely or less informative distribution, i.e., the distribution that maximizes entropy subject to the set of constraints. Let $p_i(\omega)$ be the probability that $\tau_i$ takes value $\omega$. The probability distribution $p_i(\omega)$ is obtained by maximizing the Hamiltonian, $H_i$:

$$H_i = \int_0^{\theta_i} p_i(\omega)\ln p_i(\omega)d\omega + \sum_{m=1}^{N_K} G_{mi}\lambda_m\left(E\left[\tau_i + \sum_{j=1,(j\neq i)}^{N_T} G_{mj}\tau_j\right] - k_m\right). \tag{5}$$

The first term in the right hand side of Eq. (5) is the entropy of the distribution of $\tau_i$ and the second term in the right hand side of Eq. (5) is the set of constraints (Eq. 3), where $\lambda_m$ is the Lagrange coefficient of constraint *m*.

If $g_m^r$ is large, $k_m$ is the sum of many elements and it is acceptable to assume that that $\tau_i$ is independent of $\tau_j$, for any $j \neq i$ and $i, j = 1,..., S_U$. We consider that this always holds. Therefore, we can simplify constraint *m* as follows:

$$E\left[\tau_i + \sum_{j=1,(j\neq i)}^{N_T} G_{mj}\tau_j\right] = E[\tau_i] + E\left[\sum_{j=1,(j\neq i)}^{N_T} G_{mj}\tau_j\right].$$

Recalling the standard definition of expected value:

$$E[\tau_i] = \int_0^{\theta_i} p_i(\omega)\omega d\omega,$$

Equation 5 can be rewritten as:

$$H_i = \int_0^{\theta_i} p_i(\omega)\ln p_i(\omega)d\omega + \sum_{m=1}^{N_K} G_{mi}\lambda_m \int_0^{\theta_i} p_i(\omega)\omega d\omega + \Omega,$$

where Ω is a term that does not depend neither on $\omega$ nor on $p_i(\omega)$. Maximization of of the previous equation, $\partial H_i / \partial p_i = 0$, for $\omega \in (0, \theta_i)$, leads to the standard result (Jaynes, 1957; 1983):

$$p_i(\omega) = \frac{(\Lambda_i)\exp(-\omega\Lambda_i)}{1-\exp(-\theta_i\Lambda_i)} \ . \tag{6}$$

That is, $\tau_i$ possesses a bounded exponential distribution, whose expected value is:

$$E[\tau_i] = f(\Lambda_i;\theta_i) = \theta_i\left(\frac{1}{\Lambda_i\theta_i} - \frac{1}{\exp(\Lambda_i\theta_i)-1}\right), \tag{7}$$

where $\Lambda_i = \sum_{m=1}^{N_K} G_{mi}\lambda_m$, $\Lambda_i \in (-\infty,+\infty)$, $\lambda_m \in (-\infty,+\infty)$ and $\theta_i > 0$. Function $f(\Lambda_i;\theta_i)$ is not defined for $\Lambda_i\theta_i = 0$, but using a 2nd order Taylor expansion and D' Alembert's rule it is easy to show by continuity that $\lim_{\Lambda_i \to 0} f(\Lambda_i;\theta_i) = \theta_i/2$.

To provide some intuition behind Eq. (7) let us consider unknown flow 1. Figure 1 shows how $t_1 = E[\tau_1]$ depends on $\Lambda_i\theta_i$.

INSERT FIGURE 1 HERE.

That is, $E[\tau_1]$ is monotonically decreasing with $\phi_1$; if $\phi_1 = 0$, then $E[\tau_i] = \theta_i/2$; if $\phi_1 >> 0$, then $E[\tau_1] << \theta_1$; and if $\phi_1 << 0$, then $E[\tau_1] \approx \theta_1$. Figure 2 shows the distribution of $p_1$ (Eq. 6) under these several scenarios.

INSERT FIGURE 2 HERE.

If $E[\tau_1] << \theta_1$, then the bounded exponential is well approximated by an unbounded exponential, and $\omega$ is biased toward low values. If $E[\tau_1] \approx \theta_1$, the opposite happens, $\omega$ is biased toward high values. If $E[\tau_i] = \theta_1/2$, $\tau_1$ possesses uniform distribution.

(The solution of the dice problem is as follows. The range of possible results is the set of integers from 1 to 6. Therefore, for the dice not to be unbiased, the expected value should be 1 + (6 – 1)/2 = 3.5. Since the expected value 3 is smaller than 3.5, the dice is biased toward low values and is therefore more likely to deliver a 1.)

Parameters $\lambda_m$ are unknown but they can be determined from the constraints, through simultaneous application of the maximum entropy principle to all estimated variables $\tau_i$. The system is well determined since there is a total of $N_K$ unknown parameters $\lambda_m$, and there are $N_K$ constraints $k_m$, related as $\mathbf{k} = \mathbf{G}f(\mathbf{\Lambda},\boldsymbol{\theta})$, where $f(\cdot,\cdot)$ is given by Eq. (7), $\mathbf{\Lambda} = \mathbf{G'}\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ is given by Eq. (4).

However, the system is non-linear and must be solved numerically. We found that a genetic optimization algorithm is suitable to solve this problem.

Once vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ have been determined, the properties of $\boldsymbol{\tau}$ are known. The vectors of estimated international flows and corresponding error are $\mathbf{t} = E[\boldsymbol{\tau}]$ and $\boldsymbol{\varepsilon}^{\mathbf{t}} = Var[\boldsymbol{\tau}]$, where $\varepsilon$ denotes error and $Var[\cdot]$ denotes variance.

## 3 Material and methods

In the present Section we report the I-O source data (Dimaranan, 2006; UNFCCC, 2005); the data preparation procedure; and the numerical algorithm for solving the MEP method.

## 3.1 Source data

We use the Key GHG Data of the United Nations Framework Convention on Climate Change for local emissions (UNFCCC, 2005). This document reports local emissions of different pollutants for a maximum of 9 sectors per nation and for a total of 160 nations. The document reports emissions for the period 1990-2003, and we chose to use 2003 for the emissions of Annex I countries and the most recent year for which data was available for Annex II countries. Emissions are reported in Gg ($10^9$ g) of $CO_2$-equivalent. The reference years for the emissions of the different regions are reported in Appendix I of the present paper. We considered acceptable to use different reference years between Annex I and Annex II countries in order to make use of the most consistent and recent data set.

Around 80% of GHG emissions stem from $CO_2$ (if accounted by gas type) and by the energy sector (if accounted by sector) (UNFCCC, 2005). According to Panayatou et al. (2002), "[t]here is a great deal more uncertainty surrounding the measurement of emissions from land use relative to fossil fuels, and even greater uncertainty regarding other sources of $CO_2$, particularly livestock and solid waste" (p. 442, paragraph 5). We considered only emissions from the energy sector and industrial processes, which we lumped together. We believe this to be reasonable since "[t]he energy sector is composed of two parts: a) fuel combustion and b) fugitive emissions. The fuel combustion part includes the following components (or sub-sectors): energy industries, manufacturing industries and construction, transport, other sectors, and other." (UNFCCC, 2005, p. 16, note 8). Therefore, the "energy" sector actually comprises energy, transport and what is conventionally called the secondary sector of the economy.

We use the GTAP 6.0 database for the economic source data (Dimaranan, 2006). The GTAP 6.0 database represents the state of the world economy in 2001, as a system of flows of goods and services, reported as money values, in 2001 million USD. The GTAP 6.0 model partitions the world into 87 regions, and 57 sectors per region. We considered a total of 80 regions and 3 sectors; we also consider higher-level aggregations for comparison purposes. The number of regions considered was the intersection of the sets of regions for which both monetary and emissions data were available.

The industries considered were "energy", as defined above, agriculture and services. Given the strong aggregation of the "energy" industry, whose emissions data cannot be disaggregated (unless we used another emission data source), we considered that it would not make sense to consider a higher disaggregation for the rest of the economy. The details of the aggregation procedure can be found in Appendix I of the present paper.

The GTAP database provides 19 source tables that contain information relative to inputs and outputs of firms, from which the monetary I-O data was built. The network of interactions described by these tables is organized in Table 1. Table 1a identifies the source tables and Table 1b presents the total aggregated value (sum over all sectors and regions) of those tables.

INSERT TABLE 1 HERE.

Table 1 is an I-O table that displays sales in rows and purchases in columns between 7 aggregate sectors and an external sector 0. These sectors are: 1 – Firms' sector; 2 – domestic trade sector; 3 – export sector; 4 – international transport sector; 5 – customs; 6 – general import sector; 7 – intermediate import sector.

These are imaginary sectors, which allow the systematic discrimination of who pays how much in taxes and transport margins to whom. GTAP considers three types of prices: agents' prices; market prices and world prices; these reflect the impact of taxes and margins; the definitions of prices differ from domestic to international trade (Dimaranan, 2007; McDonald and Thierfelder, 2004).

For domestic trade, firms sell products of value VDFM (market price; flow 1→2 in Table 1) and pay products of value VDFA (agents' price; 2→1); taxes paid (or subsidies received) on those products are accounted in matrix ISEP1 (0→2).

For international trade, firms sell products of value VXMD (exports at market price; 1→3) and transport firms sells transport margins of value VTWR (at world prices; 1→4). After paying taxes to the exporting region (MFRV and XTRV; 0→3), exports are now delivered to the importing region at value VXWD (exports at world prices; 3→4). The importing region receives imports at value VIWS (imports at world prices, including transport margins; 4→5). Imports pay taxes to the importing region (TFRV; 0→5) and then enter the domestic market at value VIMS (imports at market price; 5→6). These imports are then divided between private and government final expenditure (VIPM and VIGM, at market price; 6→0) and imports by firms (VIFM, at market price; 6→7). Firms' receive subsidies/pay taxes ISEP2 (0→7) on intermediate imports (VIFA, at agent's price; 7→1).

Besides these payments, firms also receive subsidies/pay taxes OSEP and deliver factor payments EVFA to added value (0→1). Firms deliver VDPM (private consumption, market price), VDGM (government consumption, market price) VDFA(CGDS) (investment from domestic production, agents' price) and VIFA(CGDS) (investment from imports, agents' price) to final expenditure (1→0).

## 3.2 Data preparation

Let $N_R$ and $N_I$ denote, respectively, the number of regions and production industries. GTAP distinguishes composite regions from nation-regions, because the former have international trade with themselves. Let $N_C$, $N_C \leq N_R$, be the *number of composite regions* and consider that composite regions are indexed before nation-regions. Let also energy (including transport) industry be indexed before other industries.

Let *P* and *M* denote, respectively, *product* and *margin*, and let *F* and *I* denote *final* and *intermediate*. Given the structure of the data available, the I-O disaggregate data is best described by $N_R N_I$ *production* industries, and $N_R N_I$ *direct import* industries (i.e., industry that deliver imports to final expenditure). Let production industries be indexed before direct import industries. The disaggregate I-O data is:

$$\mathbf{Q} = \mathbf{S} + \mathbf{T}, \mathbf{S} = \begin{pmatrix} \mathbf{S}^{\mathbf{I}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \mathbf{T} = \begin{pmatrix} \mathbf{T}^{\mathbf{I}} & \mathbf{T}^{\mathbf{F}} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} \mathbf{y}^{\mathbf{I}} \\ \mathbf{y}^{\mathbf{F}} \end{pmatrix}, \mathbf{v} = \begin{pmatrix} \mathbf{v}^{\mathbf{I}} \\ \mathbf{v}^{\mathbf{F}} \end{pmatrix} \text{ and } \mathbf{e}^{\mathbf{L}} = \begin{pmatrix} \mathbf{e}^{\mathbf{I}} \\ \mathbf{0} \end{pmatrix}.$$

Local emissions are assigned to production industries; final expenditure and added value have components from and to production industries and direct import industries.

The matrix of inter-industry flows, **Q**, is decomposed in a matrix of domestic flows **S** and in a matrix of international flows **T**. All domestic inter-industry flows are intermediate (i.e., from and to production industries). International flows comprise both intermediate and final (i.e., direct exports to final demand) flows. International flows are also decomposed in flows of products (goods and services) and transport margins. That is: $\mathbf{T}^{\mathbf{I}} = \mathbf{T}^{\mathbf{IP}} + \mathbf{T}^{\mathbf{IM}}$ and $\mathbf{T}^{\mathbf{F}} = \mathbf{T}^{\mathbf{FP}} + \mathbf{T}^{\mathbf{FM}}$.

Matrices **S** and **T** *record transactions at sellers' prices*, therefore entry (*ij*) shows the monetary payment that industry *i* receives from industry *j*. If industry *i* has to pay taxes on that transaction, they appear as a fraction of $v_i^I$.

Matrices $\mathbf{T^{IP}}$, $\mathbf{T^{IM}}$, $\mathbf{T^{FP}}$ and $\mathbf{T^{FM}}$ are to be estimated, $\mathbf{Q}$ and all vectors are known.

$\mathbf{y^I}$ accounts for domestic final demand, and results from the addition of tables VDPM, VDGM and row CGDS of table VDFA.

$\mathbf{y^F}$ accounts for imported final demand, and results from the addition of tables VIPM, VIGM and row CGDS of table VIFA.

$\mathbf{v^I}$ accounts for added value (factor payments plus taxes minus subsidies), and results from the addition of tables EFVA, OSEP, MFRV, XTRV and ISEP. (The sign of data from tables OSEP and ISEP had to be reversed, as it refers to subsidies.)

$\mathbf{v^F}$ accounts for taxes paid by final imports. Taxes on imports are provided in TFRV, without discrimination between final and intermediate imports. We assume that all such taxes are paid by final imports.

For intelligibility let $T^{*}_{((m-1)N_I+i)N_RN_I+((n-1)N_I+j)} \equiv T^{*(m,n)}_{(i,j)}$, where * = *IP*, *IM*, *FP* or *FM*.

Table $\mathbf{T^*}$ is composed by a total of $(N_R)^2$ square sub-matrices $\mathbf{T^{*(m,n)}}$ of length $N_I$ that describe trade from region *m* to region *n*. (And likewise for matrix $\mathbf{S^I}$.)

Table $\mathbf{S^I}$ is obtained from VDFA. Sub-matrices $\mathbf{S^{I(m,m)}}$, $m = 1,..., S_R$ are the regional I-O tables and $\mathbf{S^{I(m,n)}} = \mathbf{0}$, if $m \neq n$.

The structure of $\mathbf{Q^I}$, $\mathbf{T^{IP}}$, $\mathbf{T^{IM}}$, $\mathbf{T^{FP}}$ and $\mathbf{T^{FM}}$ is illustrated in Figure 3.

INSERT FIGURE 3 HERE

For all these four tables, sub-matrices $\mathbf{T^{*(m,m)}} = \mathbf{0}$, if $m > N_C$, corresponding to domestic trade of nation-regions. For the other sub-matrices, ($m, n = 1,..., N_R$) possible non-zero entries are the following:

1) All entries of $\mathbf{T}^{\mathbf{IP(m.n)}}$. ($\mathbf{T}^{\mathbf{IP}}$ is a "full" matrix, as opposed to the remaining three matrices, which are sparse);

2) The first row of $\mathbf{T}^{\mathbf{IM(m,n)}}$ and $\mathbf{T}^{\mathbf{FM(m,n)}}$, corresponding to transport services provided by the transport sector;

3) The main diagonal of $\mathbf{T}^{\mathbf{FP(m,n)}}$, corresponding to the import of product *i* from region *m* and its delivery to final demand of product *i* of region *n*.

The information of these matrices can be more conveniently stored in matrices $\widetilde{\mathbf{T}}^{\mathbf{IP}}$ and $\widetilde{\mathbf{T}}^{\mathbf{IM}}$ (both of size $(N_R N_I)^2$), and $\widetilde{\mathbf{T}}^{\mathbf{FP}}$ and $\widetilde{\mathbf{T}}^{\mathbf{FM}}$ (both of size $N_R.(N_R N_I)$). They are defined by:

1) $\widetilde{\mathbf{T}}^{\mathbf{IP}} = \mathbf{T}^{\mathbf{IP}}$;

2) $T_{1,j}^{IM(m,n)} = \sum_{i=1}^{N_I} \widetilde{T}_{ij}^{IM(m,n)}$, $j = 1,...,N_I$, $m,n = 1,...,N_R$ (since there are transport margins from the transaction of products involving non-transport sectors);

3) $T_{jj}^{FP(m,n)} = \widetilde{T}_{j}^{FP(m,n)}$ and $T_{1,j}^{FM(m,n)} = \widetilde{T}_{j}^{FM(m,n)}$, $j = 1,...,N_I$, $m,n = 1,...,N_R$ (therefore $\widetilde{\mathbf{T}}^{\mathbf{FP}}$ and $\widetilde{\mathbf{T}}^{\mathbf{FM}}$ are flattened versions of $\mathbf{T}^{\mathbf{FP}}$ and $\mathbf{T}^{\mathbf{FM}}$, from which null entries were squeezed out);

Note that in spite of the more compact formulation, all ~ matrices still have null entries, corresponding to nation-regions domestic trade. Those entries could be removed, but that gain would be out weighted by the loss of intelligibility.

**Construction of the constraint vector and aggregator matrix**

There are 6 partial constraint vectors that provide information on international flows.

1) Total imports for final demand (length $N_R N_I$): $\mathbf{k}^{\mathbf{1}} = \mathbf{y}^{\mathbf{F}} - \mathbf{v}^{\mathbf{F}}$ (total final expenditure minus taxes paid) equals column sums of $\mathbf{T}^{\mathbf{FP}}$ and $\mathbf{T}^{\mathbf{FM}}$. That is:

$$k^1_{(n-1)N_I+i} = \sum_{m=1}^{N_R} \left( \widetilde{T}_i^{FP(m,n)} + \widetilde{T}_i^{FM(m,n)} \right),\ i = 1,...,N_I,\ m,n = 1,...,N_R .$$

2) Net imports by firms on an industry basis (length $(N_R)(N_I)^2$): $\mathbf{k^2}$ is a linearization of table VIFA that specifies the imports from sector *i* to sector *j* of region *n*, but not the source region *m*. VIFA is measured at buyers' price so taxes on imports (ISEP2) must be subtracted from the total value (on an import sector and region basis). We assume that imports from all regions pay the same tax rate. $\mathbf{k^2}$ equals point-wise sums of columns of $\mathbf{T^{IP}}$ and $\mathbf{T^{IM}}$, that is:

$$k^2_{((n-1)N_I+i)N_RN_I+j} = \sum_{m=1}^{N_R} \left( \widetilde{T}_{ij}^{IP(m,n)} + \widetilde{T}_{ij}^{IM(m,n)} \right),\ i,j = 1,...,N_I,\ n = 1,...,N_R .$$

3) Margins on an export region basis (length $N_R$): $\mathbf{k^3}$ is table VST, which specifies the total of transport margins delivered by a region. $\mathbf{k^3}$ equals row sums of the regional sub-matrices of $\mathbf{T^{IM}}$ and $\mathbf{T^{FM}}$. That is:

$$k^3_m = \sum_{n=1}^{N_R} \left( \sum_{j=1}^{N_I} \left( \sum_{i=1}^{N_I} \widetilde{T}_{ij}^{IM(m,n)} + \widetilde{T}_j^{FM(m,n)} \right) \right),\ m = 1,...,N_R .$$

4) Margins on an import region and product (length $(N_R)^2$): $\mathbf{k^4}$ is table VTWR, that specifies the transport margin of import to industry *j* of region *n*, but neither source industry nor source region (because the region providing a transport margin can be different from the region providing the export). $\mathbf{k^4}$ equals row sums of $\mathbf{T^{IM}}$ and $\mathbf{T^{FM}}$ regional sub-matrices. That is:

$$k^4_{(1-N_I)n+j} = \sum_{m=1}^{N_R} \left( \sum_{i=1}^{N_I} \widetilde{T}_{ij}^{IM(m,n)} + \widetilde{T}_j^{FM(m,n)} \right),\ j = 1,...,N_I,\ n = 1,...,N_R .$$

5) Net bilateral imports (length $N_I(N_R)^2$): $\mathbf{k^5}$ is table VIWS (imports in world prices) minus VTWR (margins of those imports), that specifies the imports from region *m* to

industry *j* of region *n*, but not source industry. $\mathbf{k}^{5}$ equals column sums of regional sub-matrices of $\mathbf{T}^{\mathbf{IP}}$ and entries of $\mathbf{T}^{\mathbf{FP}}$. That is:

$$k^{5}_{(1-(N_R)^2)j+(1-N_R)k+l} = \sum_{i=1}^{N_I} \widetilde{T}_{ij}^{IP(k,l)} + \widetilde{T}_{j}^{FP(k,l)},\ j = 1,...,N_I,\ k,l = 1,...,N_R .$$

6) Bilateral exports (length $N_I(N_R)^2$): $\mathbf{k}^{6}$ is table VXWD (imports in world prices) that specifies the exports from industry *i* of region *m* to region *n*, but not target industry. Note that this way it is the seller that pays export taxes MFRV and XTRV. $\mathbf{k}^{6}$ equals row sums of regional sub-matrices of $\mathbf{T}^{\mathbf{IP}}$ and entries of $\mathbf{T}^{\mathbf{FP}}$. That is:

$$k^{6}_{(1-(S_R)^2)i+(1-S_R)m+l} = \sum_{j=1}^{S_I} \widetilde{T}_{ij}^{IP(m,n)} + \widetilde{T}_{j}^{FP(m,n)},\ i = 1,...,N_I,\ m,n = 1,...,N_R .$$

Recall that all entries corresponding to domestic trade of nation-states ( $m = n > N_C$ ) are zero.

Each non-zero international trade flow is subject to three constraints, as summarized in Table 2.

INSERT TABLE 2 HERE.

Table 2 also shows that the total magnitude of international trade is 7.14 $10^{12}$ USD 2001 and that the different constraints verify the same global balance. Table 3 also shows that margins represent only a marginal contribution to total trade (3.3%).

Therefore, the total constraint vector **k** is the concatenation of the 6 partial constraint vectors, of total length $(N_R N_I) + N_R(N_I)^2 + N_R + (N_R)^2 + 2N_I(N_R)^2$. The estimated vector **t** is the concatenation of the four linearized ~ matrices and has total length $N_T = 2(N_R N_I)^2 + 2N_I(N_R)^2$. The aggregator matrix **G** that relates them is obtained from the set of constraints displayed in the previous paragraphs.

Finally, let $\mathbf{t}^{\mathbf{r}}$ be the vector of row sums of $\mathbf{T}^{\mathbf{I}}$ and $\mathbf{T}^{\mathbf{F}}$, with length ($N_R N_I$), obtained from the row sums of VIWS, and $\mathbf{s}^{\mathbf{r}}$ and $\mathbf{s}^{\mathbf{c}}$ be row and column sums of $\mathbf{S}$. Total output of the intermediate sector is $\mathbf{x}^{\mathbf{I}} = \mathbf{t}^{\mathbf{r}} + \mathbf{s}^{\mathbf{r}} + \mathbf{y}^{\mathbf{I}}$.

Let $\mathbf{t}^{\mathbf{c}}$ be the vector of column sums of $\mathbf{T}^{\mathbf{I}}$ and $\mathbf{T}^{\mathbf{F}}$, with length ($2N_R N_I$). The second half of this vector must equal ($\mathbf{y}^{\mathbf{F}} - \mathbf{v}^{\mathbf{F}}$) – direct imports – and the first half must equal the column sums of VIFA minus ISEP2 – imports by firms net of import taxes/subsidies. Total input of the intermediate sector is $\mathbf{x}^{\mathbf{I}} = \mathbf{t}^{\mathbf{Ic}} + \mathbf{s}^{\mathbf{c}} + \mathbf{v}^{\mathbf{I}}$, where $\mathbf{t}^{\mathbf{Ic}}$ is the first half of $\mathbf{t}^{\mathbf{Ic}}$.

Vectors $\mathbf{t}^{\mathbf{r}}$ and $\mathbf{t}^{\mathbf{c}}$ are the RAS constraints required to balance the estimated matrix $\mathbf{T}$.

The information used by the 3 estimation methods is summarized in Table 3.

INSERT TABLE 3 HERE.

Method 2 uses a subset of constraints $\mathbf{k}$. Aggregate bilateral imports (a square matrix $\mathbf{T}^{\mathbf{R}}$ of length $N_R$) are obtained from $\mathbf{k}^{5}$ aggregating over industries. The set of intermediate imports by industry is $\mathbf{k}^{2}$. Intermediate imports are obtained by multiplying trade shares with intermediate imports by industry. Direct imports to final demand are obtained by multiplying trade shares with imports to final demand by industry ($\mathbf{y}^{\mathbf{F}} - \mathbf{v}^{\mathbf{F}}$). Recall Section 2.2 for the details of this procedure.

For methods 1 to 3 margins were lumped with the product they transport.

## 3.3 Numerical calculations

The problem at hand is the determination of $\boldsymbol{\lambda}$, in the state-space $(-\infty,+\infty)^{S_K}$ that satisfies $\mathbf{k} = \mathbf{G}\mathbf{t}$, $\mathbf{t} = f(\boldsymbol{\Lambda},\boldsymbol{\theta})$, where $f(\cdot;\cdot)$ is given by Eq. (6), and $\boldsymbol{\Lambda} = \mathbf{G}'\boldsymbol{\lambda}$, and $\mathbf{k}$, $\mathbf{G}$ and $\boldsymbol{\theta}$ (given by Eq. 4) are known.

This is a non-linear algebraic system of equations that must be solved numerically. We approach this problem by defining potential $\pi(\boldsymbol{\lambda})$,

$$\pi(\boldsymbol{\lambda}) = (N_K)^{-1}(\boldsymbol{\chi})^2, \text{ and } \boldsymbol{\chi} = (\mathbf{1} - (\mathbf{G}\,\mathbf{t}) \div \mathbf{k}). \quad (8)$$

This potential, $\pi > 0$, is the normalized sum of the squares of the relative errors of each constraint, and $\boldsymbol{\chi}$ is vector of relative error of the constraints. If the relative error of each constraint is 100%, e.g., if $\boldsymbol{\lambda} = \mathbf{0}$, then the potential is 1, $\pi(\mathbf{0}) = 1$. If the relative error of each constraint is 0%, $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ that satisfies $\mathbf{k} = \mathbf{G}\mathbf{t}^*$, then $\boldsymbol{\chi} = \mathbf{0}$ and the potential is 0, $\pi(\boldsymbol{\lambda}^*) = 0$.

The definition of the potential as a sum of relative errors gives the same weight to verification of every constraint. The potential could be defined as a sum of absolute errors, yielding more weight to larger constraints. The solution reached should be close using either option but convergence is faster using relative errors.

We want to find $\boldsymbol{\lambda}^*$, such that $\pi(\boldsymbol{\lambda}^*) = 0$. This is a minimization (or optimization) problem, a class of problems for which different methods are available. Local optimization analytical methods (such as Newton-Raphson and its extension, Press et al., 1992) are not feasible because it is not possible to compute the Jacobian analytically – the Jacobian is the multi-dimensional slope around a solution, $\{\partial\pi/\partial\lambda_i\}_{i=1}^{N_K}$. Local numerical deterministic methods operate by estimating the Jacobian and therefore also unpractical because that estimation is computationally very intensive, given the large dimensionality of the problem (Beveridge and Schechter, 1970). Therefore we investigated stochastic algorithms.

A local stochastic minimization algorithm works as follows. Let $\boldsymbol{\lambda}_i$ be a solution. Generate stochastically a displacement vector $\Delta\boldsymbol{\lambda}$. If $\pi(\boldsymbol{\lambda}_i+\Delta\boldsymbol{\lambda})<\pi(\boldsymbol{\lambda}_i)$, that is, if the

displacement leads to a solution of lower potential, the new solution is accepted: $\lambda_{i+1}=\lambda_i+\Delta\lambda$. Otherwise, a new displacement vector $\Delta\lambda$ is generated. This process is repeated until the solution converges to a local minimum. A global optimization algorithm avoids getting trapped in a local minimum by occasionally accepting a displacement that raises the potential of the solution. These algorithms belong to the Metropolis Monte-Carlo family (such as simulated annealing) (Metropolis et al., 1953; Solis and Wets, 1981). These methods did not prove effective in our case. We believe this happens because local minima tend be located wide apart in the phase space, because Eq. (7) is the inverse of a sum. Therefore tracking a local minimum does not necessarily bring the solution closer to the global minimum.

After testing several algorithms, we found that a genetic algorithm (Holland, 1975) proved successful in our problem. A genetic algorithm works as follows. There is a mother population of individual solutions, where each individual is defined by a genome or sequence of chromosomes. Individuals exchange chromosomes (cross-over) and suffer mutations, giving birth to a daughter population. The daughter population is ranked, and the fittest individuals replace the least fit. We believe that the genetic algorithm is successful because it explores the "natural" topology of the problem. By specifying chromosomes as coordinates in the phase space (entries of vector $\lambda$), the cross-over operation in fact explores vertices in a hypercube, whose size is altered by mutation. Therefore, the algorithm is able to efficiently explore the state-space.

Before describing the genetic algorithm proper we describe with some detail the calculation of potential $\pi(\lambda)$, which is the most computationally intensive operation of the algorithm.

Numerical calculations were performed in FORTRAN 90. Software Mathematica was used for processing and analyzing data. Source code is available from the authors upon request.

**Calculation of the potential**

Recall from the previous Subsection that $N_K$ (=$N_R$ ($N_I$(1 + $N_I$) + 1 + $N_R$(1 + 2$N_I$)) is the length of vectors **k** and $\boldsymbol{\lambda}$; and $N_T$ (=2($N_R^2 N_I$)(1 + $N_I$)) is the length of vector **t**.

These vectors can be quite large. For example if $N_R$ = 80 and $N_I$ = 3, then $N_K$ = 45840 and $N_T$ = 153600. If $N_R$ = 87 and $N_I$ = 57 (the full GTAP model), $N_K$ = 1158144 and $N_T$ = 50046228. So the problem is computationally demanding.

The range of constraints values is quite large, if $N_R$ = 80 and $N_I$ = 3; max(**k**) = 5.482x$10^{11}$ USD 2001 and min(**k**) = 1.x$10^2$ USD 2001, thus spanning 9 orders of magnitude. Inverse calculation involving numbers much larger and much smaller than 1 can lead to computational errors, and so all calculations were performed in a $10^{12}$ USD 2001 basis, so that all constants are smaller than 1.

For computational reasons, real vector $\boldsymbol{\lambda}$ was stored as a $2N_T$-vector $\boldsymbol{\gamma}$, whose entry $\gamma_i \in (0,1)$, $i = 1,...,2N_T$. $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are related as:

$\lambda_i = h(\gamma_i; \gamma_{i+N_K}) = h_1(\gamma_i) h_2(\gamma_{i+N_K})$, with:

$$h_1(\gamma_i) = \log\left(\frac{\gamma_i}{1-\gamma_i}\right) \text{ and}$$

$h_2(\gamma_{i+N_K}) = (\gamma_{i+N_K})^{-(\gamma_{i+N_K})^{-1}}$, with $i = 1,..., N_K$.

Figure 4 shows the behavior of function $h(\cdot;\cdot)$.

INSERT FIGURE 4 HERE.

Function $h_1(\cdot)$ is a mapping from (0,1) to $(-\infty, \infty)$, such that in half of its range (from $\gamma_i = .25$ to $\gamma_i = .75$), the region from $-\log(3)$ to $\log(3)$ is explored.

Function $h_2(\cdot)$ sets the scale at which function $h_1(\gamma_i)$ explores the phase-space. This scale is fine grained if $\gamma_{i+SK}$ is large and becomes coarser as it decreases ($h_2(1) = 1$; $h_2(.5) = 4$; $h_2(.25) = 256$, $h_2(.1) = 10^{10}$ and $h_2(.01) = 10^{200}$).

Working with $\boldsymbol{\gamma}$ instead of $\boldsymbol{\lambda}$ has two main advantages. First, the range of each coordinate is (0,1), which is more tractable than $(-\infty, \infty)$, especially given that we systematically use random numbers, generated from a uniform distribution of range (0,1).

Second, $\boldsymbol{\gamma}$ describes every coordinate in the $\boldsymbol{\lambda}$ space-state with a fine grained and a coarse grained parameter. Tuning the fine grained parameter operates a local exploration while tuning the coarse grained parameter operates a quick exploration of a distant region in the state space.

Other functional forms also provide this behavior; the specific functional form of $h(\cdot;\cdot)$ was chosen because it provided good results.

From solution $\boldsymbol{\gamma}$, application of $h(\cdot;\cdot)$ leads to $\boldsymbol{\lambda}$, and with stored vector $\boldsymbol{\theta}$ and matrix **G**, vector $\boldsymbol{\Lambda}$ is computed. Application of function $f(\cdot;\cdot)$ (Eq. 7) leads to **t**, and application of **G'** and **k** leads to constraint relative error vector $\boldsymbol{\chi}$ (Eq. 8), and from that scalar potential $\pi$.

Function $f(\cdot;\cdot)$ (Eq. 7) is inconvenient for numerical purposes, because it is discontinuous at the origin and because it may involve the calculation of the exponential of large numbers. In numerical calculations we replaced Eq. (7) by the branch equation:

$$\begin{cases} f_1(\phi_i;\theta_i) = \theta_i\left(\dfrac{1}{\phi_i} - \dfrac{\exp(-\phi_i)}{1-\exp(-\phi_i)}\right) & , \quad \phi_i > .1; \\ f_2(\phi_i;\theta_i) = \theta_i\left(\dfrac{1}{2} - \dfrac{\phi_i}{12}\right) & , \quad -.1 \le \phi_i \le .1; \\ f_3(\phi_i;\theta_i) = \theta_i\left(\dfrac{1}{\phi_i} - \dfrac{1}{\exp(\phi_i)-1}\right) & , \quad \phi_i < -.1. \end{cases}$$

The expression of the middle branch was obtained from a 2nd order Taylor expansion around the origin. This branch equation ensures that only the exponential of small numbers is calculated.

Therefore, the algorithm that computes $\boldsymbol{\gamma}$ to $\pi$ requires as input the location vector $\boldsymbol{\gamma}$, and data $\boldsymbol{\theta}$, **G** and **k**. Intermediary vectors computed are $\boldsymbol{\lambda}$, $\boldsymbol{\Lambda}$, **t** and $\boldsymbol{\chi}$.

The most computationally demanding operations of this algorithm are the two matrix-vector multiplications, that can be optimized using the following technique, used in the field computational fluid mechanics.

Let *g* be the total of non-empty entries of **G**. Since matrices **G** and **G'** are sparse, it is convenient not to perform the full matrix-vector operation (that would involve $N_K N_T$ operations) but to multiply only its non-null entries (involving $g << N_K N_T$ operations).

This is achieved by compressing matrices **Z** and **Z'** into vectors $\mathbf{z}^{\mathbf{c}}$ and $\mathbf{z}^{\mathbf{r}}$ (respectively of length $N_T$ and $N_K$) and $\mathbf{z}^{\mathbf{C}}$ and $\mathbf{z}^{\mathbf{R}}$ (both of length *z*).

Vectors $\mathbf{z}^{\mathbf{c}}$ and $\mathbf{z}^{\mathbf{r}}$ state how many non-zero entries exist in each column and row of **Z**. Vectors $\mathbf{z}^{\mathbf{C}}$ and $\mathbf{z}^{\mathbf{R}}$ state where those entries are located.

That is, the first $z^r_1$ entries of $\mathbf{z}^{\mathbf{R}}$ are the *i* values for which $Z_{1,i} = 1$. Entries $z^r_1 + 1$ to $z^r_1 + z^r_2$ are the *i* values for which $Z_{2,i} = 1$. And so forth.

$\mathbf{z}^{\mathbf{c}}$ and $\mathbf{z}^{\mathbf{C}}$ are defined in an analogous way to $\mathbf{z}^{\mathbf{r}}$ and $\mathbf{z}^{\mathbf{R}}$.

**Genetic algorithm**

Let $N_P$ be the number of elements in the population. Let $\boldsymbol{\Gamma}$ be a matrix whose $N_P$ rows are $(2N_K)$-vectors that represent individual solutions $\boldsymbol{\gamma}^{\mathbf{i}}$. Let $\boldsymbol{\pi}$ denote the vector of length $N_P$ that indexes the potential of each individual solution and $\boldsymbol{\rho}$ the vector of length $N_P$ that ranks the solutions in ascending order; i.e., the best individual (with lowest potential) has rank $N_P$, the worst individual (with highest potential) has rank 1.

Let $\boldsymbol{\Gamma}_{\mathbf{w}}$ be the mother generation. A daughter generation is obtained from the mother generation via 3 operations.

*Cross-over*: A total of $p_C N_P/2$ pairs of individuals is selected. For a selected pair of individuals $m$ and $n$, every coordinate $i$, $i = 1,...,N_T$, is exchanged with probability $p_X$.

*Mutation*: In every cross-over event (exchange of coordinates) a mutation occurs with probability $p_M$. If a mutation occurs in coordinate $i$ of individual $n$, $\gamma_i^n$ is replaced by $\gamma *_i^n$ defined as:

$$\gamma *_i^n = \gamma_i^n \left(1 + 4 p_R (2\omega - 1)\gamma_i^n \left(1 - \gamma_i^n\right)\right),$$

where $p_R$ is a parameter of range (0,1) and $\omega$ is a random number taken from a uniform distribution in range (0,1). This formulation for the mutation range ensures that $\gamma *_i^n \in (0,1)$, irrespective of the value of $\gamma_i^n$. The larger parameter $p_R$ is, the larger is the mutation range.

*Translocation*: In every cross-over event a translocation occurs with probability $p_T$. If a translocation occurs in coordinate $i$ of individual $n$, a coordinate $j$ of the same individual is chose with uniform probability and coordinates $i$ and $j$ are exchanged.

After performing these operations to all individuals and coordinates it becomes necessary to update the potential and rank vector. The potential vector is updated using the technique described under the heading "calculation of the potential" above in this Subsection. The rank vector is updated using a conventional ranking algorithm.

Now the operation of *selection* is performed. We consider that the single best solution of the mother generation is always selected (elitist selection). We also consider that the $p_S N_P$ worst individuals are replaced by copies of the $p_S N_P$ best individuals, where $p_S$ is a selection parameter.

A daughter generation $\mathbf{\Gamma_{w+1}}$ has been obtained. The algorithm proceeds by repeating these steps until there is convergence to the global minimum. The best solution and corresponding potential of every generation $m$, $\mathbf{\Lambda^{*m}}$ and $\pi^{*m}$, are stored, thus recording the history of the algorithm's travel in the state-space. (It is more convenient to store $\mathbf{\Lambda^{*m}}$ rather than the corresponding $\boldsymbol{\gamma}$ or $\boldsymbol{\lambda}$ vectors it is more compact).

When a local minimum is found (i.e., no displacement occurs for a consecutive number of generations, which we defined as 50), the whole population is replaced by mutants (with only a small fraction of mutations and translocations) without elitist selection, so that a worse individual is chosen. This operation successfully allowed the algorithm to escape local minima (in which the algorithm rarely became trapped in the first place).

We considered that global convergence occurred when two conditions were simultaneously satisfied.

One condition is that there is convergence in the first place. That is, if $m >> n$, then $\pi^m < \pi^n$ but $\mathbf{t^m} \approx \mathbf{t^n}$. More specifically, we imposed that only if $\pi^m / \pi^n < .1$ and

$\max\left\{\left|1 - t_i^m / t_i^n\right|\right\}_{i=1}^{N_T} < .1$, we accepted convergence (less than 10% error after 90% potential reduction).

The second condition is that convergence satisfies the global constraints (row and column sums). Thus we accepted the solution if $\left(1 - \left(\mathbf{t}^{\mathbf{r}}, \mathbf{y}^{\mathbf{F}} - \mathbf{v}^{\mathbf{F}}\right) / \left(\mathbf{t}^{\mathbf{c}}\right)\right)^2 / \left(2N_R N_I\right) < 0.001$. That is, if the sum of relative error of the global constraints was less than 0.1%.

We note that minimization was occurring on potential $\pi$, thus there is no guarantee that such a close approximation to the global constraints occurs.

The genetic algorithm involves 7 parameters. Population size, integer $N_P$, and six real parameters of range (0,1). The parameters reported in Table 5 provide fast convergence.

INSERT TABLE 4 HERE.

The total computation time to arrive at the global minimum was of the order of 1 week, running in a single processor in the case ($N_R$ = 80 and $N_I$ = 3).

The most computationally intensive part of the algorithm are matrix-vector multiplication, with large matrices and vectors. For such operations, computation time increases super-linearly with the size of matrices and vectors. Thus parallel computation is expected to accelerate the algorithm, since every processor operates with smaller matrices and vectors.

Thus, for the general problem ($N_R$ = 87 and $N_I$ = 57), which we have not performed, computation should run in parallel.

# 4 Results

We compare the estimation methods; we analyze GHG embodied emissions and intensities, as provided by the MEP method; and we discuss the results.

## 4.1 Comparison of estimation methods

We calculated the GHG embodied emissions and GHG intensity of industries for 80 regions and 3 industries per region. We calculated upstream quantities relative to domestic production and to direct imports; and downstream quantities relative to domestic production. Thus, there is a total of 720 (= 80x3x3) estimated embodied emissions and 720 estimated intensities.

We used 4 different methods to estimate the matrix of international trade **T**. Method 1 and 2 (the "homogeneous intensity" method and the "trade share" method) are presented in Section 2.2. Method 3 is the MEP method using the average estimated trade flows, as given by Eq. 7. Method 4 is the full MEP method, described as follows.

100 random copies of the **T** matrix are generated, using Eq. 6, and the estimated quantities are computed each time for each copy. From this sample, the expected value and the 90% confidence intervals (CI) were computed. (The 90%-CI lower and upper values are, respectively, the realizations that are smaller and larger than 95% of all realizations).

Figures 5 and 6 show the scatter-plots and linear regressions of the logarithm of 720 estimated embodied emissions and the logarithm of 720 intensities using methods 1, 2 and 3 against the average from method 4.

INSERT FIGURE 5 HERE.

INSER FIGURE 6 HERE.

We see from Figures 5 and 6 that the correlation between methods 1 and 4 is very weak, that the correlation between methods 2 and 4 is better but still weak and that the correlation between methods 3 and 4 is very good. This shows that method 3 – computationally less demanding – is an excellent approximation to method 4, while methods 1 and 2 are less so.

Now we examine whether the differences between methods are within the 90%-CIs. Figure 7 shows the 240 *relative* estimated upstream embodied emissions of consumption from domestic production, of the expected values of all methods and the 90%-CIs of method 4. We now explain the meaning of this graphic.

Let $e_{ik}$ be an arbitrary estimated quantity $i$, as estimated by method $k$. The graphic shows $(1 - e_{i4}/ e_{ik})$. Therefore, if method $k$ and method 4 yield the same value, the graphic shows 0%, if method $k$ yields a value 2 times large as method 4, the graphic shows + 100%. The data points are ranked by decreasing magnitude of the 90%-CI.

INSERT FIGURE 7 HERE.

Figure 7 shows that, for *circa* 90% of data points, the 90%-CI is within a 20 % range of the average value. Method 1 is more often than not outside the 90%-CI; method 2 is often outside those bounds and method 3 is almost always inside them.

Method 3 is almost always within the 90%-CI of Method 4 and that Methods 1 and 2 are often outside those bounds.

INSERT FIGURE 8 HERE.

Figure 8 shows the magnitude of the confidence bounds, as a function of the size rank of the estimated embodied emissions. That is, for 90% of 100% of the largest estimated data points, the 90%-CIs have a relative width smaller than 50%. For 50%

of 100% of the largest estimated data points, the 90%-CIs have a relative width smaller than 20%. For 50% of 50% of the largest estimated data points, the 90%-CIs have a relative width smaller than 10%.

However, many of the estimated data points are small, if we look at accumulated total embodied emissions the picture is as follows. For 80% of the data points accounting for 90% of total estimated embodied emissions, the 90%-CIs have a relative width smaller than 20%. For 50% of the data points accounting for 90% of total estimated embodied emissions, the 90%-CIs have a relative width smaller than 10%.

Thus, we see that the confidence intervals are narrow.

## 4.2 GHG embodied emissions

We aggregated the GHG embodied emissions of all industries of a region to define its consumer GHG responsibility (upstream embodied emissions of final expenditure), producer GHG responsibility (downstream embodied emissions of added value) and total GHG responsibility (arithmetic average of the other two). The world total of each of these three quantities is the total sum of local emissions, $23.736 \times 10^9$ $KgCO_2$-equivalent.

Tables 5, 6 and 7 compare consumer, producer and total responsibility against local emissions, respectively in 2, 8 and 80-region models. All tables display the share of total emissions of each region. Table 7 also displays the ranking of regions, in decreasing order.

In the 2-region model, the two regions are Annex I and Annex II countries as defined in the Kyoto Protocol. The details of the aggregation of the 8 and 80-region models can be found in the Appendix.

INSERT TABLE 5 HERE.

INSERT TABLE 6 HERE.

INSERT TABLE 7 HERE.

The ranking of regions does not change significantly, according to the indicator used (local vs. embodied emissions). However, the fraction that each region has of total emissions does change significantly.

Figures 9 and 10 (and Table 8, the key to Figure 10) show the difference of embodied to local emissions, relative to local emissions. E.g., +10% means that the embodied emissions of that region are 10% larger than its local emissions.

INSERT FIGURE 9 HERE.

INSERT FIGURE 10 HERE.

INSERT TABLE 8 HERE.

Figures 9 and 10 show that the difference from embodied to local emissions can be large and that it can differ across regions.

## 4.3 GHG emission intensities

GHG emission intensity is an amount of emissions by economic value (measured in Kg $CO_2$-eq/USD2001). We examine four types of intensities: local GHG emission intensities are local emissions over total output; upstream intermediate GHG emission intensities are upstream emissions embodied in domestic products over final expenditure of such products; upstream final GHG emission intensities are upstream emissions embodied in products imported for final demand over final expenditure of such products; downstream GHG emission intensities are downstream emissions embodied in the payments of domestic products over value added of such products.

Local intensities are defined for regions; other intensities are defined for industries (or products) and average intensities can be defined for regions. Table 9 shows local intensities and the ranking of regions according to local and average intensities in decreasing order.

INSERT TABLE 9 HERE.

Figures 11 to 13 show the industry-level upstream intermediate, upstream final and downstream intensities.

INSERT FIGURE 11 HERE.

INSERT FIGURE 12 HERE.

INSERT FIGURE 13 HERE.

In Figures 11 and 13 we see that there are significant variations in the intensities of domestic products across regions. We also see that the energy sector is much more intensive than either agriculture or services.

In Figure 12 we see that there is a smaller variation of intensities across regions for imported products. This happens because the intensity of imported products is an average of the intensities of the domestic products of the exporting regions. Figure 12 also shows that imported agricultural products have much higher intensity than domestic agricultural products. This happens because international transport margins are an important component of the final cost of imported agricultural products, thus approximating the intensity of agricultural products to the intensity of energy products.

## 4.4 Discussion

The main qualitative patterns identified in this study are as follows:

Regarding the comparison of estimation methods, the homogeneous intensity method yields results very different from the full MEP method, the trade share method a bit less so and the average MEP method yields results very similar to the full MEP method. The full MEP method itself yields narrow confidence intervals, with 80% of data points corresponding to 90% of global emissions having 90%-CI smaller than 20% of the expected value and 50% of data points corresponding to 90% of global emissions having 90%-CI smaller than 10% of the expected value.

Regarding GHG emissions, the ranking of regions either by local or embodied emissions does not change significantly but the fraction of total emissions allocated to each region does. Regarding GHG intensities, the energy sector is much more intensive than either agriculture or services, the upstream intensity of imported products has a smaller variation range than domestic produced ones and imported agricultural products are significantly more intensive than domestically produced ones.

We believe that these qualitative patterns are robust, but that the quantitative details of our results are not, because of the data used. A minor point is that the data refers to 2001, thus not reflecting important changes that occurred in recent years (mainly the growth of China's role in global emissions and economy). A major point refers to the high level of industry aggregation used.

The data set used (80 regions and 3 industries) is appropriate for the main purpose of this study: to compare the performance of different methods of estimation of international flows. However, it is not appropriate to make policy inferences because it neglects the whole network of domestic transactions. For an accurate estimation of either region or industry-level intensities or embodied emissions it makes more sense to consider a larger number of industries and a smaller number of regions, and

defining regions such that the main trade partners of the main region under consideration are disaggregated. In fact this is what is commonly done (Hertwich and Peters, 2006), because it reduces the error resulting from aggregation and loss of homogeneity (Murray, 1998; Lenzen, 2001; Lahr and Stevens, 2002).

Therefore we make the cautionary remark that the errors resulting from the estimation of international inter-industry flows can be small compared to the errors from aggregation and from the original source data.

# 5 Final remarks

In the present paper we have presented a new method for the estimation of international inter-industry flows, based on the Maximum Entropy Principle (or MEP, Jaynes, 1957). Unlike existing methods, this method allows for the incorporation of arbitrary information in the estimation process, and for the quantification of the estimation error.

This method involves calculating the solution of a large non-linear algebraic system of equations, and we therefore also presented a genetic optimization algorithm that implements the estimation method.

Using the GTAP database as source data, we developed a 80-region 3-industry model that we used as a case-study to compare GHG embodied emissions and intensities, computed using different estimation methods.

We observed that the conventional methods (homogeneous intensity and trade share methods) are not good approximations of the full MEP method, but that a simplification using average values is. We also concluded that the confidence interval of most estimated data points is small.

As directions of future research we intend to provide a matrix of international inter-industry flows estimated with the MEP method, for the whole GTAP dataset (87 regions and 57 sectors) and the quantification of source data and aggregation errors.

## Acknowledgements

We acknowledge the support of FCT via scholarship SFRH/BD/9055/2002 (to J. R.) and grant POCTI/MGS/47731/2002.

We thank Rosa Trancoso, João Paulo Magalhães and Marco Reis for invaluable help in programming.

## References

Ahmad, N., Wyckoff, A.W. (2003) Carbon dioxide emissions embodied in international trade of goods, STI Working Paper DSTI/DOC, 15. Organisation for Economic Co-operation and Development (OECD), Paris, France.

Beveridge, G. S. G., Schechter, R. S. (1970) Optimization: theory and practice. (New York, McGraw-Hill).

Dimaranan, B. V., Editor (2006) Global Trade, Assistance, and Production: The GTAP 6 Data Base (Center for Global Trade Analysis, Purdue University).

Duchin, F. (2005) A world trade model based on comparative advantage with m regions, n goods, and k factors, Economic Systems Research, 17, pp. 141–162.

Ghosh, A. (1958) Input-output approach in an allocation system, Economica, 25, pp. 58-64.

Gilchrist, D. A., St-Louis, L. V. (2004) An algorithm for the consistent inclusion of partial information in the revision of input–output tables, Economic Systems Research, 16 (2), pp. 149-156.

Hertwich, E. G. (2005) Life cycle approaches to sustainable consumption: a critical review, Environmental Science and Technology, 39 (13), pp. 4673-4684.

Hoekstra, R., Janssen, M.A. (2006) Environmental responsibility and policy in a two-country dynamic input–output model, Economic Systems Research, 18, pp. 61–84.

Holland, J. H. (1975) Adaptation in Natural and Artificial Systems (Ann Arbor, University of Michigan Press).

Jaynes, E. T. (1957) Information theory and statistical mechanics I, Physical Review, 106, pp. 620-630.

Jaynes, E. T. (1983) Papers on Probability, Statistics and Statistical Physics (Dordrecht, Kluwer Academic Publishers).

Kainuma, M., Matsuoka, Y., Morita, T. (2000) Estimation of embodied $CO_2$ emissions by general equilibrium model, European Journal of Operational Research, 122 (2), pp. 392-404.

Lahr, M. L., Stevens, B. H. (2002) A study of the role of regionalization in the generation of aggregation error in regional input-output models, Journal of Regional Science, 42 (3), pp. 477-507.

Lee, P. M. (1989) Bayesian statistics: an introduction (New York, Oxford University Press).

Lenzen, M. (2001) Error in conventional and input-output-based life-cycle inventories, Journal of Industrial Ecology, 4 (4), pp. 127-148.

Lenzen, M., Pade, L.-L., Munksgaard, J. (2004) $CO_2$ multipliers in multi-region input–output models, Economic Systems Research, 16, pp. 391–412.

Leontief, W. (1970) Environmental repercussions and the economic structure: an input-output approach, Review of Economics and Statistics, 52 (3), pp. 262-271.

Lutz, C., Meyer, B., Wolter, M.I. (2005) GINFORS-Model. Gesellschaft für Wirtschaftliche Strukturforschung mbH (GWS), Osnabrück, Germany. MOSUS Workshop, IIASA Laxenburg; April 14.–15.

McDonald, S., Thierfelder, K. (2004) Deriving a global social accounting matrix from GTAP Versions 5 and 6 data. GTAP Technical Paper No. 22. Center for Global Trade Analysis, Purdue University.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A. H., Teller, E. (1953) Equation of State Calculations by Fast Computing Machines, Journal of Chemical Physics, 21, pp. 1087-1092.

Miller, R.E., Blair, P.D. (1985) Input-Output Analysis: Foundations and Extensions (Englewood Cliffs, NJ, Prentice-Hall).

Munksgaard, J., Wier, M., Lenzen, M., Dey, C. (2005) Using Input-Output analysis to measure the environmental pressure of consumption at different spatial levels, Journal of Industrial Ecology, 9 (1-2), pp. 169-185.

Murray, A. T. (1998) Minimizing aggregation error in input-output models, Environment and Planning A, 30 (6), PP. 1125-1128.

Peters, G. P., Hertwich, E. G. (2006) Structural analysis of international trade: environmental impacts of Norway, Economic Systems Research, 18, pp. 155–181.

Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. (1992) Numerical Recipes in Fortran: The Art of Scientific Computing (Cambridge, Cambridge University Press).

Proops, J. L. R., Atkinson, G., Schlotheim, B. F. v., Simon, S. (1999) International trade and the sustainability footprint: a practical criterion for its assessment, Ecological Economics, 28, pp. 75-97.

Rodrigues, J., Domingos, T., Giljum, S., Schneider, F. (2006) Designing an indicator of environmental responsibility, Ecological Economics, 59 (3), pp. 256-266.

Schmitt, L. M. (2001) Theory of genetic algorithms, Theoretical Computer Science, 259, pp. 1-61.

Solis, F. J., Wets, R. J.-B. (1981) Minimization by random search techniques, Mathematical Operations Research, 6, pp. 19-30.

Thomas, M. U., Barr, D. R., 1977. An approximate test of Markov chain lumpability. Journal of the American Statistical Association, 72 (357): 175-179.

Tukker, A., Jansen, B. (2006) Environmental impacts of products: a detailed review of studies, Journal of Industrial Ecology, 10 (3), pp. 159-182.

Tribus, M. (1961) Thermostatics and Thermodynamics (Princeton, NJ, D. Van Nostrand Company, Inc.).

Wiedmann, T., Lenzen, M., Turner, K., Barrett, J., (2007) Examining the global environmental impact of regional consumption activities — Part 2: Review of input–output models for the assessment of environmental impacts embodied in trade, Ecological Economics, 61, pp. 15-26.

United Nations (1994) System of National Accounts 1993. Studies in Methods, Series F., No. 2, Rev. 4, Sales No. 94.XVII.4 (New York, United Nations).

United Nations (1999) Handbook of input-output compilation and analysis. Studies in Methods: Handbook of national accounting. ST/ESA/STAT/SER.F/74 (New York, United Nations).

UNFCCC (2005) Key GHG Data: Greenhouse Gas (GHG) Emissions Data for 1990-2003 submitted to the UNFCCC. (Bonn, United Nations Framework Convention on Climate Change). http://unfccc.int/resource/docs/publications/key_ghg.pdf

## Appendix – Industry and region aggregation

The equivalence between industries is presented in Table A1.

INSERT TABLE A1 HERE.

Regarding the number of regions, there were GTAP regions for which no emissions were reported, and that were therefore aggregated into other GTAP regions as follows: Hog-Kong into chn (China); Taiwan into xea (rest of East Asia); xna (Bermuda; Greenland; Saint Pierre and Miquelon) into xcb (rest of Caribbean); Venezuela into xsm (rest of South America); Cyprus into xer (rest of Europe); Turkey into xme (Rest of Middle East) and Mozambique into xss (rest of Sub-Saharan Africa). Table A2 shows the legend of the reference codes and the reference year of emissions (composite regions may have different emission years, always in period 1990-2003). Table A3 lists the countries aggregated in composite regions.

INSERT TABLE A2 HERE.

INSERT TABLE A3 HERE.

In the paper, besides the 80 region model, we considered two aggregate models, with 2 and 8 regions. The 2 region model corresponds to Annex I and Annex II countries, as defined in the Kyoto Protocol (UNFCCC, 2005).

The rest of the former Soviet Union was aggregated in Annex I, while in fact only the Russian Federation and Ukraine belong to Annex I under the Kyoto Protocol. This option was taken because Ukraine is aggregated with the other former member of the Soviet Union as xsu, in the GTAP database.

Cyprus is aggregated in XAI instead of EU27, because in the 80 region model it was aggregated in sector xer.

Table A4 identifies the aggregated regions.

INSERT TABLE A4 HERE.

## Figures

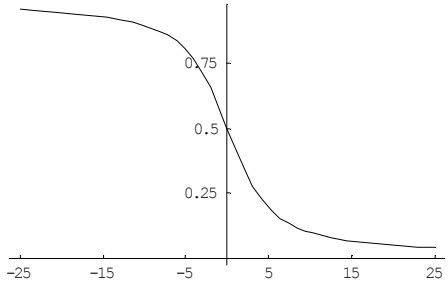
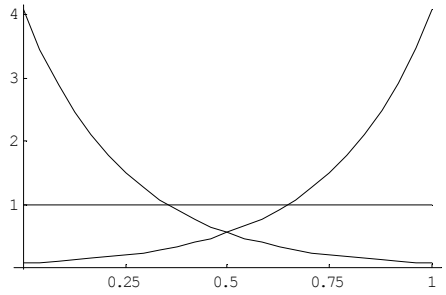



**Figure 1.** Figure 1 shows $t_i/\theta_i$ as a function of $\Lambda_i\theta_i$ (Eq. 6). $t_i/\theta_i$ is bounded between 0 and 1.




**igure 2.** Figure 2 shows $p(\omega)$ as a function of $\omega$, assuming $\theta_i = 1$, for 3 values of $\Lambda_i$. The upward sloping curve is for $\Lambda_i = -4$; the downward sloping curve for $\Lambda_i = 4$ and the flat curve for $\Lambda_i = 0$.

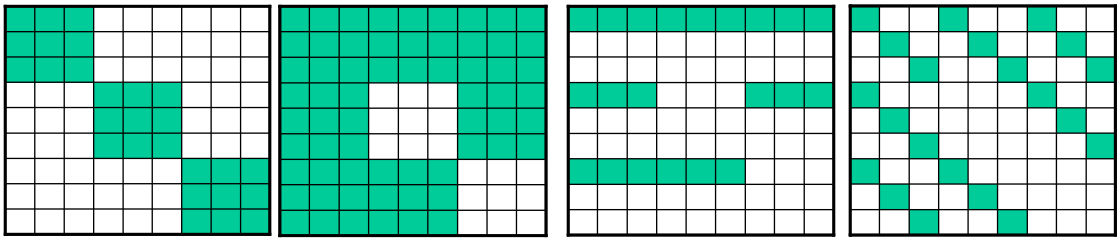


**Figure 3.** Let $N_R = N_I = 3$ and $N_C = 1$. Figure 3 shows the non-empty entries in the sub-matrices of **Q**. Figure 3a shows $\mathbf{S^I}$; figure 3b shows $\mathbf{T^{IP}}$, figure 3c shows $\mathbf{T^{IM}}$ and $\mathbf{T^{FM}}$, figure 3d shows $\mathbf{T^{FP}}$.

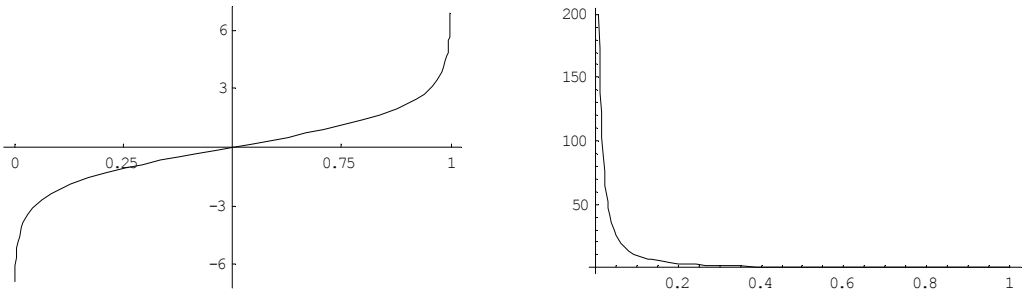



**Figure 4**. Figure 4a shows $h_1$ and Figure 4b shows $\mathrm{Log}_{10}(h_2)$. E.g., $h_2(0.1)$ is $10^{10}$, $h_2(0.01)$ is $10^{200}$.

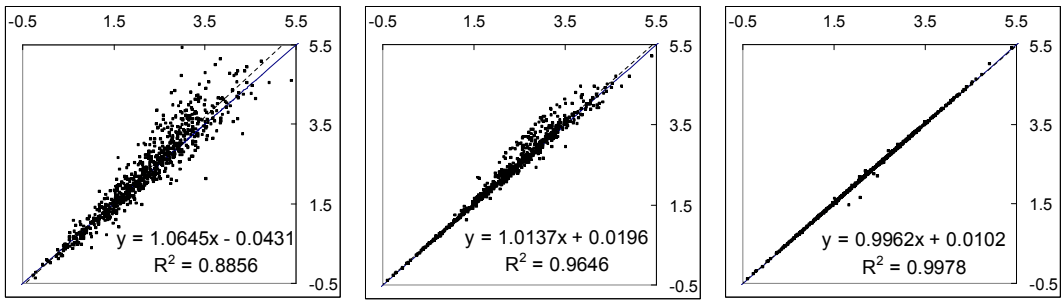



**Figure 5**. Regressions of logarithm of embodied GHG emissions estimated using different methods, for 720 data points: 5a) method 1 vs. method 4; 5b) method 2 vs. method 4; 5a) method 3 vs. method 4. Besides data points, the identity line is full, the regression line is dashed; a linear regression and correlation coefficient are also displayed.

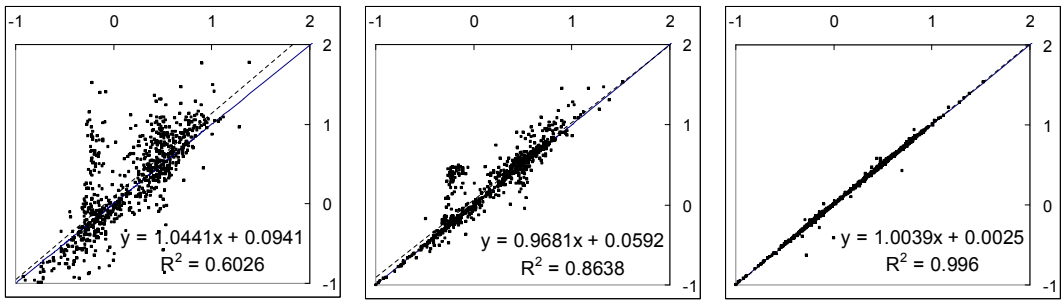



**Figure 6**. Regressions of logarithm of embodied GHG intensities estimated using different methods, for 720 data points: 5a) method 1 vs. method 4; 5b) method 2 vs. method 4; 5a) method 3 vs. method 4. Besides data points, the identity line is full, the regression line is dashed; a linear regression and correlation coefficient are also displayed.

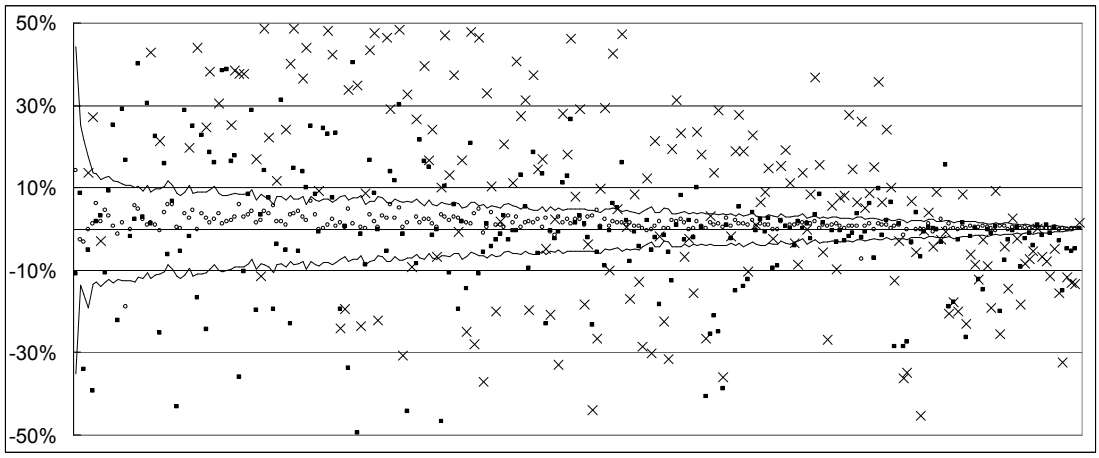


**Figure 7**. Relative error of different methods. The *x*-axis is an ordering of the 240 data points of upstream embodied GHG emissions of final expenditure from domestic production, according to decreasing width of the 90% confidence bound. Full lines are the upper and lower 90%-confidence bounds. Crosses are the relative errors of estimates from method 1. Full squares are the relative errors of estimated from method 2. Empty circles (close to 0 %) are the relative errors of estimates from method 3.

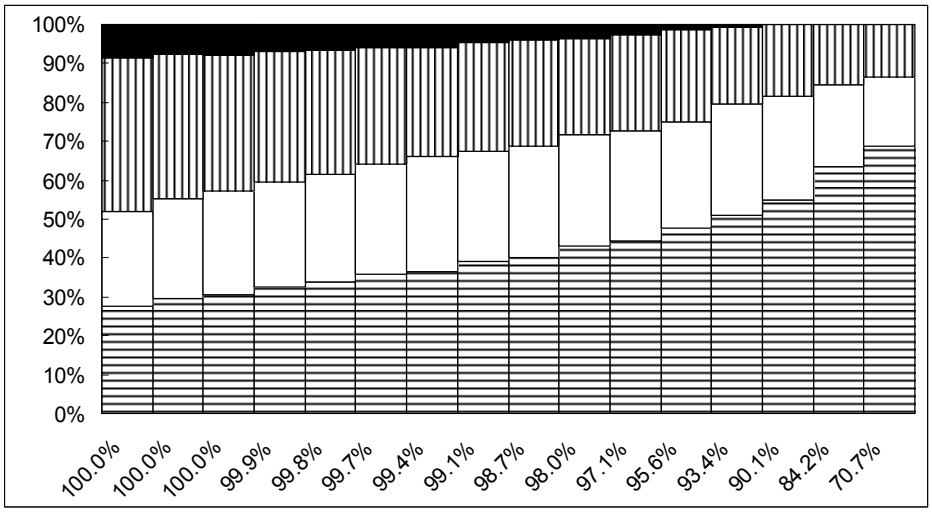



**Figure 8**. Relative width of the 90% confidence bounds of estimated embodied emissions. The *x*-axis is the percentage of total estimated embodied emissions above that point. The *y*-axis is the probability that the 90% confidence bound has a relative width smaller than a given threshold: horizontal stripes are the 10% threshold; blank columns are the 20% threshold and vertical stripes are the 50% threshold.

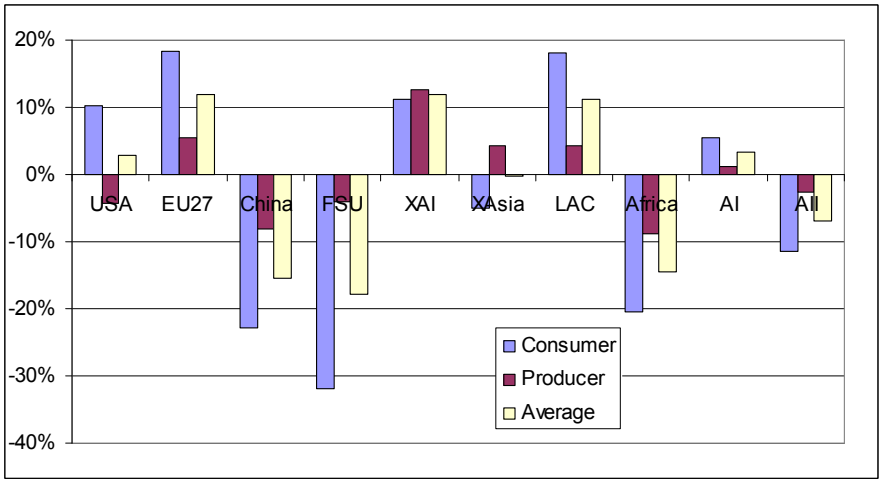



**Figure 9**. Embodied emissions relative to direct emissions in the 8-region and 2-region models.

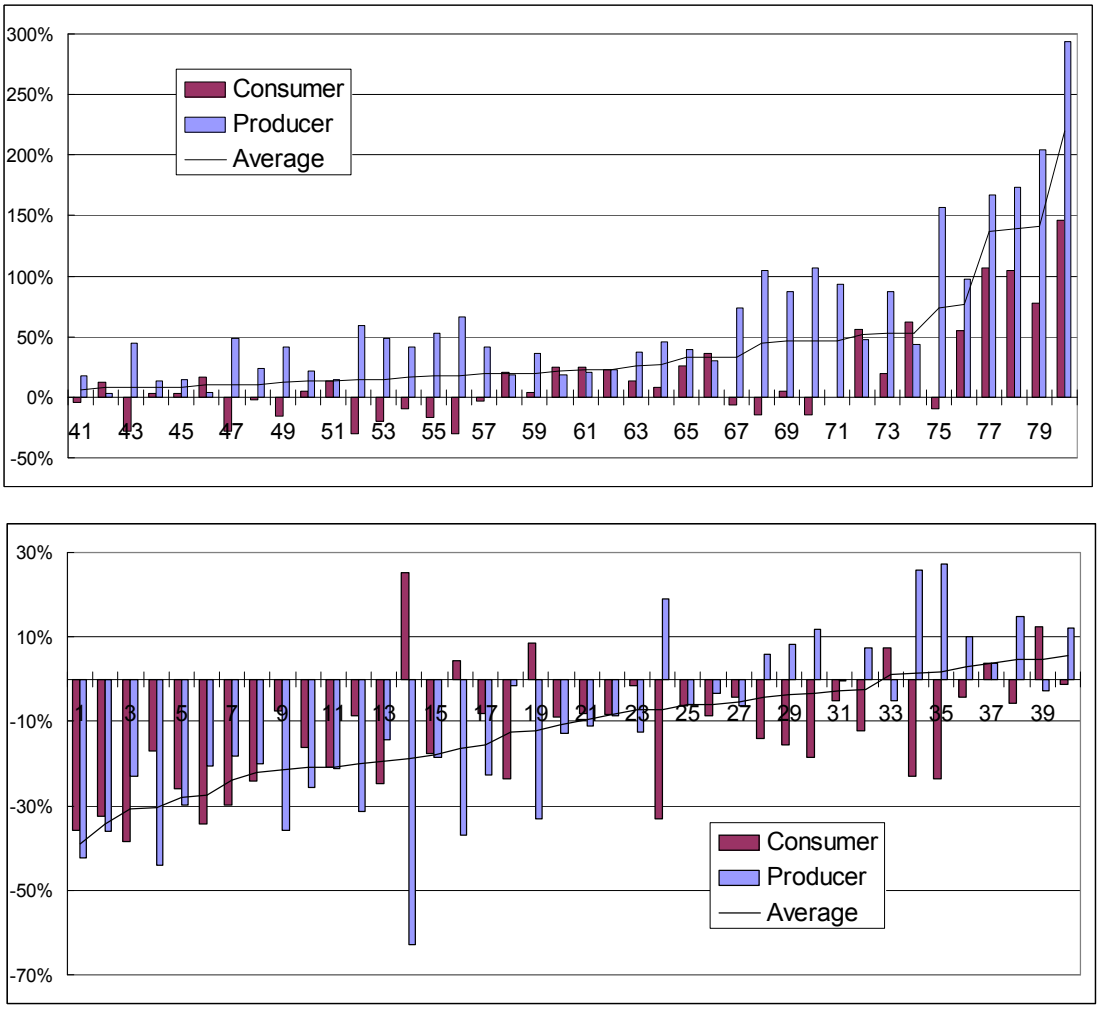



**Figure 10**. Embodied emissions relative to direct emissions in the 80-region model.

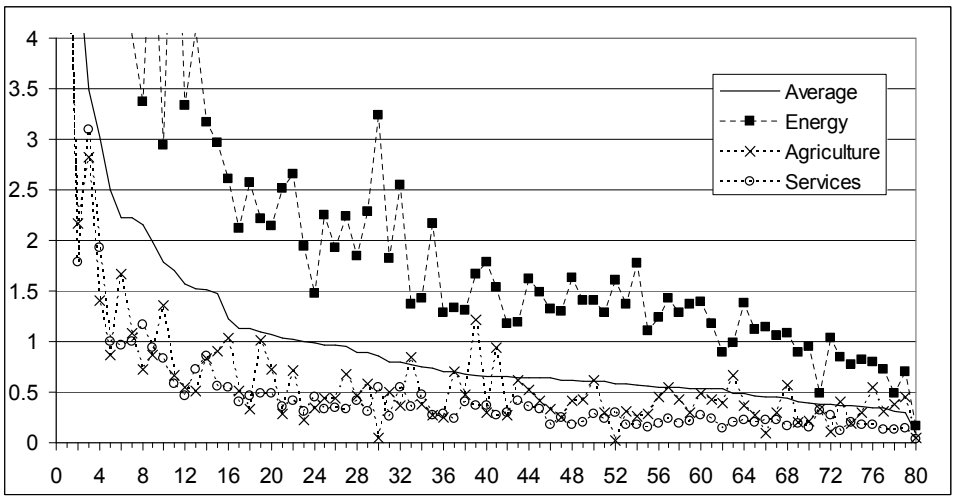



**Figure 11**. Upstream GHG intensity of domestic products (Kg $CO_2$-eq/USD2001).

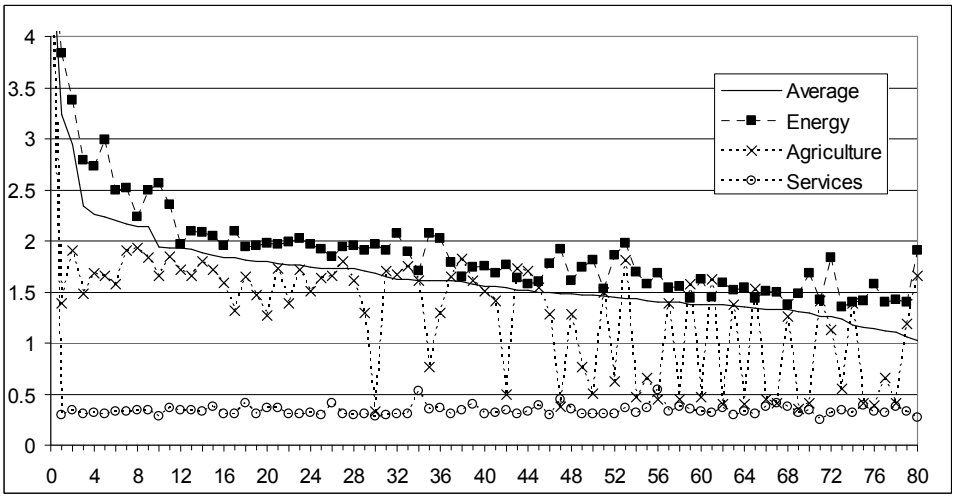



**Figure 12**. Upstream GHG intensity of imports to final demand (Kg $CO_2$-eq/USD2001).

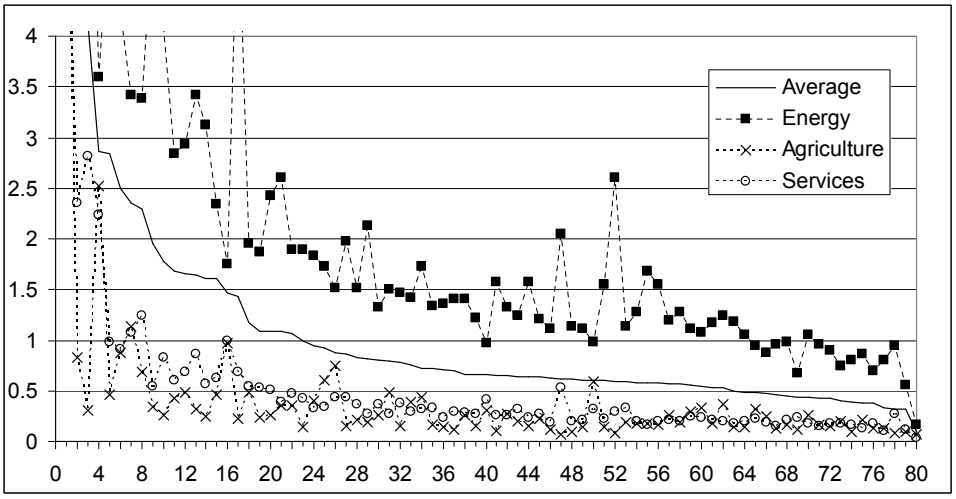



**Figure 13**. Downstream GHG intensity of domestic products (Kg $CO_2$-eq/USD2001).

## Tables

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | VDFM | VXMD | VTWR | | | | VDPM | VDGM | VDFA (CDGS) | VIFA (CDGS) |
| 2 | VDFA | | | | | | | | | | |
| 3 | | | | VXWD | | | | | | | |
| 4 | | | | | VIWS | | | | | | |
| 5 | | | | | | VIMS | | | | | |
| 6 | | | | | | | VIFM | VIPM | VIGM | | |
| 7 | VIFA | | | | | | | | | | |
| 0 | OSEP EVFA | ISEP1 | MFRV XTRV | | TFRV | | ISEP2 | | | | |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 29.47 | 6.90 | 0.23 | | | | 16.91 | 5.06 | 5.66 | 1.09 | 65.32 |
| 2 | 29.86 | | | | | | | | | | | 29.86 |
| 3 | | | | 6.91 | | | | | | | | 6.91 |
| 4 | | | | | 7.14 | | | | | | | 7.14 |
| 5 | | | | | | 7.38 | | | | | | 7.38 |
| 6 | | | | | | | 5.57 | 1.66 | 0.14 | | | 7.38 |
| 7 | 5.67 | | | | | | | | | | | 5.67 |
| 0 | 1.05 | 0.39 | 0.01 | | 0.23 | | 0.10 | | | | 30.52 | 129.7 |
| | 28.73 | | 0.01 | | | | | 30.52 | | | | |
| | 65.32 | 29.86 | 6.91 | 7.14 | 7.38 | 7.38 | 5.67 | 129.7 | | | | 99.14 |

**Table 1.** Table 1a reports the GTAP code and Table 1b represents the aggregated sum of that table (over all regions and industries). Table 1 shows the quantitative relationship between different tables in the GTAP database.

| Constraint | IP | IM | FP | FM | Flows (in $10^{12}$ USD 2001) | |
|---|---|---|---|---|---|---|
| 1 | | | X | X | FP + FM | 2.66941666 |
| 2 | X | X | | | IP + IM | 4.47525390 |
| 3 | | X | | X | IM + FM | 0.23444167 |
| 4 | | X | | X | IM + FM | 0.23444167 |
| 5 | X | | X | | IP + FP | 6.91022883 |
| 6 | X | | X | | IP + FP | 6.91022900 |
| 1 + 2 | | | | | Total | 7.14467056 |
| 3 + 5 | | | | | Total | 7.14467056 |

**Table 2.** The left side of Table 2 shows which constraint set affects which unknown set. The right hand side shows the aggregate value (sum over all regions and industries) of those sets. The two bottom lines show that the different datasets are consistent.

| Method | Information used |
|---|---|
| 1 – Homogeneous intensity | $\mathbf{t}^{\mathbf{r}}$ and $\mathbf{t}^{\mathbf{c}}$ (single region) |
| 2 – Trade shares | $\mathbf{t}^{\mathbf{r}}$, $\mathbf{t}^{\mathbf{c}}$, $\mathbf{k}^{\mathbf{2}}$ and $\mathbf{T}^{\mathbf{R}}$ (aggregation of $\mathbf{k}^{\mathbf{5}}$) |
| 3 – Maximum entropy principle | $\mathbf{t}^{\mathbf{r}}$, $\mathbf{t}^{\mathbf{c}}$, $\mathbf{k}$ |

**Table 3.** The information used by different estimation methods.

| Parameter | $N_P$ | $p_C$ | $p_X$ | $p_M$ | $p_R$ | $p_T$ | $p_S$ |
|---|---|---|---|---|---|---|---|
| Value | 100 | 0.9 | 0.5 | 0.005 | 0.2 | 0. 0001 | 0.1 |

**Table 4.** Parameters of the genetic optimization algorithm.

| Region | % of local | % of consumer | % of producer | % of average |
|---|---|---|---|---|
| Annex I | 67.5 | 71.2 | 68.3 | 69.8 |
| Annex II | 32.5 | 28.8 | 31.7 | 30.2 |

**Table 5.** Fraction of local and embodied GHG emissions in the 2-region model.

| Region | % of local | % of consumer | % of producer | % of average |
|---|---|---|---|---|
| USA | 26.4 | 29.1 | 25.3 | 27.2 |
| EU27 | 18.9 | 22.4 | 20.0 | 21.2 |
| China | 13.9 | 10.7 | 12.8 | 11.7 |
| XUSSR | 11.6 | 7.9 | 11.1 | 9.5 |
| XAnnex I | 10.6 | 11.8 | 11.9 | 11.8 |
| XAsia | 10.4 | 9.9 | 10.9 | 10.4 |
| LAC | 4.3 | 5.1 | 4.5 | 4.8 |
| Africa | 3.8 | 3.0 | 3.5 | 3.3 |

**Table 6.** Fraction of local and embodied emissions in the 8-region model.

**Table 7. (following page)** Rank and fraction of local and embodied emissions in the 80-region model.

| Region | Local | | Consumer | | Producer | | Average | | Region | Local | | Consumer | | Producer | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | % | Rank | % | Rank | % | Rank | % | | Rank | % | Rank | % | Rank | % | Rank | % |
| usa | 1 | 26.44 | 1 | 29.13 | 1 | 25.32 | 1 | 27.22 | bgr | 41 | 0.25 | 47 | 0.20 | 48 | 0.16 | 48 | 0.18 |
| chn | 2 | 13.86 | 2 | 10.70 | 2 | 12.75 | 2 | 11.73 | xef | 42 | 0.21 | 37 | 0.32 | 37 | 0.33 | 38 | 0.32 |
| rus | 3 | 7.40 | 4 | 4.96 | 3 | 8.03 | 4 | 6.50 | irl | 43 | 0.20 | 42 | 0.26 | 40 | 0.27 | 42 | 0.26 |
| jpn | 4 | 5.37 | 3 | 6.59 | 4 | 6.61 | 3 | 6.60 | svk | 44 | 0.19 | 53 | 0.16 | 50 | 0.14 | 51 | 0.15 |
| xsu | 5 | 4.18 | 8 | 2.93 | 7 | 3.10 | 8 | 3.01 | che | 45 | 0.19 | 29 | 0.51 | 33 | 0.39 | 30 | 0.45 |
| deu | 6 | 3.86 | 5 | 4.57 | 5 | 4.66 | 5 | 4.62 | xfa | 46 | 0.19 | 40 | 0.28 | 43 | 0.21 | 43 | 0.24 |
| ind | 7 | 3.57 | 7 | 3.11 | 6 | 3.24 | 6 | 3.18 | chl | 47 | 0.17 | 49 | 0.19 | 47 | 0.17 | 47 | 0.18 |
| can | 8 | 2.75 | 11 | 2.41 | 9 | 2.70 | 9 | 2.56 | nzl | 48 | 0.15 | 54 | 0.14 | 49 | 0.14 | 52 | 0.14 |
| gbr | 9 | 2.50 | 6 | 3.44 | 8 | 2.84 | 7 | 3.14 | per | 49 | 0.14 | 51 | 0.17 | 51 | 0.13 | 50 | 0.15 |
| ita | 10 | 2.18 | 9 | 2.66 | 12 | 2.29 | 11 | 2.48 | mar | 50 | 0.13 | 57 | 0.12 | 53 | 0.12 | 55 | 0.12 |
| fra | 11 | 1.87 | 10 | 2.62 | 10 | 2.35 | 10 | 2.49 | xap | 51 | 0.13 | 44 | 0.25 | 46 | 0.19 | 46 | 0.22 |
| xme | 12 | 1.84 | 12 | 2.22 | 11 | 2.30 | 12 | 2.26 | vnm | 52 | 0.12 | 50 | 0.18 | 55 | 0.09 | 54 | 0.14 |
| aus | 13 | 1.71 | 15 | 1.52 | 14 | 1.57 | 15 | 1.55 | xcb | 53 | 0.12 | 48 | 0.20 | 56 | 0.08 | 53 | 0.14 |
| esp | 14 | 1.46 | 13 | 1.72 | 15 | 1.40 | 14 | 1.56 | xca | 54 | 0.12 | 46 | 0.22 | 52 | 0.12 | 49 | 0.17 |
| pol | 15 | 1.40 | 18 | 1.11 | 19 | 1.11 | 18 | 1.11 | sgp | 55 | 0.11 | 36 | 0.34 | 44 | 0.20 | 40 | 0.27 |
| mex | 16 | 1.40 | 14 | 1.60 | 13 | 1.59 | 13 | 1.59 | hrv | 56 | 0.11 | 58 | 0.11 | 54 | 0.09 | 56 | 0.10 |
| zaf | 17 | 1.38 | 21 | 0.89 | 17 | 1.28 | 19 | 1.08 | zwe | 57 | 0.09 | 62 | 0.07 | 57 | 0.07 | 59 | 0.07 |
| xss | 18 | 1.25 | 20 | 0.93 | 20 | 1.05 | 20 | 0.99 | est | 58 | 0.08 | 66 | 0.05 | 62 | 0.05 | 65 | 0.05 |
| bra | 19 | 1.13 | 16 | 1.27 | 18 | 1.12 | 17 | 1.20 | tun | 59 | 0.08 | 60 | 0.08 | 59 | 0.06 | 60 | 0.07 |
| kor | 20 | 1.12 | 17 | 1.17 | 16 | 1.30 | 16 | 1.23 | zmb | 60 | 0.07 | 71 | 0.04 | 61 | 0.06 | 63 | 0.05 |
| idn | 21 | 0.97 | 27 | 0.61 | 21 | 1.01 | 23 | 0.81 | svn | 61 | 0.07 | 61 | 0.08 | 60 | 0.06 | 61 | 0.07 |
| xea | 22 | 0.84 | 22 | 0.79 | 22 | 0.90 | 22 | 0.85 | bdg | 62 | 0.07 | 56 | 0.13 | 58 | 0.07 | 57 | 0.10 |
| ndl | 23 | 0.80 | 19 | 1.08 | 23 | 0.83 | 21 | 0.96 | ltu | 63 | 0.06 | 59 | 0.09 | 64 | 0.05 | 62 | 0.07 |
| xnf | 24 | 0.71 | 23 | 0.74 | 24 | 0.74 | 24 | 0.74 | xer | 64 | 0.05 | 55 | 0.13 | 63 | 0.05 | 58 | 0.09 |
| arg | 25 | 0.62 | 26 | 0.62 | 25 | 0.59 | 25 | 0.61 | lux | 65 | 0.05 | 63 | 0.06 | 65 | 0.04 | 64 | 0.05 |
| tha | 26 | 0.61 | 31 | 0.42 | 26 | 0.56 | 28 | 0.49 | lva | 66 | 0.03 | 65 | 0.05 | 70 | 0.02 | 69 | 0.04 |
| cze | 27 | 0.57 | 34 | 0.36 | 34 | 0.38 | 34 | 0.37 | tza | 67 | 0.03 | 72 | 0.04 | 71 | 0.02 | 72 | 0.03 |
| bel | 28 | 0.56 | 24 | 0.65 | 28 | 0.53 | 26 | 0.59 | lka | 68 | 0.03 | 70 | 0.04 | 68 | 0.03 | 70 | 0.04 |
| rom | 29 | 0.53 | 30 | 0.42 | 32 | 0.40 | 31 | 0.41 | xoc | 69 | 0.03 | 64 | 0.06 | 73 | 0.02 | 67 | 0.04 |
| grc | 30 | 0.51 | 28 | 0.57 | 29 | 0.41 | 29 | 0.49 | ury | 70 | 0.03 | 69 | 0.04 | 72 | 0.02 | 71 | 0.03 |
| mys | 31 | 0.43 | 52 | 0.16 | 27 | 0.54 | 35 | 0.35 | xsm | 71 | 0.02 | 73 | 0.03 | 69 | 0.03 | 73 | 0.03 |
| xsa | 32 | 0.41 | 33 | 0.38 | 35 | 0.38 | 33 | 0.38 | uga | 72 | 0.02 | 74 | 0.03 | 75 | 0.01 | 74 | 0.02 |
| aut | 33 | 0.34 | 25 | 0.64 | 31 | 0.41 | 27 | 0.52 | xsd | 73 | 0.02 | 68 | 0.05 | 66 | 0.03 | 66 | 0.04 |
| fin | 34 | 0.32 | 39 | 0.31 | 36 | 0.36 | 36 | 0.34 | bwa | 74 | 0.02 | 78 | 0.02 | 74 | 0.02 | 76 | 0.02 |
| hun | 35 | 0.29 | 45 | 0.25 | 42 | 0.22 | 44 | 0.23 | xsc | 75 | 0.02 | 75 | 0.02 | 76 | 0.01 | 75 | 0.02 |
| col | 36 | 0.28 | 41 | 0.27 | 39 | 0.27 | 41 | 0.27 | mwi | 76 | 0.02 | 79 | 0.01 | 77 | 0.01 | 79 | 0.01 |
| prt | 37 | 0.28 | 32 | 0.39 | 41 | 0.25 | 37 | 0.32 | alb | 77 | 0.01 | 76 | 0.02 | 78 | 0.01 | 77 | 0.02 |
| dnk | 38 | 0.26 | 38 | 0.31 | 38 | 0.33 | 39 | 0.32 | xse | 78 | 0.01 | 67 | 0.05 | 67 | 0.03 | 68 | 0.04 |
| phl | 39 | 0.26 | 43 | 0.25 | 45 | 0.19 | 45 | 0.22 | mlt | 79 | 0.01 | 77 | 0.02 | 79 | 0.01 | 78 | 0.02 |
| swe | 40 | 0.25 | 35 | 0.36 | 30 | 0.41 | 32 | 0.39 | mdg | 80 | 0.01 | 80 | 0.01 | 80 | 0.01 | 80 | 0.01 |

| Region | Rank | Region | Rank | Region | Rank | Region | Rank | Region | Rank |
|---|---|---|---|---|---|---|---|---|---|
| est | 1 | chn | 17 | xea | 33 | lux | 49 | fra | 65 |
| cze | 2 | phl | 18 | tza | 34 | ita | 50 | irl | 66 |
| mwi | 3 | rus | 19 | mdg | 35 | mex | 51 | ury | 67 |
| zmb | 4 | ind | 20 | usa | 36 | alb | 52 | xoc | 68 |
| xsu | 5 | aus | 21 | xnf | 37 | ltu | 53 | bdg | 69 |
| bgr | 6 | xsa | 22 | bel | 38 | prt | 54 | mlt | 70 |
| svk | 7 | can | 23 | fin | 39 | xsc | 55 | xca | 71 |
| rom | 8 | uga | 24 | bra | 40 | xcb | 56 | xef | 72 |
| zaf | 9 | nzl | 25 | esp | 41 | lka | 57 | aut | 73 |
| xss | 10 | mar | 26 | bwa | 42 | deu | 58 | swe | 74 |
| pol | 11 | col | 27 | lva | 43 | ndl | 59 | xer | 75 |
| tha | 12 | hrv | 28 | xsm | 44 | dnk | 60 | xap | 76 |
| hun | 13 | tun | 29 | chl | 45 | xme | 61 | che | 77 |
| mys | 14 | grc | 30 | kor | 46 | jpn | 62 | xsd | 78 |
| zwe | 15 | arg | 31 | vnm | 47 | gbr | 63 | sgp | 79 |
| idn | 16 | svn | 32 | per | 48 | xfa | 64 | xse | 80 |

**Table 8.** Ordering of regions displayed in Figure 10.

| Region | Local Intensity | Rank | | | | Region | Local Intensity | Rank | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Local | Up Int | Up fin | Down | | | Local | Up Int | Up fin | Down |
| rus | 2.654 | 1 | 2 | 46 | 2 | nzl | 0.338 | 41 | 52 | 13 | 34 |
| zmb | 2.098 | 2 | 4 | 8 | 3 | usa | 0.334 | 42 | 48 | 58 | 51 |
| zaf | 1.358 | 3 | 6 | 21 | 5 | fin | 0.328 | 43 | 45 | 24 | 31 |
| rom | 1.322 | 4 | 5 | 6 | 6 | xsm | 0.316 | 44 | 66 | 34 | 42 |
| zwe | 1.252 | 5 | 9 | 10 | 9 | esp | 0.311 | 45 | 49 | 66 | 54 |
| est | 1.236 | 6 | 7 | 9 | 7 | xme | 0.308 | 46 | 56 | 70 | 36 |
| xsu | 1.213 | 7 | 1 | 1 | 1 | chl | 0.308 | 47 | 47 | 15 | 39 |
| mwi | 1.186 | 8 | 11 | 7 | 17 | mex | 0.303 | 48 | 55 | 64 | 48 |
| xss | 0.987 | 9 | 13 | 76 | 10 | bra | 0.300 | 49 | 53 | 69 | 58 |
| ind | 0.945 | 10 | 12 | 35 | 13 | xsc | 0.296 | 50 | 43 | 4 | 45 |
| chn | 0.907 | 11 | 8 | 56 | 8 | xea | 0.288 | 51 | 59 | 57 | 37 |
| pol | 0.866 | 12 | 15 | 17 | 14 | lka | 0.280 | 52 | 69 | 18 | 67 |
| cze | 0.805 | 13 | 14 | 33 | 11 | kor | 0.267 | 53 | 51 | 49 | 35 |
| idn | 0.782 | 14 | 18 | 36 | 12 | mlt | 0.266 | 54 | 39 | 43 | 46 |
| svk | 0.727 | 15 | 10 | 27 | 15 | lux | 0.261 | 55 | 57 | 73 | 63 |
| xsa | 0.659 | 16 | 21 | 51 | 20 | bel | 0.257 | 56 | 64 | 68 | 57 |
| aus | 0.586 | 17 | 22 | 67 | 21 | ita | 0.254 | 57 | 58 | 42 | 61 |
| uga | 0.555 | 18 | 30 | 11 | 52 | prt | 0.253 | 58 | 37 | 59 | 53 |
| hun | 0.552 | 19 | 20 | 12 | 22 | deu | 0.244 | 59 | 61 | 54 | 43 |
| tha | 0.550 | 20 | 23 | 75 | 18 | ndl | 0.224 | 60 | 60 | 60 | 62 |
| hrv | 0.534 | 21 | 17 | 31 | 19 | irl | 0.212 | 61 | 42 | 80 | 40 |
| can | 0.489 | 22 | 31 | 62 | 24 | dnk | 0.212 | 62 | 67 | 77 | 60 |
| mys | 0.486 | 23 | 24 | 78 | 16 | xfa | 0.210 | 63 | 73 | 20 | 71 |
| col | 0.479 | 24 | 35 | 65 | 29 | aut | 0.208 | 64 | 50 | 23 | 59 |
| grc | 0.471 | 25 | 25 | 39 | 27 | xoc | 0.204 | 65 | 44 | 41 | 70 |
| xnf | 0.454 | 26 | 28 | 45 | 23 | gbr | 0.204 | 66 | 68 | 71 | 64 |
| ltu | 0.439 | 27 | 16 | 2 | 25 | mdg | 0.187 | 67 | 72 | 53 | 78 |
| lva | 0.412 | 28 | 19 | 14 | 33 | ury | 0.182 | 68 | 65 | 38 | 77 |
| tun | 0.407 | 29 | 46 | 19 | 32 | xca | 0.182 | 69 | 74 | 30 | 72 |
| alb | 0.403 | 30 | 27 | 37 | 44 | fra | 0.178 | 70 | 70 | 61 | 68 |
| bwa | 0.393 | 31 | 41 | 3 | 26 | xer | 0.176 | 71 | 63 | 29 | 75 |
| tza | 0.387 | 32 | 32 | 25 | 47 | jpn | 0.169 | 72 | 77 | 50 | 74 |
| xcb | 0.368 | 33 | 29 | 16 | 56 | bdg | 0.167 | 73 | 62 | 26 | 76 |
| mar | 0.367 | 34 | 34 | 52 | 28 | xef | 0.161 | 74 | 75 | 28 | 66 |
| phl | 0.365 | 35 | 40 | 47 | 41 | swe | 0.139 | 75 | 76 | 74 | 65 |
| vnm | 0.358 | 36 | 26 | 72 | 38 | xap | 0.108 | 76 | 78 | 63 | 79 |
| arg | 0.357 | 37 | 54 | 44 | 55 | xsd | 0.103 | 77 | 71 | 79 | 69 |
| bgr | 0.344 | 38 | 3 | 5 | 4 | sgp | 0.102 | 78 | 38 | 48 | 50 |
| per | 0.341 | 39 | 36 | 32 | 49 | che | 0.094 | 79 | 79 | 55 | 73 |
| svn | 0.340 | 40 | 33 | 22 | 30 | xse | 0.020 | 80 | 80 | 40 | 80 |

**Table 9.** Value of local GHG intensity and the ranking of average intensities: Up int = Upstream domestic products; Up fin = Upstream imports for final demand; Down = Downstream. There rankings are the orderings of Figures 11 to 13.

## Tables of the Appendix

| Sectors | GTAP 6.0 Database |
|---|---|
| Energy (including manufacture and transport) | Coal; Oil; Gas; Electricity; Gas manufacture and distribution; Other minerals; Textiles; Wearing apparel; Leather products; Wood products; Paper products, publishing; Petroleum, coal products; Chemical, rubber, plastic products; Sea transport; Air transport; Other transport; Construction; Mineral products; Ferrous metals; Other metals; Metal products; Motor vehicles and parts; Transport equipment; Electronic equipment; Machinery and equipment; Manufactures |
| Agriculture | Paddy rice; Wheat; Cereal grains; Vegetables, fruit, nuts; Oil seeds; Sugar cane, sugar beet; Plant-based fibers; Crops; Cattle, sheep, goats, horses; Animal products; Raw milk; Wool, silk-worm cocoons; Forestry; Fishing; Meat: cattle, sheep, goats, horse; Meat products; Vegetable oils and fats; Dairy products; Processed rice; Sugar; Food products; Beverages and tobacco products |
| Services | Water; Trade; Communication; Financial services; Insurance; Business services; Recreation services; Health and Education; Dwellings |

**Table A1 – Equivalence between sectors in the present work and the GTAP database.**

| Code | Region | Reference year | Code | Region | Reference year |
|---|---|---|---|---|---|
| aus | Australia | 2003 | irl | Ireland | 2003 |
| nzl | New Zealand | 2003 | ita | Italy | 2003 |
| xoc | Rest of Oceania | 1994 | lux | Luxembourg | 2003 |
| chn | China+Hong Kong | 1994 | nld | Netherlands | 2003 |
| jpn | Japan | 2003 | prt | Portugal | 2003 |
| kor | Korea | 1990 | esp | Spain | 2003 |
| xea | Rest of East Asia | several | swe | Sweden | 2003 |
| idn | Indonesia | 1994 | che | Switzerland | 2003 |
| mys | Malaysia | 1994 | xef | Rest of EFTA | 2003 |
| phl | Philippines | 1994 | xer | Rest of Europe | several |
| sgp | Singapore | 1994 | alb | Albania | 1994 |
| tha | Thailand | 1994 | bgr | Bulgaria | 2003 |
| vnm | Vietnam | 1994 | hrv | Croatia | 2003 |

| | | | | | |
|---|---|---|---|---|---|
| xse | Rest of Southeast Asia | several | cze | Czech Republic | 2003 |
| bgd | Bangladesh | 1994 | hun | Hungary | 2003 |
| ind | India | 1994 | mlt | Malta | 2000 |
| lka | Sri Lanka | 1995 | pol | Poland | 2002 |
| xsa | Rest of South Asia | 1994 | rom | Romania | 2003 |
| can | Canada | 2003 | svk | Slovakia | 2003 |
| usa | United States | 2003 | svn | Slovenia | 2003 |
| mex | Mexico | 1990 | est | Estonia | 2003 |
| col | Colombia | 1994 | lva | Latvia | 2003 |
| per | Peru | 1994 | ltu | Lithuania | 2003 |
| xap | Rest of Andean Pact | several | rus | Russian Federation | 1999 |
| arg | Argentina | 1997 | xsu | Rest of Former Soviet Union | several |
| bra | Brazil | 1994 | xme | Rest of Middle East | several |
| chl | Chile | 1994 | mar | Morocco | 1994 |
| ury | Uruguay | 1998 | tun | Tunisia | 1994 |
| xsm | Rest of South America | several | xnf | Rest of North Africa | several |
| xca | Central America | several | bwa | Botswana | 1994 |
| xfa | Rest of FTAA | several | zaf | South Africa | 1994 |
| xcb | Rest of Caribbean | 1996 | xsc | Rest of South African CU | 1994 |
| aut | Austria | 2003 | mwi | Malawi | 1994 |
| bel | Belgium | 2003 | tza | Tanzania | 1994 |
| dnk | Denmark | 2003 | zmb | Zambia | 1994 |
| fin | Finland | 2003 | zwe | Zimbabwe | 1994 |
| fra | France | 2003 | xsd | Rest of SADC | 1995 |
| deu | Germany | 2003 | mdg | Madagascar | 1994 |
| gbr | United Kingdom | 2003 | uga | Uganda | 1994 |
| grc | Greece | 2003 | xss | Rest of Sub-Saharan Africa | several |

**Table A2 – Legend of codes and reference year for emissions.**

Composite region Member nations

| | |
|---|---|
| Rest of Oceania | American Samoa; Cook Islands; Fiji; French Polynesia; Guam; Kiribati; Marshall Islands; Micronesia, Federated St; Nauru; New Caledonia; Norfolk Island; Northern Mariana Islands; Niue; Palau; Papua New Guinea; Samoa; Solomon Islands; Tokelau; Tonga; Tuvalu; Vanuatu; Wallis and Futuna |
| Rest of East Asia | Macau; Mongolia; Korea, Democratic People's; Taiwan |
| Rest of Southeast Asia | Brunei Darussalam; Cambodia; Lao People's Democratic Republic; Myanmar; Timor Leste |
| Rest of South Asia | Afghanistan; Bhutan; Maldives; Nepal; Pakistan |
| Rest of North America | Bermuda; Greenland; Saint Pierre and Miquelon |
| Rest of Andean Pact | Bolivia; Ecuador |
| Rest of South America | Falkland Islands (Malvinas); French Guiana; Guyana; Paraguay; Suriname; Venezuela |
| Central America | Belize; Costa Rica; El Salvador; Guatemala; Honduras; Nicaragua; Panama |
| Rest of FTAA | Antigua & Barbuda; Bahamas; Barbados; Dominica; Dominican Republic; Grenada; Haiti; Jamaica; Puerto Rico; Saint Kitts and Nevis; Saint Lucia; Saint Vincent and the Grenadines; Trinidad and Tobago; Virgin Islands, U.S. |
| Rest of the Caribbean and North America | Anguilla; Aruba; Cayman Islands; Cuba; Guadeloupe; Martinique; Montserrat; Netherlands Antilles; Turks and Caicos; Virgin Islands, British; Bermuda; Greenland; Saint Pierre and Miquelon |
| Rest of EFTA | Iceland; Liechtenstein; Norway |
| Rest of Europe | Andorra; Bosnia and Herzegovina; Faroe Islands; Gibraltar; Macedonia, the former Yugoslav republic of; Monaco; San Marino; Serbia and Montenegro; Cyprus |
| Rest of Former Soviet Union | Armenia; Azerbaijan; Belarus; Georgia; Kazakhstan; Kyrgyzstan; Moldova, Republic of; Tajikistan; Turkmenistan; Ukraine; Uzbekistan |
| Rest of Middle East | Bahrain; Iran, Islamic Republic of; Iraq; Israel; Jordan; Kuwait; Lebanon; Palestinian Territory, Occupied; Oman; Qatar; Saudi Arabia; Syrian Arab Republic; United Arab Emirates; Yemen; Turkey |
| Rest of North Africa | Algeria; Egypt; Libyan Arab Jamahiriya |

| | |
|---|---|
| Rest of South African CU | Lesotho; Namibia; Swaziland |
| Rest of SADC | Angola; Congo, the Democratic Republic of; Mauritius; Seychelles |
| Rest of Sub-Saharan Africa | Benin; Burkina Faso; Burundi; Cameroon; Cape Verde; Central African Republic; Chad; Comoros; Congo; Cote d'Ivoire; Djibouti; Equatorial Guinea; Eritrea; Ethiopia; Gabon; Gambia; Ghana; Guinea; Guinea-Bissau; Kenya; Liberia; Mali; Mauritania; Mayotte; Niger; Nigeria; Reunion; Rwanda; Saint Helena; Sao Tome and Principe; Senegal; Sierra Leone; Somalia; Sudan; Togo; Mozambique |

**Table A3 – Member nations of composite regions.**

| Aggregate region (2 region model) | Aggregate region (8 region model) | Disaggregate region (81 region model) |
|---|---|---|
| Annex I | United States of America (USA) | usa |
| | European Union (EU27) | aut; bgr; bel; dnk; fin; fra; deu; gbr; grc; irl; ita; lux; nld; prt; esp; swe; cze; hun; mlt; pol; rom; svk; svn; est; lva; ltu |
| | Rest of Annex I (XAI) | jpn; can; aus; nzl; che; xef; xer; alb; hrv |
| | Former Soviet Union (FSU) | rus; xsu |
| Annex II | China | chn |
| | Rest of Asia and Oceania (XAO) | xoc; kor; xea; idn; mys; phl; sgp; tha; vnm; xse; bgd; ind; lka; xsa; xme |
| | Latin America and the Caribbean (LAC) | mex; col; per; xap; arg; bra; chl; ury; xsm; xca; xfa; xcb |
| | Africa | mar; tun; xnf; bwa; zaf; xsc; mwi; tza; zmb; zwe; xsd; mdg; uga; xss |

**Table A3 – Regions composing aggregate models.**