



**HAL**  
open science

# Material and Energy Flows of the French Economy - Case study of NegaWatt and Ademe's scenarios

Alexandre Borthomieu

► **To cite this version:**

Alexandre Borthomieu. Material and Energy Flows of the French Economy - Case study of NegaWatt and Ademe's scenarios. Modeling and Simulation. 2020. hal-03016346

**HAL Id: hal-03016346**

**<https://inria.hal.science/hal-03016346>**

Submitted on 20 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Master Operations Research and Combinatorial Optimization**  
Master Informatique / Master Mathématiques & Applications

# **Material and Energy Flows of the French Economy**

Case study of NegaWatt and Ademe's scenarios

**Alexandre Borthomieu**

June 25, 2020

Research project performed at GSCOP and INRIA

Under the supervision of:

Vincent Jost, GSCOP

Jean-Yves Courtonne, INRIA

Defended before a jury composed of:

Prof. Van-Dat Cung

Prof. Matej Stehlik

Prof. Nadia Brauner

Prof. Marie-Laure Espinouse



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Context . . . . .	1
1.2	STA: a new line of research of the STEEP team . . . . .	3
1.3	Computational issues and operations research . . . . .	6
<b>2</b>	<b>State of the Art and Methods</b>	<b>9</b>
2.1	Input Output Analysis (IOA) . . . . .	9
2.2	Material and Energy Flow Analysis . . . . .	11
2.3	Supply Use Tables . . . . .	13
<b>3</b>	<b>Implementation, experiment, result</b>	<b>15</b>
3.1	Implementation of the SU formalism, IOA and AMC on the Negawatt scenario. . . . .	15
3.2	Coupling the Negawatt (energy) and Ademe (materials) reports . . . . .	17
3.3	Organization of the Excel file . . . . .	18
3.4	Sectors' Merge . . . . .	19
3.5	Units Problem and Layers . . . . .	21
<b>4</b>	<b>Conclusion</b>	<b>25</b>
	<b>Bibliography</b>	<b>27</b>
	<b>Appendix</b>	<b>29</b>



# Introduction

## 1.1 Context

In a situation of global pandemic, our way of life is particularly affected. While half the planet is locked down, we are witnessing a huge decrease of industrial and commercial activities, and wild animals are regaining cities cleared out of their inhabitants. Those phenomena appeared in only two months of lock down. What environmental effects can we expect with a change of our way of life for the next thirty years when two months produce already some visible results? Indeed, considering the environmental impacts of our modes of production and consumption, a shift is increasingly urgent. Those questionings can not stay unanswered. But the answers' spectrum is vast from individual actions to global awareness and political responses.

A recent report from *Carbone 4* [Dugast and Soyeux, 2019], an independent consulting firm specialized in low carbon strategies and environmental changes adaptation, shows that individual actions are necessary but largely insufficient to achieve the objectives of reduction of greenhouse gas emission. Thus, what is the best choice to make? And does such a choice exist? What alternatives exist and how do they compare in terms of compromises between socio-economic and environmental criteria?

For many decades, our societies have grown exponentially and their complexity with it. So much that it becomes a challenge to have an accurate understanding of the macro-economic inter-dependencies within an economy. However, we have the knowledge to develop models that represent our complex economy and the technology to run them.

In 1972, a research team from MIT published a report *The Limit to Growth* commissioned by the *Club of Rome*. This report describes the model developed by the research team: *World3*, which simulates the interactions between the earth (environment) and human systems. The model was based on five global variables “population, food production, industrialization, pollution and consumption of nonrenewable natural resources” [Meadows et al., 1972]. The different scenarios explored by the model lead to a description of the mechanism of collapse: i) exponential economic growth, ii) overshoot of the carrying capacity of the environment, iii) erosion of the carrying capacity during the overshoot, iv) inertia of the economic system which equals slowness of the political response. They concluded that if we change the growth trends there is a possibility where the variables described above would stabilize without overshoot or collapse. This was 1972.

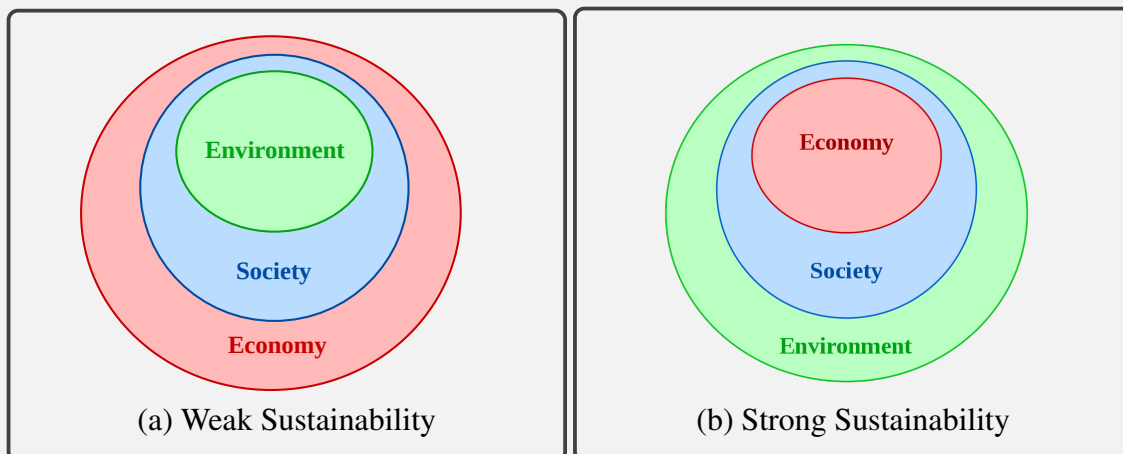
Since then, scientists and economists have developed hundred of models to describe the structure of the world economy, of a country's economy or of a specific sector. Yet, a lot of

those models are designed for economic purposes only and few of them really take into account all the interactions between their components, and pollution nor the use of natural resources do not appear in dependencies. Those missing parts contribute to growth trends. However, the necessity to have models that can help us understand our impact on the environment is growing.

Most economic models work with monetary flows or monetary units. Using monetary units to assess physical flows is a challenging task: the value of the resources can be constantly changing. Assessing real impacts on our environment from those monetary models is then extremely difficult. Thus, we need models that will take into account physical flows and energy flows: an (objective part of the) reality behind monetary flows.

The purpose of this work is thus to explore different models and formalisms in order to precisely depict our economy and the environmental pressures associated with it, through a multisectorial model. Since this internship is designed to be followed by a thesis, we will present here just one formalism implemented and applied to an example. It is important to present the « big picture », i.e., the context and objectives of the whole research axis, in order to understand how the internship (and the potential thesis following it) are contributing to it.

### Strong sustainability vs weak sustainability [Boisvert and Vivien, 2006]



Weak sustainability (figure (a)) is a concept which explains that environmental issues could be fixed with more economic growth. To be precise, it relies on an economic definition of sustainability: sustainability means the transmission to the future generation of a capacity to produce well-being at least equal to the present generation's one. Therefore, weak sustainability wants to substitute natural capital with technological capital to create economic wealth. That is why, we prefer monetary units with this prism.

On the other hand, the concept of strong sustainability (figure (b)) puts priority on ecosystems. They must persist, not only for moral reason. According to Herman Daly, sustainability relies on three principles.

- The consumption rate of renewable resources must not exceed renewing rate of those same resources.
- The consumption pace of non renewable resources must not exceed the pace of

finding and developing renewable substitutes.

- The emission rate of pollutants must not exceed the capacity of the environment to absorb and assimilate those pollutants.

Physical units are here required to compute the extracted quantities from the environment.

## 1.2 STA: a new line of research of the STEEP team

Considering the need for an understanding of a given economy coupled to an environmental approach, the STEEP team developed this line of research called Socio-Technical Alternatives (STA). Its goal is to study an economy by a physical approach, to analyse the impact on the environment but also to evaluate and propose alternative economic structures by its modes of production and consumption. Two axis can arise from that: analysis of specific sectors and systemic analysis of the whole economy while taking into account interactions between sectors.

The first axis focuses on specific supply chains and allows a detailed analysis of upstream and downstream vulnerabilities, dependencies to natural and fossil resources as well as computation of environmental footprints. It must be stressed here that footprints can be computed from a production or from a consumption perspective and the supply chain approach makes it possible to connect the two. Indeed, only analyzing the impacts of our production could mislead us about the real impact of our system. Here is a figure that shows the differences that can occur if we propagate the environmental impacts of our whole economy in France (figure 1.1, p.4). On the right part, the "footprint of the national territory" is slowly decreasing over the years. The left part is the greenhouse gas emissions, it contains also emissions owed to imported side of our economy and it was growing until 2010 then stabilizing (maybe?). One can see that production and consumption footprints convey different messages and should be both considered during the decision-making process.

The importance of taking into account all interactions in our system in a precise way is clearly visible.

Two softwares were designed by the STEEP team for supply chain studies: one for supply chain material flow analysis (SC-MFA) and one for datavisualization through Sankey diagrams<sup>1</sup>. Every Sankey diagram in this report was made by this tool. SC-MFA is based on data reconciliation and downscaling techniques, allowing to meet mass conservation but also geographical aggregation constraints. [Courtonne, 2016] explains that sometimes data can be missing or inaccurate to describe the whole sector. The collect of data to fill our model can be constraining and sometimes unsuccessful. Thus, a tool that can highlight missing part and propagate uncertainty was needed.

Here is a figure that describes the principle of data reconciliation. On the left we have information about some flows but not all of them. Fortunately, those missing flows can be computed as shown on the right part of this figure to have something which is coherent.

There are two kinds of works in physical units on sectors, one undertakes good inventories of specific sectors in physical units, one focuses on the overall economy with highly aggregated

---

<sup>1</sup>that can be found here [www.flux-biomasse.fr](http://www.flux-biomasse.fr)



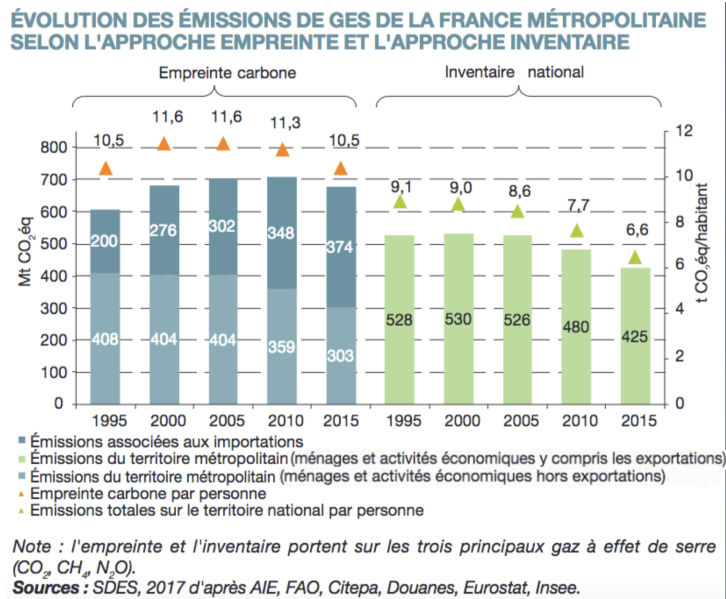


Figure 1.1

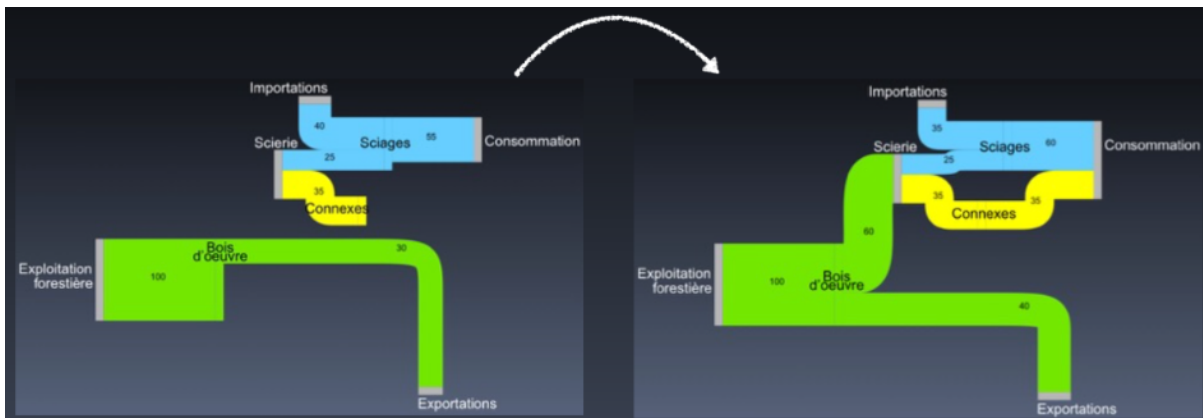


Figure 1.2 – Principle of data reconciliation

data. But currently very few models or works share information on precise interactions between sectors. In the STEEP Team, we have quite precise data on wood and agri-food supply chains<sup>2</sup>. Nevertheless, we lack dependencies with other sectors. If we focus on the wood and try to create a prospective scenario, we will find blind spots that could lead to unfeasibility. A lot of scenarios are based on few sectors and try to detail their evolution. Olivier Vidal reviewed some proposed energetic transition's scenarios to meet Paris' Agreement and stay below 2°C of global warming. In his paper, [Vidal et al., 2018] shows by using accounting methods on physical materials that some of those scenarios demand too much raw material which lead to a global warming above 2°C. Indeed, if we want to replace all energetic infrastructures based on fossil energy source by new infrastructures using renewable energy the number of wind turbines for example should dramatically increase. Yet, we know approximately the composition of a wind turbine and especially in precious metals : it contains (in tons by MegaWatt placed) huge part of concrete (150 t/MW), aggregate (300 t/MW), steel or ferrous metals (150 t/MW) and

<sup>2</sup>[https://www.flux-biomasse.fr/resultats/sankey\\_bois/France](https://www.flux-biomasse.fr/resultats/sankey_bois/France)

some part of copper (3 t/MW), aluminum (2 t/MW), glass (7 t/MW) or plastics (4 t/MW). Those numbers can significantly change depending on the type of wind turbine (offshore wind turbines need even more materials). Thus, the consumption of all those materials will be increasing a lot to build wind turbines. On another example, lets suppose that all fuel sources are replace by biomass to produce biofuel. One major impact is that we will have to use crops to produce large amount of biomass. This will compete with the crops needed to feed the population. Having a systemic model that integrates interactions between sectors and raw materials, consumption and production, energy and pollution becomes essential.

One objective of this work is thus to allow those sector-specific analysis to communicate and impact each other. Consequently, we will be able to conceive new alternatives for our economy that we call Socio-Technical Alternatives (STA) and analyze them. Below is a schematic description of a STA (figure 1.3, p.5).

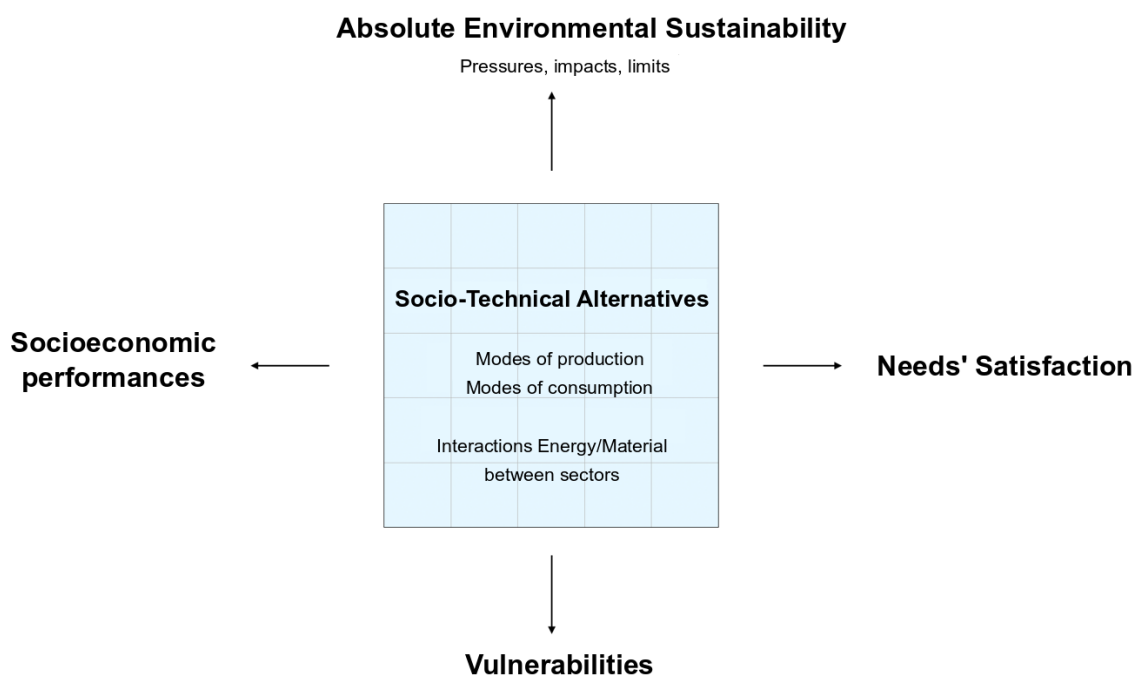


Figure 1.3 – Description of a STA

It mixes a precise comprehension of our way of production and consumption and their interactions and dependencies between sectors with different criteria gathered in four categories: Socioeconomic performances, sustainability, vulnerabilities and needs' satisfaction. On this project, we are all convince of the necessity of a strong sustainability which is reflected in this definition and in the tools used around a STA. A STA is supposed to be (almost) stationary. For now, this stationary condition has not any real meaning but it is convenient for our work.

From this definition of a STA, four objectives arise :

- We need a formalism capable of describing STAs, and thus able to integrate various datasources.
- We need to develop a methodology and linked tools allowing the multi-criteria evaluation of STA.

- We want to propose a methodology which will allow local (or not city, region, state) set of stakeholders to conceive and assess their own alternatives.
- Finally, we want to help to structure public debate by facilitating the comparison of alternatives (e.g., on food, energy transition). In other words, a method is needed to clarify and emphasize the possible compromises that society might choose as compass toward adapting to environmental constraints.

In this internship, we put our focus on the first two points.

Throughout this paper, words scenario and STA will be used. They do not refer to the same concept. A STA is an end-state of an economy or a discrepancy between an initial state and a final state while a scenario refers to a path between two points. Moreover, most scenarios fall in two categories: some focus on specific sectors and are very detailed, the others embrace the whole economy but are less precise. Our goal is to describe STAs taking all sectors into account with flexible levels of details (i.e., using a multisectorial and multiscale model).

### 1.3 Computational issues and operations research

In this project, one huge part is calculations and analysis of large systems of data. To carry this part out, several tools are needed. We mentioned one above, a data reconciliation tool. This tool uses notions of operations research. Indeed, to reconcile data, we rely on constrained optimization. The optimization part comes from the minimization of the discrepancy between the input data and the result by taking into account the reliability of the sources, while the constrained problem comes from the mass conservation law of the products and sectors involved in our system or some coefficient of transformation of some sectors. In other words, it is an inverse problem, in which we try to infer the real (hidden) world from (available) noisy data.

Thus, the objective function currently used in the SC-MFA software is the following:

$$\min \sum \frac{(\hat{x}_i - x_i)^2}{\sigma_i}$$

In which  $\hat{x}_i$  refers to the output,  $x_i$  refers to the input,  $\sigma_i$  refers to the uncertainty. The uncertainty part is made by MonteCarlo simulation. The constraints of the mass conservation law or constraints of aggregation (1) (or as well constraints set by the user (1),(2),(3)) are described below.

$$\begin{aligned} (1) \quad & \sum \alpha_i \cdot \hat{x}_i = k \\ (2) \quad & m \leq \sum \alpha_i \cdot \hat{x}_i \\ (3) \quad & \sum \alpha_i \cdot \hat{x}_i \leq M \end{aligned}$$

You will find in the appendix a paper describing my work during my last year internship.

Another important part of this project is the conflict detection. Indeed, the tools developed here process system that contains a certain quantity of data gathered from different sources. Sometimes, some constraints are inconsistent with others, which can be explained by the fact that they come from different sources. It is inevitable that our system could contain some

conflicts and we must correct them to hope find results. Thus we develop some ways of finding those conflicts as precisely as possible. The paper in appendix describes steps in this direction. Further enhancements are needed to improve performance and to semi-automatically suggest to the user solutions to remove conflicts (either by deleting or relaxing some of the constraints).



## State of the Art and Methods

### 2.1 Input Output Analysis (IOA)

Input-Output tables are one way to represent or analyze the economic structure of a country. According to [Leontief, 1970], this approach describes precisely the dependencies that sectors have with each others by analyzing consumption and production in the whole economy of several goods and services by the Leontief inverse matrix. We will be using the example in the Leontief paper to illustrate how it works: a simple economy with two sectors very aggregated, Agriculture and Manufacture. The matrix  $Z$  will represent the dependencies of sectors (depending of the example) for the rest of this work. (Here,  $Z_{1,2} = 20$ ).

From	Into	Agriculture	Manufacture	Final Demand	Total Output
Agriculture		$Z_{11}$	$Z_{12}$	$Y_1$	$X_1$
Manufacture		$Z_{21}$	$Z_{22}$	$Y_2$	$X_2$

Table 2.1 – Notation

In this kind of table, lines represent production and columns consumption (transitional consumption in  $Z$  and final consumption in  $Y$ ).

From	Into	Agriculture	Manufacture	Final Demand	Total Output
Agriculture		25	20	55	100 bushels of wheat
Manufacture		14	6	30	50 yards of cloth

Table 2.2 – Input-Output Table of an economy

From this table, we can construct another one that details what he calls the “cooking recipes of the two sectors” to produce one unit. For example, Agriculture will need  $25/100 = 0.25$  unit of wheat and  $14/100 = 0.14$  unit of cloth to produce one unit of wheat. Which gives us the following table (table 2.3, p.10).

Through this report, we will refer to those ratio as a matrix  $A$  with  $A_{ij}$  as coefficients. This table can be written with equations where  $X_1, X_2$  represent the total output of Agriculture and Manufacture and  $Y_1, Y_2$  the final demand:

$$\begin{cases} X_1 - 0.25X_1 - 0.40X_2 = Y_1 \\ X_2 - 0.14X_1 - 0.12X_2 = Y_2 \end{cases} \quad (2.1)$$

From	Into	Agriculture	Manufacture
Agriculture		0.25	0.40
Manufacture		0.14	0.12

Table 2.3 – Input Requirement per Unit of Ouput

Thus,

$$\begin{cases} 0.75X_1 - 0.40X_2 = Y_1 \\ -0.14X_1 + 0.88X_2 = Y_2 \end{cases} \quad (2.2)$$

For any  $Y_1, Y_2$  we can solve this system and find (rounded numbers) :

$$\begin{cases} X_1 = 1,457Y_1 + 0,662Y_2 \\ X_2 = 0,232Y_1 + 1,242Y_2 \end{cases} \quad (2.3)$$

Which correspond to the inverse matrix of the system (2). This inverse matrix is the Leontief Inverse matrix. We will refer to this as the matrix  $L$ .

There exists an easier way to reach the Leontief Inverse matrix  $L$  by matrix operations. We note  $\hat{M}$  the diagonal matrix formed by a vector  $M$  and  $\mathbf{1}$  the summation vector (of ones). We can find again our matrix  $A$  by simply doing the operation  $Z \cdot \hat{X}^{-1}$ . From here, since  $Y = X - Z \cdot \mathbf{1}$  like we saw above at 2.1, we can write :

$$\begin{aligned} X &= Z \cdot \mathbf{1} + Y \\ X &= A \cdot \hat{X} \cdot \mathbf{1} + Y \\ X &= A \cdot X + Y \\ X &= (I - A)^{-1} \cdot Y \end{aligned} \quad (2.4)$$

Where  $(I - A)^{-1}$  is the Leontief Inverse matrix  $L$ . Here  $L_{12}$  represents the amount of sector 1 that needs to be produced in order to enable the final consumption of 1 unit of sector 2. The  $L$  matrix is therefore useful to allocate production footprints down to consumption.

Moreover, [Leontief, 1970] shows us that the IOA framework permits to add pollutants dependencies and thus to have the impact of final demand on the pollutant emissions or be able to trace them for instance. Miller and Blair resume very well how this framework could be of great use to our work : “Input–Output Analysis provides a useful framework for tracing energy use and other related characteristics such as environmental pollution or flows of physical materials associated with inter industry activity” [Miller and Blair, 2009] (Chapter 9). IOA framework could also be extended with trade of services or the cost of labour.

Those tables are often used in most of the countries to detail the national economy. However, they use monetary units (economic flows) and it does not exist at present physical tables with pertinent units for each sectors like [Duchin, 2009] said in his introduction : “Input-output economics accords a place of privilege to physical quantities measured in physical units. However, this potential has not been fully exploited because input-output models, like other economic models, are typically implemented using databases in money values only.”

## 2.2 Material and Energy Flow Analysis

### 2.2.1 Introduction to MEFA

Material flow analysis (MFA) is a systematic assessment of the flows and stocks of materials within a system defined in space and time [Brunner and Rechberger, 2003]. It is based on the implementation of the laws of conservation. This method has been applied to substances (e.g. copper [Bonnin et al., 2012]), products (e.g. cereals [Courtonne et al., 2015]), and on the whole economy [EUROSTAT, 2001]. Material and Energy Flow Analysis (MEFA) results directly from MFA by adding the energy aspect.

Interestingly, [Nakamura et al., 2007] propose to couple MFA with the formalism of IOA to build physical IO tables and trace material flows throughout supply chains. We show how in the following section.

### 2.2.2 Coupling MEFA with IOA or AMC to trace flows

MEFA (which can be visualized with Sankey diagrams) inform on direct flows between sectors. In order to estimate indirect flows one can use IOA or AMC to trace flows. We are familiar with the Markov Chains which are a part of the ORCO program in our Optimization under Uncertainties course. We will not detail the theory behind MC, however [Kemeny and Snell, 1976] proposes a way to organize the matrix of transition of a graph by distinguishing two kind of states: transient states and ergodic states. The first represent transition states while the second represent end state in which the flow is absorbed. Lets  $n$  be the number of states and  $s$  the number of ergodic states (thus  $n - s$  is the number of transient states). We can create the matrix  $P$  in which  $P_{ij}$  represents the probability of being in state  $j$  at the next step knowing that we are at the state  $i$  in the current step.

$$P = \begin{matrix} & \begin{matrix} s & n-s \end{matrix} \\ \begin{matrix} s \\ n-s \end{matrix} & \begin{pmatrix} S & 0 \\ R & Q \end{pmatrix} \end{matrix} \quad (2.5)$$

In which:

- $S$  is the transition matrix between ergodic states.
- $R$  is the transition matrix from transient states to ergodic states.
- $Q$  is the transition matrix from transient to transient states (which implies that  $\lim_{n \rightarrow \infty} Q^n = 0$ ).<sup>1</sup>

From our definition of an ergodic states, we can deduce that  $S = I$  which gives us the canonical form of  $P$  :

$$P = \begin{matrix} & \begin{matrix} s & n-s \end{matrix} \\ \begin{matrix} s \\ n-s \end{matrix} & \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \end{matrix} \quad (2.6)$$

We can now easily compute  $P^n$  by the matrix theory.

---

<sup>1</sup>Cycles in  $Q$  may occur in theory but those cycles have no meaning in our studied systems



$$P^n = \begin{pmatrix} I & 0 \\ (I+Q+\dots+Q^n) \cdot R & Q^n \end{pmatrix} \quad (2.7)$$

Since  $Q^\infty = 0$ , we have :

$$P^\infty = \lim_{n \rightarrow \infty} P^n = \begin{pmatrix} I & 0 \\ (I-Q)^{-1} \cdot R & 0 \end{pmatrix} \quad (2.8)$$

This equation means that we have a probability of one to end up in an ergodic state and starting from a transient state  $i$ , the probability to reach ergodic state  $j$  is  $(N \cdot R)_{ij}$  in which  $N = (I - Q)^{-1}$ . We call this matrix  $B$ .

To illustrate, we will use the following example (figure 2.1, p.12),

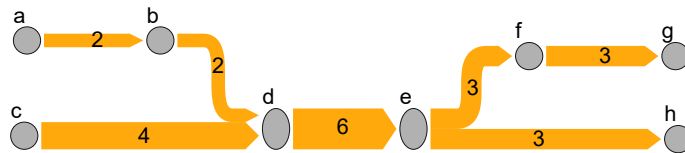


Figure 2.1 – Example for AMC

and his matrix of flows:

$$\begin{matrix} & a & b & c & d & e & f & g & h \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{matrix} & \begin{pmatrix} 2 & & & & & & & & \\ & 2 & & & & & & & \\ & & 2 & & & & & & \\ & & & 4 & & & & & \\ & & & & 6 & & & & \\ & & & & & 3 & & & 3 \\ & & & & & & 3 & & \\ & & & & & & & 3 & \\ & & & & & & & & 3 \end{pmatrix} & \end{matrix} \quad (2.9)$$

From here, we can compute matrix  $Q$  of transient states and  $R$ .

$$Q = \begin{matrix} & a & b & c & d & e & f \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 0.5 \end{pmatrix} & \end{matrix} \quad R = \begin{matrix} & g & h \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{pmatrix} & \\ & \\ & \\ & \\ & 0.5 \\ 1 & \end{pmatrix} & \end{matrix}$$

Then,  $N$  and  $B$ .

$$N = \begin{matrix} & a & b & c & d & e & f \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{pmatrix} 1 & 1 & & 1 & 1 & 0.5 \\ & 1 & & 1 & 1 & 0.5 \\ & & 1 & 1 & 1 & 0.5 \\ & & & 1 & 1 & 0.5 \\ & & & & 1 & 0.5 \\ & & & & & 1 \end{pmatrix} \end{matrix} \quad B = \begin{matrix} & g & h \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 1 & \end{pmatrix} \end{matrix}$$

Which means that every transient state (except  $f$ ) has a probability of 0.5 to end up in  $g$  or  $h$ .

In other words, Absorbing Markov Chains method allows us to know where will end up one unit of any flow and especially the primary flows.

### 2.2.3 Coupling MEFA with LCA to compute footprints

MEFA can be coupled with other frameworks. [Wackernagel et al., 2005] used MFA and the Life Cycle Assessment framework to compute the Ecological Footprint in his paper.

## 2.3 Supply Use Tables

This project aims to use large amount of data and this can become an issue without a proper way to organize it. Supply Use Tables (SUTs) seems to be the most convenient framework to solve this problem. The SUTs organize products and sectors in two tables : the Supply table which describes the origin of the products (table 2.4a) and the Use table for the destination of the products (table 2.4b). [Courtonne et al., 2015]

Lets suppose on the figure 2.1 that the nodes  $b, c, e$  and  $g$  are sectors and  $a, d, f$  and  $h$  are products. We could construct the SU tables as follow.

	$b$	$c$	$e$	$g$
$a$				
$d$	2	4		
$f$			3	
$h$			3	

(a) Supply

	$b$	$c$	$e$	$g$
$a$	2			
$d$			6	
$f$				3
$h$				

(b) Use

Table 2.4

Depending on the question and on data availability, one will rely on IOA/AMC, MEFA or on the coupling of the two. The choice between those approaches depends on the questions we want to answer and on data availability. One huge advantage of the Supply Use framework is that both previous approach can be applied to this formalism. Indeed, constructing a squared matrix from both tables as follow (table 2.5) is enough to build a square matrix that can be processed with IOA or AMC.

	Sectors	Products
Sectors	0	$S^t$
Products	$U$	0

Table 2.5 – Squared matrix from Supply Use tables

## Implementation, experiment, result

All along this internship, we focused on some aspects only of the STA research line described above. The first question was to determine a formalism suited to describe a STA which could be easily used to evaluate them. Since we do not have yet a STA available we decided to base our approach on an existing scenario: NegaWatt.

### 3.1 Implementation of the SU formalism, IOA and AMC on the Negawatt scenario.

NegaWatt is a scenario about the energy transition of France developed by an independent association with the same name. This association gathers about twenty experts involved in professional activities linked to the energy sector. In 2003, they published their first scenario which aims to zero greenhouse gas emission in 2050. Since this first one, they publish new scenario update every five years.

The scenario is based on several hypotheses about the consumption and production of energy in France from now to 2050. For this work, we were only interested in two years: 2015 and 2050 described by two sankeys figure 1 (p.35) and 2 (p.36). Indeed to produce a STA, end-states and data are needed (cf comment at the end of this section). Those two years gives us both. Our first work was thus to express those diagrams into our formalism.

The formalism selected is the SC-MFA formalism of STEEP extended to different units. For this example, all flows are in the same units but since our economy uses many kinds of materials (better expressed in their own « natural » unit), it is relevant to have a formalism in which all data can appear and be understood in its own unit. A first part of the work was to make the Negawatt diagrams fit the SU format by explicitly dividing the nodes of the diagram into products and sectors. Sometimes we created ad hoc products or sectors to shape our model the closest to reality.

Once both diagrams were integrated to our formalism, we started to test it with the different analysis to deduce dependencies between consumption and production sectors. Like we said before, the advantage of the Supply Chain formalism is that we can apply IOA and AMC at the same time. To understand correctly the interaction between producer and consumer, we wanted to know the proportion of resources uses by consumption sectors. To get this, the table 2.5, p.14 gave us  $P$  the matrix of probability of the AMC method. From here, we obtained  $N$  and  $B$  easily

with  $Q$  and  $R$  in  $P$  as [Courtonne, 2016] showed us<sup>1</sup>. Then, our interest was only on the primary resources which gave us those reduced tables (figures 3.1, p.16 and 3.2, p.16):

	Pertes	Exportation	Usage Chaleur Agriculture	Usage Chaleur Industrie	Usage Chaleur Residentiel	Usage Chaleur Tertiaire	Usage Mobilité Agricolture	Usage Mobilité des personnes	Usage Transport de marchandises	Usages Electricité spécifiques Agriculture	Usages Electricité spécifiques Energie	Usages Electricité spécifiques Industrie	Usages Electricité spécifiques Residentiel	Usages Electricité spécifiques Tertiaire	Total	Total Quantité (en TWh)
Biomasse liquide	0.03	0.00	0.01	0.04	0.05	0.02	0.05	0.55	0.26	0.00	0.00	0.00	0.00	0.00	1	42.5
Biomasse solide	0.01	0.08	0.01	0.24	0.59	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1	129.3
Chaleur environnement	0.01	0.00	0.00	0.04	0.92	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1	24.2
Déchets	0.57	0.02	0.00	0.10	0.12	0.08	0.00	0.00	0.00	0.00	0.01	0.04	0.03	0.03	1	13.4
Energie Eolienne	0.07	0.11	0.00	0.00	0.17	0.08	0.00	0.02	0.00	0.02	0.04	0.20	0.13	0.16	1	21.1
Energie Marémotrice	0.07	0.11	0.00	0.00	0.17	0.08	0.00	0.02	0.00	0.02	0.04	0.20	0.13	0.16	1	0.5
Energie Potentielle Primaire	0.07	0.11	0.00	0.00	0.17	0.08	0.00	0.02	0.00	0.02	0.04	0.20	0.13	0.16	1	108.6
Energie Solaire	0.06	0.09	0.00	0.00	0.29	0.07	0.00	0.01	0.00	0.02	0.03	0.17	0.11	0.13	1	8.7
Source d'Uranium	0.69	0.04	0.00	0.00	0.06	0.03	0.00	0.01	0.00	0.01	0.01	0.07	0.04	0.05	1	1226.2
Source de Charbon	0.17	0.01	0.00	0.69	0.05	0.03	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	1	90.7
Source de Gaz fossile	0.14	0.01	0.04	0.24	0.35	0.17	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	1	504
Source de Pétrole	0.07	0.00	0.00	0.06	0.10	0.05	0.04	0.46	0.22	0.00	0.00	0.00	0.00	0.00	1	812.6

Figure 3.1 – From Resources to Consumers Year 2015 (French) Matrix  $B = N \cdot R$  reduced to primary resources

	Pertes	Exportation	Usage Chaleur Agriculture	Usage Chaleur Industrie	Usage Chaleur Residentiel	Usage Chaleur Tertiaire	Usage Mobilité Agricolture	Usage Mobilité des personnes	Usage Transport de marchandises	Usages Electricité spécifiques Agriculture	Usages Electricité spécifiques Energie	Usages Electricité spécifiques Industrie	Usages Electricité spécifiques Residentiel	Usages Electricité spécifiques Tertiaire	Total	Total Quantité (en TWh)
Biomasse liquide	0.43	0.00	0.01	0.03	0.05	0.03	0.06	0.26	0.12	0.00	0.00	0.00	0.00	0.00	1	169.3
Biomasse solide	0.23	0.00	0.04	0.12	0.24	0.09	0.03	0.11	0.12	0.00	0.01	0.01	0.01	0.01	1	352.8
Chaleur environnement	0.04	0.00	0.00	0.12	0.59	0.19	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	1	97.1
Déchets	0.58	0.00	0.00	0.04	0.12	0.22	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	1	5.5
Energie Eolienne	0.21	0.04	0.00	0.02	0.10	0.06	0.01	0.10	0.07	0.03	0.06	0.14	0.07	0.09	1	246.6
Energie Marémotrice	0.21	0.04	0.00	0.02	0.10	0.06	0.01	0.10	0.07	0.03	0.06	0.14	0.07	0.09	1	13.7
Energie Potentielle Primaire	0.55	0.02	0.00	0.01	0.05	0.04	0.01	0.06	0.04	0.02	0.03	0.08	0.04	0.05	1	108.2
Energie Solaire	0.19	0.03	0.00	0.05	0.13	0.07	0.01	0.09	0.06	0.03	0.06	0.13	0.06	0.09	1	166
Source d'Uranium	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0
Source de Charbon	0.09	0.01	0.00	0.55	0.06	0.06	0.01	0.04	0.03	0.01	0.02	0.06	0.03	0.04	1	0.3
Source de Gaz fossile	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0.1
Source de Pétrole	0.14	0.01	0.00	0.46	0.05	0.05	0.02	0.11	0.03	0.01	0.02	0.05	0.02	0.03	1	3

Figure 3.2 – From Resources to Consumers Year 2050 (French) Matrix  $B = N \cdot R$  reduced to primary resources

Read those tables as follow : On the left we can see natural resources. The ratio correspond to the proportion of one unit of natural resources that end up in each consumption sector. Here everything is in TeraWatt Hour. Thus, as an example, you can say that for 1 TWh of Uranium extracted in 2015, 0.69 TWh will end up in losses. While in 2050, for 1 TWh of heat extracted from environment, 0.59 will end up in heat for houses. Those table were created thanks to the IOA method.

One important question could be the inverse situation. In other words: What resources (here primary energies) are necessary to satisfy one unit of consumption of each final consumption sector? On our example with NegaWatt, we could ask ourselves what is the decomposition in primary resources of 1 TeraWatt Hour which end up in the industry sector. One way of doing that is by using the IOA method<sup>2</sup>. Starting from table 2.5, p.14, we obtained those results by getting the Leontief matrix  $L$  and multiplying  $L$  by the matrix of final demand  $Y$ .  $(L \cdot Y)_{ij}$  is the production of sector  $i$  needed to sustain final demand of sector  $j$ . From here, we reduce the matrix to primary resources only and finally, we get figures 3.3, p.17 and 3.4, p.17 by normalizing and transposing the results.

Same as before, each line correspond to the decomposition of one unit of a sector on the left. The heat used by the industry sector comes mostly from fossil gas in 2015 while in 2050 it comes from the biomass for instance.

<sup>1</sup>We put the tables from the excel file of the year 2015 in the appendix 37

<sup>2</sup>tables of the excel file of year 2015 in appendix 40

	Biomasse liquide	Biomasse solide	Chaleur environnement	Déchets	Energie Eolienne	Energie Marémotrice	Energie Potentielle Primaire	Energie Solaire	Source d'Uranium	Source de Charbon	Source de Gaz fossile	Source de Pétrole	Total	Total quantité (en TWh)
Pertes	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,84	0,02	0,07	0,06	1	1012,32
Exportation	0,00	0,14	0,00	0,00	0,03	0,00	0,16	0,01	0,58	0,01	0,06	0,01	1	76,76
Usage Chaleur Agriculture	0,01	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,95	0,00	1	20,33
Usage Chaleur Industrie	0,01	0,12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,24	0,45	0,18	1	263,49
Usage Chaleur Résidentiel	0,00	0,17	0,05	0,00	0,01	0,00	0,04	0,01	0,15	0,01	0,39	0,18	1	462,31
Usage Chaleur Tertiaire	0,01	0,04	0,00	0,01	0,01	0,00	0,05	0,00	0,18	0,02	0,48	0,20	1	180,42
Usage Mobilité Agriculture	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,94	1	35,21
Usage Mobilité des personnes	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,92	1	409,83
Usage Transport de marchandises	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,94	1	187,55
Usages Electricité spécifiques Agriculture	0,00	0,00	0,00	0,00	0,03	0,00	0,18	0,01	0,67	0,01	0,07	0,01	1	12,30
Usages Electricité spécifiques Energie	0,00	0,00	0,00	0,00	0,03	0,00	0,18	0,01	0,67	0,01	0,07	0,01	1	23,72
Usages Electricité spécifiques Industrie	0,00	0,00	0,00	0,00	0,03	0,00	0,18	0,01	0,67	0,01	0,07	0,01	1	122,58
Usages Electricité spécifiques Résidentiel	0,00	0,00	0,00	0,00	0,03	0,00	0,18	0,01	0,67	0,01	0,07	0,01	1	80,29
Usages Electricité spécifiques Tertiaire	0,00	0,00	0,00	0,00	0,03	0,00	0,18	0,01	0,67	0,01	0,07	0,01	1	94,68

Figure 3.3 – From Consumers to Resources Year 2015 (French) Matrix  $L \cdot Y$  reduced to primary resources, normalized and transposed

	Biomasse liquide	Biomasse solide	Chaleur environnement	Déchets	Energie Eolienne	Energie Marémotrice	Energie Potentielle Primaire	Energie Solaire	Source d'Uranium	Source de Charbon	Source de Gaz fossile	Source de Pétrole	Total	Total Quantité (en TWh)
Pertes	0,24	0,26	0,01	0,01	0,17	0,01	0,19	0,10	0,00	0,00	0,00	0,00	1	306,07
Exportation	0,01	0,06	0,02	0,00	0,48	0,03	0,12	0,29	0,00	0,00	0,00	0,00	1	13,85
Usage Chaleur Agriculture	0,09	0,79	0,00	0,00	0,06	0,00	0,02	0,04	0,00	0,00	0,00	0,00	1	16,30
Usage Chaleur Industrie	0,06	0,55	0,16	0,00	0,07	0,00	0,02	0,11	0,00	0,00	0,00	0,02	1	74,23
Usage Chaleur Résidentiel	0,04	0,41	0,28	0,00	0,12	0,01	0,03	0,11	0,00	0,00	0,00	0,00	1	202,67
Usage Chaleur Tertiaire	0,06	0,37	0,20	0,01	0,17	0,01	0,04	0,13	0,00	0,00	0,00	0,00	1	89,12
Usage Mobilité Agriculture	0,34	0,29	0,01	0,00	0,19	0,01	0,05	0,12	0,00	0,00	0,00	0,00	1	25,70
Usage Mobilité des personnes	0,34	0,29	0,01	0,00	0,19	0,01	0,05	0,12	0,00	0,00	0,00	0,00	1	130,58
Usage Transport de marchandises	0,22	0,44	0,01	0,00	0,17	0,01	0,04	0,11	0,00	0,00	0,00	0,00	1	94,69
Usages Electricité spécifiques Agriculture	0,01	0,06	0,02	0,00	0,48	0,03	0,12	0,29	0,00	0,00	0,00	0,00	1	15,00
Usages Electricité spécifiques Energie	0,01	0,06	0,02	0,00	0,48	0,03	0,12	0,29	0,00	0,00	0,00	0,00	1	31,59
Usages Electricité spécifiques Industrie	0,01	0,06	0,02	0,00	0,48	0,03	0,12	0,29	0,00	0,00	0,00	0,00	1	72,78
Usages Electricité spécifiques Résidentiel	0,01	0,06	0,02	0,00	0,48	0,03	0,12	0,29	0,00	0,00	0,00	0,00	1	36,39
Usages Electricité spécifiques Tertiaire	0,01	0,06	0,02	0,00	0,48	0,03	0,12	0,29	0,00	0,00	0,00	0,00	1	48,49

Figure 3.4 – From Consumers to Resources Year 2050 (French) Matrix  $L \cdot Y$  reduced to primary resources, normalized and transposed

Another comment is that those data are only about the energy production and consumption. This could not be a STA by itself. The next question was then to determine how can we add data from different sources into our formalism and the NegaWatt data. This is why we used the year 2015 which is not an end-state but this a year on which we can have access to inventoried data.

## 3.2 Coupling the Negawatt (energy) and Ademe (materials) reports

Thanks to the NegaWatt's partners, we were able to have access to a report from ADEME. ADEME is an agency of the Environment and the mastery of Energy<sup>3</sup> which publishes several reports about sectors closely linked to pollution or natural resources and proposes various expert assessment on the environment. This report ([Rauzier et al., 2019]) proposes a focus on the physical quantities in the current economy with a prospective part. Indeed, the focus is on nine materials and six sectors that consume the most energy in France. We choose one sector (building sector) and two materials (steel and clinker) plus the energy sector (NegaWatt 2015, as studied in the previous section) to analyze and to integrate into our formalism. These sectors and materials were chosen because they are linked to one another and are central in the energy transition debate (e.g. renovation and insulation policies).

<sup>3</sup>[www.ademe.fr](http://www.ademe.fr)

## ADEME's Report

- **Six sectors** : Mechanical-electricity-textile sectors, chemical products sector, packing sector, building sector, energy production sector and transport sector
- **Nine materials** : Steel, clinker, glass, chemical products, plastics, ammonia, paper and cardboard and sugar

So, we extracted and put into our formalism the data from the report for each sector and material. After reconciliation, we obtained the following Sankey diagrams (figures 3.5, p.18 and 3.6, p.19). Every flows are in kilotons unit. We can see aggregate on those diagrams, like we described above, in the steel sector diagram with the node "Tout secteur" and on the other hand we have desaggregated sectors in the building sector diagram on the right side.

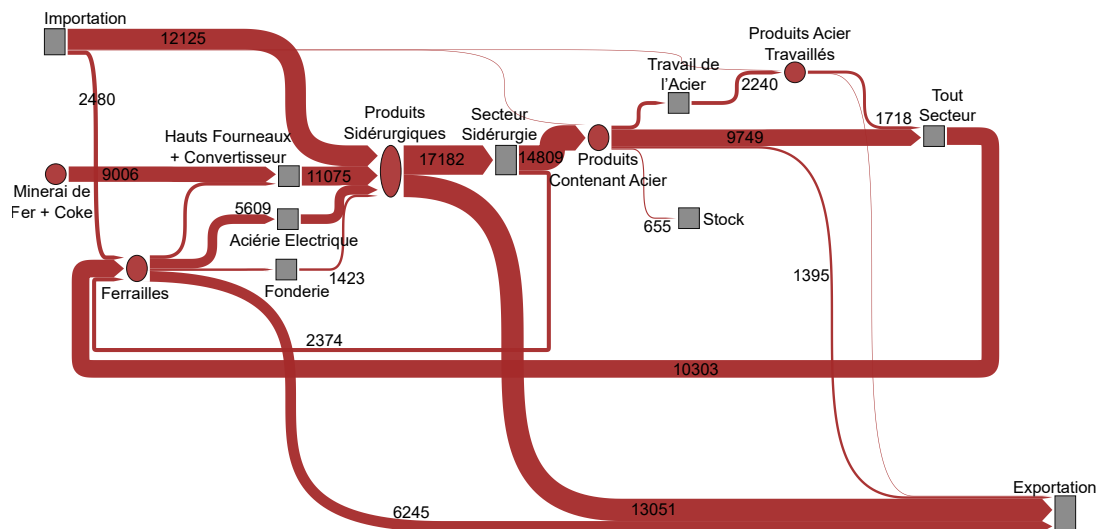


Figure 3.5 – Steel Sector Sankey in kt of iron ADEME data (French)

From this point, we realized we needed a way to merge two excel files in which our sectors were describe in order to meet our objective of having all dependencies in our economy into our model.

### 3.3 Organization of the Excel file

STEEP developed an organization of excel files based on the SU tables that we will describe here. It contains several sheets called : *dim\_products*, *dim\_sectors*, *existing\_flows*, *data* and others that do not interest us for now.

The first two sheets contains a list about each product and sector. On top of that, we will find an aggregation level for each product (or sector) and if it must respect the mass conservation law. The sheet *existing\_flows* details the SU tables without the value, which means those tables are filled with ones and zeros, one if the flow exists and zero otherwise. The last interesting sheet for this work is *data* that will describe the value of each flow, its source, its destination and on which table (supply or use) it will be represented.

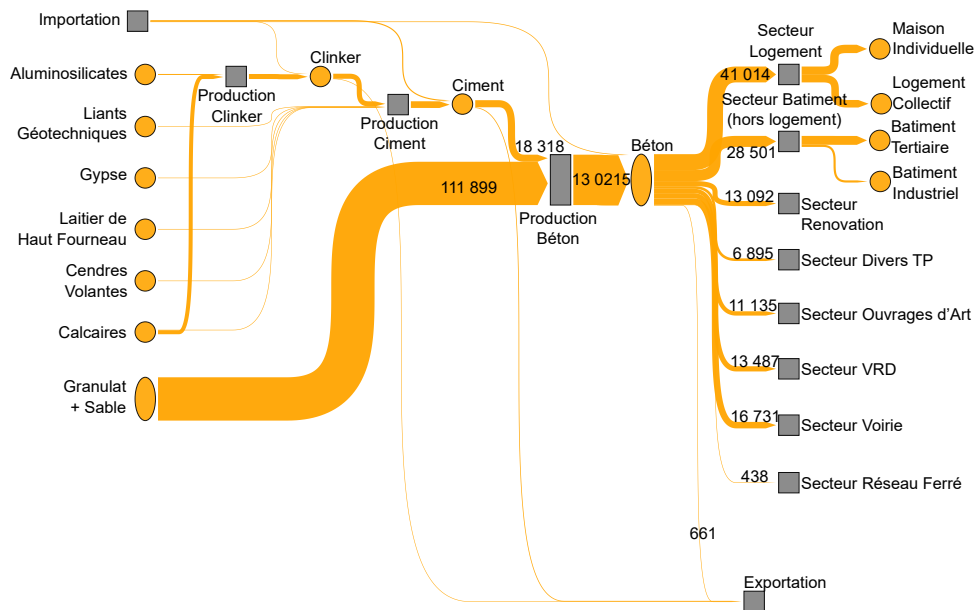


Figure 3.6 – Building Sector Sankey in kt ADEME data (French)

### 3.4 Sectors' Merge

As we said before, the excel files contain different sheets. The first ones detail the aggregation level and the mass' conservation constraint of each sectors and products that we call "dims". To merge those sheets, we need to rebuild the aggregation tree diagram to put the right level of aggregation (root is 1). A second problem involves the existing flows sheet. Since aggregation levels are likely to change with the merge, we need in addition to concatenate the table to propagate to superior level of aggregation the existing flows data as well as merge together products and sectors which appear in both excel files.

To do so, a bit of code was already made, the part in which data are extracted from the excel file. What we added is a simple algorithm to recreate the aggregation tree. In our code, the data are extracted so that an element knows both his parents and his children and other parameters. Thus, when we merge we need to update them at the right time. First, we extract  $D_1$ ,  $D_2$  from the excel files. We browse through the list  $D_2$ , and three cases exist :

- The element is not in  $D_1$  meaning we have to add it.
- The element is in  $D_1$  as a single root (with no child) meaning we can overwrite it with the data in  $D_2$ .
- The element is in  $D_1$  not as a single root meaning we have to update some attributes according to his type in both lists.



---

**Algorithm 1:**

---

**Result:** Merged aggregation tree in  $D_1$

$D_1 \leftarrow \text{Excel}_1$ ;

$D_2 \leftarrow \text{Excel}_2$ ;

**for** For every element  $e \in D_2$  **do**

**if**  $e \notin D_1$  **then**  
        | Add  $e$  to  $D_1$

**else**

**if**  $e$  is a root in  $D_1$  **then**

            |  $e \in D_1 \leftarrow e \in D_2$

**else**

            | update  $e$  according to his type in  $D_1$  and  $D_2$

**end**

**end**

**end**

---

Once the aggregation tree is recreated, we want to write element next to each other in ascending order of aggregation's level. It is also a simple algorithm that starts from a root element, writes his first child and the child of his child and so on ... By remembering when you should go back up in the aggregation tree, the algorithm does not miss any child element. It is simply the Depth First Search of a tree.

To deal with the existing flows sheet, we simply apply an existing algorithm of propagation after a concatenation of table and a merge for data in both file.

We saw above the result of our reconciliation on both Steel and Concrete sectors. Now, here is the result (figure 3.7, p.20).

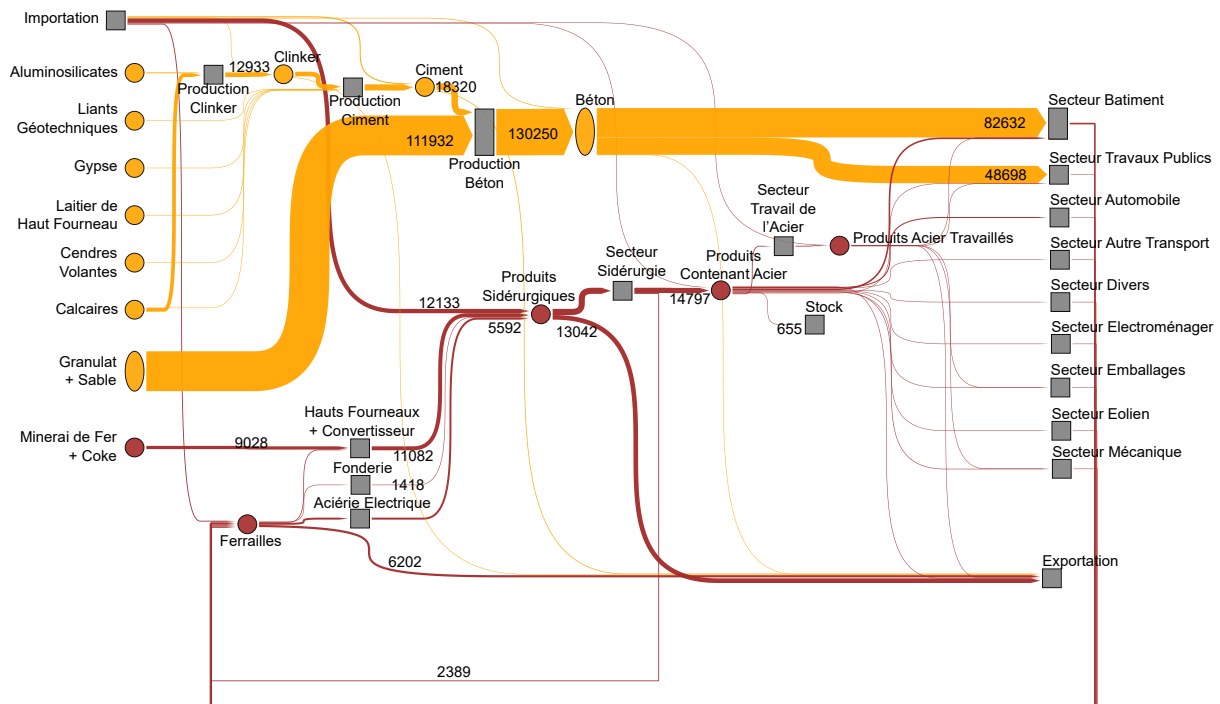


Figure 3.7 – Steel and Building Sectors Sankey in kt ADEME data(French)

One important thing to mention is that our formalism permits to have the same result if you reconcile individually each file or if you merge them into one.

### Remark

The merge with NegaWatt scenario of those sectors has been made. However, our tool is not yet adequate to visualize massive sankey diagrams. Indeed, each diagram needs to be adjusted by hand. Thus, we can not present for now a result with the NegaWatt scenario. Improving our automatic design of Sankey diagrams is one of our objective for the future.

A second issue is that NegaWatt scenario uses TeraWatt Hour while the data in the ADEME's report uses Kilotons. In this special case, there is no reconciliation issue due to the fact that data from both sources are not really intricated. However, we believe that in many case, data could have different units if they come from different sources.

## 3.5 Units Problem and Layers

To explain this issue, here is a simple example. Suppose that we have two sectors : a power station (uses coal and produces electricity) and a factory (uses steel and electricity to produces any end product). On the following graph, we draw the interaction between every product and sector. This is the translation of the adjacency matrix (figure 3.8, p.21).

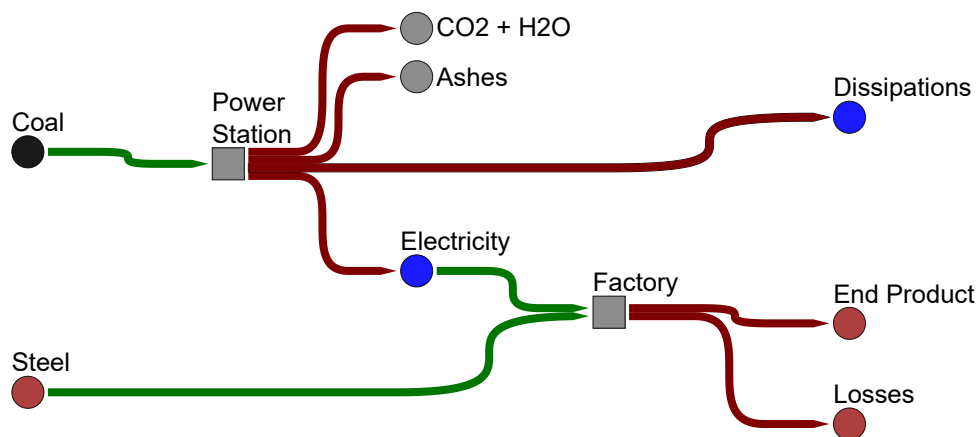


Figure 3.8 – Input data interaction graph of our example

Suppose now that we have knowledge about those data (figure 3.9, p.22).

Some flows are missing and some nodes do not respect the mass conservation law. Thus, we need to use our tool to reconcile this input. Yet, the next diagram shows us that the reconciliation does not work (figure 3.10, p.22). This sankey diagram represents nothing in our reality. The reconciliation tries to minimize changes and simultaneously respects constraints.

This issue comes from the way we reconcile our data. Our formalism permits to store data with several units but to process them and compute the correct solution of mass conservation equations, we need values in the same unit. The first idea that comes to mind is to use conversion

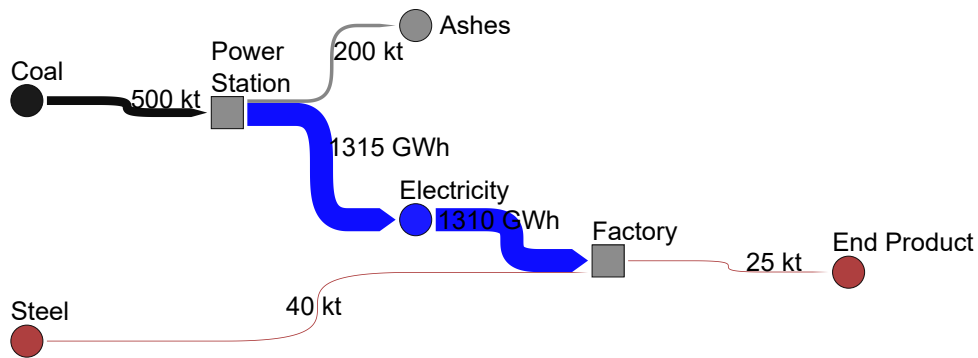


Figure 3.9 – Input data without any treatment

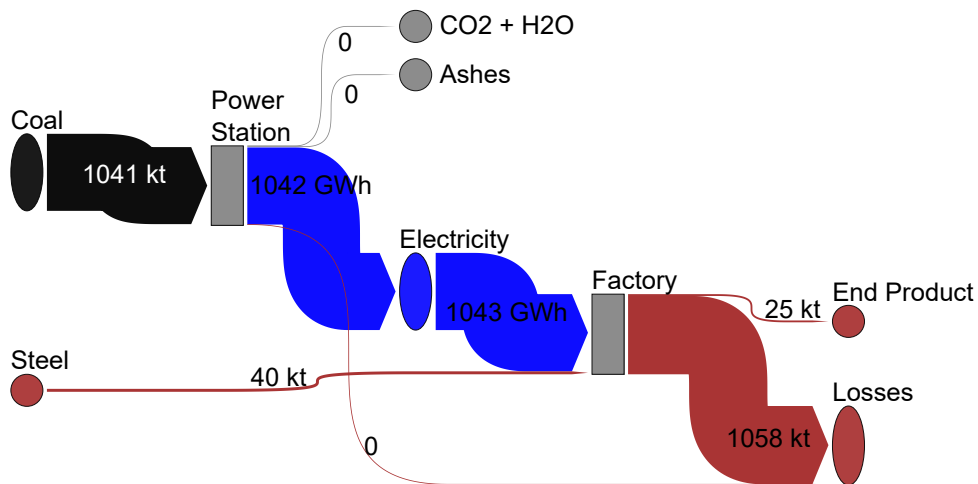


Figure 3.10 – Input data after reconciliation

factors. However, we can not convert every flows in the same units. But this way, we could add information about flows and nodes. For instance, 500 kt of coal go into our power station and this coal carries energy which is 4068 GWh approximately (computed by a conversion factor called the *Higher Heating Value*, HHV). Lets focus only on the energy part, it is a layer energy. If we add the HHV of coal value, we can reconcile on this energy's layer. That gives us (figure 3.11, p.22).

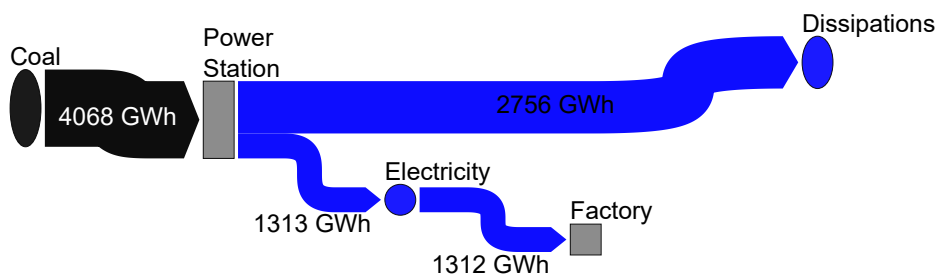


Figure 3.11 – Sankey energy (in GWh) after use of layers, HHV and reconciliation

We can also do a mass' layer and reconcile the mass part (figure 3.12, p.23).

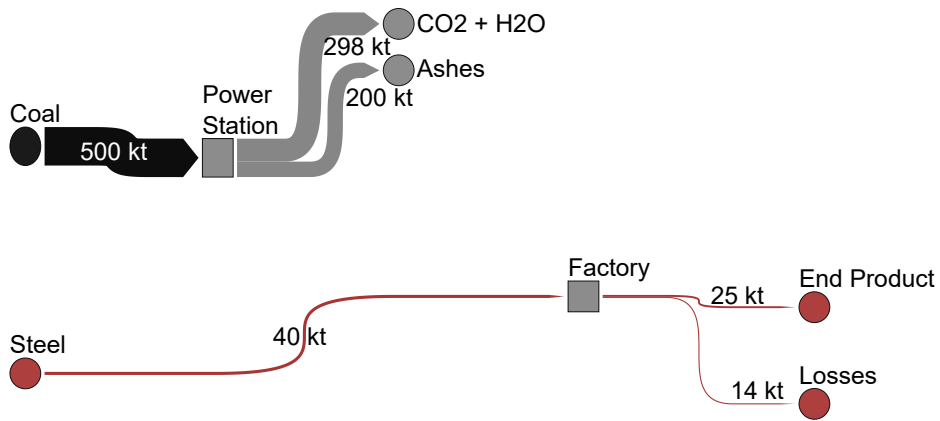


Figure 3.12 – Sankey Mass (in kt) after layers and reconciliation

Thanks to this conversion and the use of layers, we are now able to create a new sankey diagram that represents all interactions in our example (figure 3.13, p.23). Thus, there are 4068 GWh **and** 500 kt of coal that go into our power station and everything is balanced.

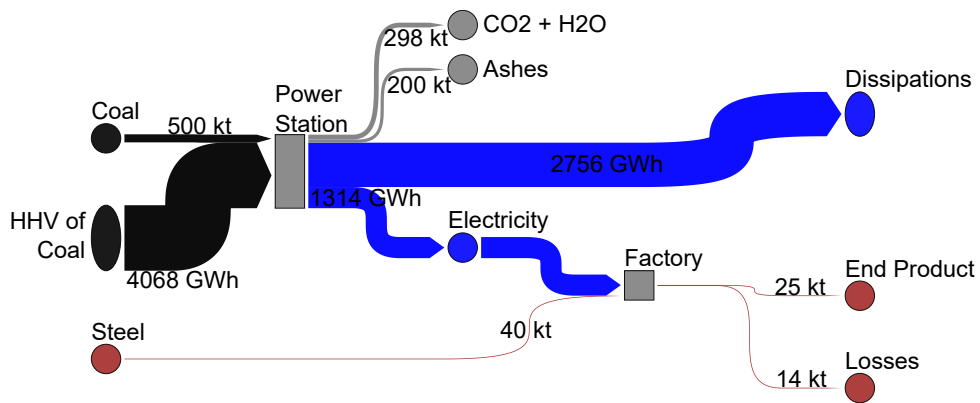


Figure 3.13 – All layers merged into one sankey



## Conclusion

This interdisciplinary internship allowed us to get acquainted with the basic concepts of environmental assessment and with the most common methods used in this field. This demanding task was a prerequisite to determine which methods will be adequate to our problem. We tested one formalism on actual data and many examples. We started to gather together some expert assessment on sectors (such as the ADEME or NegaWatt ones) and we developed a tool to do help us. We noted issues that will improve our work (such as our Sankey tool). Finally we worked on a major issue which is the units problem to find a simple solution for now that will need to be conceptualized and automated. We believe that this first approach of this work will allow us to greatly understand the next biggest stakes.

The major difficulty in this project is its interdisciplinarity and his ambitious scale. In the last month of the internship (after the completion of this manuscript), the work will be to integrate new supply chains in the model, still using the ADEME report and emphasize new issues that may arise.



## Bibliography

- [Boisvert and Vivien, 2006] Boisvert, V. and Vivien, F.-D. (2006). Le développement durable: une histoire de controverses économiques.
- [Bonnin et al., 2012] Bonnin, M., Azzaro-Pantel, C., Pibouleau, L., Domenech, S., and Villeneuve, J. (2012). Development of a dynamic material flow analysis model for french copper cycle. In Bogle, I. D. L. and Fairweather, M., editors, *22nd European Symposium on Computer Aided Process Engineering*, volume 30 of *Computer Aided Chemical Engineering*, pages 122 – 126. Elsevier.
- [Brunner and Rechberger, 2003] Brunner, P. and Rechberger, H. (2003). Practical handbook of material flow analysis, 2a edición. 43-68. *CRC, Boca Ratón, Fla.*
- [Courtonne, 2016] Courtonne, J.-Y. (2016). *Environmental assessment of territories through supply chain analysis : biophysical accounting for deliberative decision-aiding*. Theses, Université Grenoble Alpes.
- [Courtonne et al., 2015] Courtonne, J.-Y., Alapetite, J., Longaretti, P.-Y., Dupré, D., and Prados, E. (2015). Downscaling material flow analysis: The case of the cereal supply chain in france. *Ecological Economics*, 118:67 – 80.
- [Courtonne et al., 2016] Courtonne, J.-Y., Longaretti, P.-Y., Alapetite, J., and Dupré, D. (2016). Environmental pressures embodied in the french cereals supply chain. *Journal of Industrial Ecology*, 20(3):423–434.
- [Duchin et al., 2007] Duchin, Faye, Steenge, E., A., et al. (2007). Mathematical models in input-output economics. *Rensselaer Polytechnic Institute, Troy, NY.*
- [Duchin, 2009] Duchin, F. (2009). *Input-Output Economics and Material Flows*, pages 23–41. Springer Netherlands, Dordrecht.
- [Dugast and Soyeux, 2019] Dugast, C. and Soyeux, A. (2019). Faire sa part ? pouvoir et responsabilité des individus, des entreprises et de l'état face à l'urgence climatique. *Publication Carbone 4.*
- [EUROSTAT, 2001] EUROSTAT (2001). *Economy-wide material flow accounts and derived indicators: a methodological guide*. Office for Official Publ. of the Europ. Comm.



- [Kemeny and Snell, 1976] Kemeny, J. G. and Snell, J. L. (1976). *Markov chains*. Springer-Verlag, New York.
- [Leontief, 1970] Leontief, W. (1970). Environmental repercussions and the economic structure: An input-output approach. *The Review of Economics and Statistics*, 52(3):262–271.
- [Meadows et al., 1972] Meadows, D. H., Meadow, D. L., Randers, J., and III, W. W. B. (1972). *The Limits to Growth*. Universe Books.
- [Miller and Blair, 2009] Miller, R. E. and Blair, P. D. (2009). *Input-output analysis: foundations and extensions*. Cambridge university press.
- [Nakamura et al., 2007] Nakamura, S., Nakajima, K., Kondo, Y., and Nagasaka, T. (2007). The waste input-output approach to materials flow analysis. *Journal of Industrial Ecology*, 11(4):50–63.
- [Rauzier et al., 2019] Rauzier, E., Verzat, B., Letz, T., Rieser, T., Metivier, S., Moteau, S., and Julien, C. (2019). *Transition industrielle - Prospective énergie matière : vers un outil de modelisation des niveaux de production*. ADEME.
- [Vidal et al., 2018] Vidal, O., Le Boulzec, H., and François, C. (2018). Modelling the material and energy costs of the transition to low-carbon energy. In *EPJ Web of Conferences*, volume 189, page 00018. EDP Sciences.
- [Wackernagel et al., 2005] Wackernagel, M., Monfreda, C., Moran, D., Wermer, P., Goldfinger, S., Deumling, D., and Murray, M. (2005). National footprint and biocapacity accounts 2005: the underlying calculation method.

# Appendix

# Efficient computation of solution space and conflicts detection for linear systems

Alexandre BORTHOMIEU  
 University of Grenoble  
 INRIA - STEEP Team  
 Supervised by Jean-Yves Courtonne  
 alexandre.borthomieu@etu.univ-grenoble-alpes.fr

## ABSTRACT

In this paper, we analyze a system of linear inequations  $l \leq A.x \leq u$ , used for the purpose of Material Flow Analysis, with three different but complementary goals : (i) given some known variables  $x_i$ , efficiently compute the solution space of unknown variables, (ii) if the set of constraints is infeasible, efficiently identify the conflicts, (iii) efficiently classify variables to determine whereas they are redundant, just measured, determinable or non-determinable. In each case we compare the efficiency of different algorithms or languages.

## 1. INTRODUCTION

In a context of global environmental changes which questions our modes of production and consumption, the STEEP team of INRIA aims at developing decision-making tools to promote ecological transition at local (sub-national) levels. One of these tools, called Supply Chain Material Flow Analysis (MFA) tries to provide a consistent image of a given chain (e.g., the wood supply chain) in order to stress the potential competition between different types of uses (e.g., energy vs. construction) as well as the material dependency on imports from the rest of the world. In practice, flows are the variables of the problem (some are measured, some are unknown) and the constraints are given by the law of mass conservation and by other rules linking variables (e.g., yields of transformation). The MFA follows the following steps :

1. Reconcile the data with constraint optimization, i.e., minimize the distance between measured flows (inputs) and model results (outputs) given the set of constraints. Reconciliation is necessary because the data comes from various sources which are generally inconsistent. Constraint optimization moreover “fills the gap” for missing flows. However, it provides a unique solution for each variable even when the solution is not unique: for the so-called “free” variables, other values would also respect constraints and lead to the same value of the objective function. It is thus important to identify them.
2. For this purpose, variables are classified based on the set of constraints and on the measured variables in order to determine which ones are determinable (not measured but deductible), free (not measured and not deductible), redundant (measured and staying determinable if the measure is dropped), just measured (measured and becoming free if the measure is dropped).

3. Knowing which variables are “free” and knowing the unique solution for other variables, an interval (solution space) is computed for each free variable.

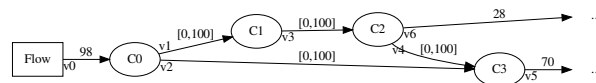
STEEP already developed a python program to conduct these 3 steps. However it proves very inefficient for large problems (about 20,000 variables). Another issue is that in practice, when one builds a large supply chain model, it often occurs that the reconciliation is infeasible because of conflicting constraints, and the manual identification of these constraints is a tedious job. The purpose of our work was therefore to make improvements on these two fronts.

In the following sections, we start by presenting the problem of interval reduction (step 3 above) before turning to conflicts identification. We finish with ongoing and future work, particularly, in relationship with variable classification.

### 1.1 Our examples

To illustrate our algorithms, we will take two examples: the first one to computing the solution space of “free” variables and the second one for the conflicts detection algorithm. Both will take the form of a simple flow-graph as a reminder of our context.

Figure 1: Without conflicts

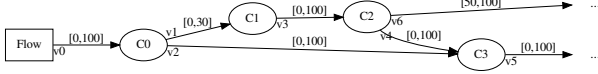


In this first example, we suppose that we have measured or determined some variables ( $v_0, v_6, v_5$ ). Others remain missing. We will apply our algorithm of Intervals Reduction (IR) with the hypothesis that every variable should be between 0 and 100.

We will use this second example with our algorithm of Conflicts Detection (CD). Indeed, in this example, we have added two constraints on  $v_1$  ( $C_4$ ) and  $v_6$  ( $C_5$ ) in order to create an unfeasible system. Those constraints can not be represented by flows because they constrain only one variable. So, we can observe them by looking at the intervals of  $v_1$  ( $[0, 30]$ ) and  $v_6$  ( $[50, 100]$ ).

## 2. INTERVALS REDUCTION

**Figure 2: With conflicts**



To compute new intervals, we work on a linear system. We try to determine for each variables a maximum and a minimum to surround the solution space. Thus, let say we have  $N$  constraints  $c_0 \dots c_{N-1}$  and  $M$  variables  $v_0 \dots v_{M-1}$  such as :

$$\forall i \in [0..N-1], c_i : l_i \leq \sum_{k=0}^{M-1} \alpha_{i,k} v_k \leq u_i \quad (1)$$

with  $l_i$  and  $u_i$  the bounds of each constraint and  $\alpha_{i,k}$  the coefficient of the variable  $k$  in the constraint  $i$ .

In our context, we said that some variables would be fixed as far as we know their values. Thus, let  $F$  be the number of variables known. All variables that are not "free" are known variables (they are uniquely determined). Let  $L$  be the number of unknown variables such as  $M = F + L$ . We can now decompose our first equation (1) :

$$\forall i \in [0..N-1], c_i : l_i \leq \sum_{f=0}^{F-1} \alpha_{i,f} v_f + \sum_{l=0}^{L-1} \alpha_{i,l} v_l \leq u_i \quad (2)$$

Our goal here is to compute the maximum and minimum values that each  $v_l$  can achieve. To do so, a few algorithms already exist.

## 2.1 Similar work

An interesting algorithm can be found in the literature. Indeed, this problem appears similar to the optimization problem of the *Simplex*<sup>1</sup>. The *Simplex* method consists in approaching the optimal solution of an objective function in a problem under duress. Many *Simplex* algorithms have been developed and are extremely efficient. Thus, in their paper, Puranik and Sahinidis [2] develop the idea that if we run two *Simplex* programs for each variable (one to compute the maximum and one for the minimum), we are able to get the interval of each unknown variable. Therefore, we have to run  $2 * L$  *Simplex* to solve our system. In term of complexity, running a such powerful tool two times on variables sub-problem could take some time. But we will detail the efficiency of that algorithm in the part 2.4.

When I joined the STEEP Team, one algorithm was already developed and was particularly long. This algorithm uses the property of each node in the flow-graph. For each constraint, if we want to isolate one unknown variable of

index  $k$  to compute his maximum or minimum, we get :

Let  $k \in [0..L-1], \forall i \in [0..N-1],$

$$c_i \begin{cases} l_i - \sum_{f=0}^{F-1} \alpha_{i,f} v_f - \sum_{l=0, l \neq k}^{L-1} \alpha_{i,l} v_l \leq \alpha_{i,k} v_k \\ \alpha_{i,k} v_k \leq u_i - \sum_{f=0}^{F-1} \alpha_{i,f} v_f - \sum_{l=0, l \neq k}^{L-1} \alpha_{i,l} v_l \end{cases} \quad (3)$$

For each  $v_l$ , we can only get an interval. In order to calculate the value of  $v_k$ , we need to use the maximum or the minimum of each variable  $v_l$  which depends on the coefficient  $\alpha_{i,k}$ . Thus, we generalize this with four cases as follow. We place ourselves in a context of one constraint with two unknown variables to simplify the inequations. With the graph, on a node, we observe in-flow and out-flow. These flows are represented by the sign of the coefficient  $\alpha_X$ . Furthermore, when a flow  $v_l$  (unknown variable) has the same sign than  $v_k$  (unknown variable for which we wish to tighten his interval), we use  $v_{l_{min}}$  (resp.  $v_{l_{max}}$ ) to compute  $v_{k_{max}}$  (resp.  $v_{k_{min}}$ ). On the other hand, if coefficients have a different sign, we use  $v_{l_{min}}$  (resp.  $v_{l_{max}}$ ) to compute  $v_{k_{min}}$  (resp.  $v_{k_{max}}$ ). In this formalization, the coefficients  $\alpha_X$  are positive.

- Case 1 :  $l \leq \alpha_k v_k + \alpha_l v_l \leq u$

$$\begin{cases} v_{k_{max}} \leq \frac{1}{\alpha_k} (u - \alpha_l v_{l_{min}}) \\ v_{k_{min}} \geq \frac{1}{\alpha_k} (l - \alpha_l v_{l_{max}}) \end{cases} \quad (4)$$

- Case 2 :  $l \leq \alpha_k v_k - \alpha_l v_l \leq u$

$$\begin{cases} v_{k_{max}} \leq \frac{1}{\alpha_k} (u + \alpha_l v_{l_{max}}) \\ v_{k_{min}} \geq \frac{1}{\alpha_k} (l + \alpha_l v_{l_{min}}) \end{cases} \quad (5)$$

- Case 3 :  $l \leq -\alpha_k v_k - \alpha_l v_l \leq u$

$$\begin{cases} v_{k_{max}} \leq \frac{1}{-\alpha_k} (l + \alpha_l v_{l_{min}}) \\ v_{k_{min}} \geq \frac{1}{-\alpha_k} (u + \alpha_l v_{l_{max}}) \end{cases} \quad (6)$$

- Case 4 :  $l \leq -\alpha_k v_k + \alpha_l v_l \leq u$

$$\begin{cases} v_{k_{max}} \leq \frac{1}{-\alpha_k} (l - \alpha_l v_{l_{min}}) \\ v_{k_{min}} \geq \frac{1}{-\alpha_k} (u - \alpha_l v_{l_{max}}) \end{cases} \quad (7)$$

In this algorithm, we loop over the unknown variables to compute a maximum and a minimum for each constraint with those properties. We repeat this operation while the intervals are being modified.

## 2.2 Our Algorithm

In order to find a new algorithm, we manually tried to find the intervals for our small example, without looking at existing algorithms. Then, we tried to automate our method in an algorithm. Finally, we compared other algorithms with our own method. We found out that our algorithm was really close to the STEEP Team's one. Actually, the differences came from the order of the for-loop. We loop first over the constraints and then over unknown variables to compute the intervals (instead of doing the opposite). Also, a list of constraints to check was added to increase efficiency. When a variable is modified, we put in this list all constraints in which our variable occurs. We pop out those constraints once they are checked. Thanks to that, we do not need to loop over every constraints each time a variable is modified.

<sup>1</sup>[https://en.wikipedia.org/wiki/Simplex\\_algorithm](https://en.wikipedia.org/wiki/Simplex_algorithm)

---

**Algorithm 1: Intervals Reduction**


---

```

1 intervals reduction ( $M, I, V, u, l$ ) :
   Input :  $M$  a matrix of coefficient,  $I$  a list of the
           intervals,  $V$  a list of the unknown variables,  $u$ 
           and  $l$  the lists of the boundaries of each
           constraint
   Output: a list of the reduced intervals
2  $C \leftarrow$  all the constraints
3 while  $I$  is modified do
4   for each constraint  $c_i$  in  $C$  do
5     Drop  $c_i$ 
6     for each unknown variable  $v_k$  in  $c_i$  do
7       Compute  $v_{k_{max}}$  and  $v_{k_{min}}$  (With the four
           cases (4) - (7))
8       if  $v_{k_{max}} \leq$  upper-bound  $v_k$  then
9         | Modify the upper-bound  $v_k$ 
10      end
11      if  $v_{k_{min}} \geq$  lower-bound  $v_k$  then
12        | Modify the lower-bound  $v_k$ 
13      end
14      if  $v_k$ 's interval is modified then
15        | Add to  $C$  every  $c_j$  in which  $v_k$  occurs
16      end
17    end
18  end
19 end
20 return  $I$ 

```

---

Here, before this algorithm we use a pre-processing to compute the new boundaries (i.e. fixed values are being subtracted from the  $l_i$ ,  $u_i$  bounds) for each constraint as follow :

$$\forall i \in [0..N-1], \begin{cases} l_i \leftarrow l_i - \sum_{f=0}^{F-1} \alpha_{i,f} v_f \\ u_i \leftarrow u_i - \sum_{f=0}^{F-1} \alpha_{i,f} v_f \end{cases} \quad (8)$$

Furthermore, our algorithm stops because either we reach a max iterations that we fixed or the intervals are unmodified which end the while loop.

With this method, we approach the optimal intervals with more efficiency. Indeed, by computing in each constrain all the unknown variables, we obtain a better restriction on the other variables. That is because the intervals of more variables are reduced in one loop. And the more there are restrictions, the more the intervals are quickly tighten to their optimal values.

Moreover, we can apply a few optimization methods easily (such as parallelism) as the form of the two for-loop are simple (thus easily parallelisable). This tool can drastically increase the speed of our algorithm. To confirm such predicate, we have to create tests that will give us the performance data for our analysis. But first, let us illustrate our algorithm with our example.

### 2.3 Application to our example

With our example (Figure 3), the system will be stored in

a matrix as follow :

$$\begin{matrix} & v_0 & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ \begin{matrix} C_0 \\ C_1 \\ C_2 \\ C_3 \end{matrix} & \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & -1 & 0 \end{pmatrix} \end{matrix}$$

$v_1, v_2, v_3$  and  $v_4$  are our unknown variables here. Once this matrix is constructed and the initialization of our intervals is done, we start the loop over the constraints.  $C_0$  will give us new boundaries on  $v_1$  and  $v_2$  such as :

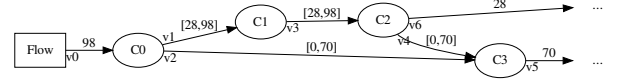
$$\begin{cases} v_{1_{max}} \leq \frac{1}{-1}(-98 + 0) \implies v_{1_{max}} \leq 98 \\ v_{1_{min}} \geq \frac{1}{-1}(-98 + 100) \implies v_{1_{min}} \geq -2 \end{cases} \quad (9)$$

We keep the upper-bound only because we want our interval tightened as possible. Thus,  $v_1$  is in  $[0, 98]$ .

$$\begin{cases} v_{2_{max}} \leq \frac{1}{1}(-98 + 0) \implies v_{2_{max}} \leq 98 \\ v_{2_{min}} \geq \frac{1}{-1}(-98 + 98) \implies v_{2_{min}} \geq 0 \end{cases} \quad (10)$$

We continue the loop over the constraints while there are still some modifications. When the algorithm ends, we get the following result :

**Figure 3: Result Intervals Reduction**



### 2.4 Efficiency

Our tests work as follow : for each algorithm we run four examples of different sizes and store the results. First, we want to warm-up our CPU. Thus, ten call to our algorithm are launched without any measurements of the execution time. Then, we run one hundred instances on our example and we compute the average of those launches. At last, we do this operation ten times and as a result, we get the average of these ten averages<sup>2</sup>. As for now, the final execution time is quite representative of the efficiency of this algorithm.

In this table, each average (computed like mentioned above) is stored depending on the size of the example and the algorithm employed. Each number is in milliseconds (except with the huge example).

	Small <sup>1</sup>	Medium <sup>2</sup>	Medium <sup>3</sup>	Huge <sup>4</sup>
Simplex	113	/	/	/
STEEP	0.303	105.30	438	~1h
STEEP (numba)	0.04	9.04	77	~11min
Our Algo	0.02	2.00	6	~1s
Our Algo (numba)	0.03	0.75	2	76ms
Our Algo (pybind)	0.004	0.12	0.4	31ms

<sup>1</sup> around 10 constraints and 4 unknown variables

<sup>2</sup> around 300 constraints and 150 unknown variables

<sup>3</sup> around 800 constraints and 400 unknown variables

<sup>4</sup> around 18 000 constraints and 16 500 unknown variables

We developed in python language all those algorithms. However, to use parallelism, we use two kind of tools. The

<sup>2</sup>Theorem Central Limit

first one we thought of was the *numba* tool. It is a python library which permits to “translate python function to optimized machine code at runtime”<sup>3</sup>. Thanks to that, we can approach the speeds of a C program. There are many constraints on types of all the arguments, pre-processing was necessary to redefine our parameters. The second tool is called *pybind*. It consists in writing a C++ code of our algorithm with some links to python types. Then compiling it in a python library with many optimization processes. Finally, we can use this library in a python program to compute our intervals faster thanks to parallelism and the C like code.

Concerning the *Simplex* program, we did not succeed to run it on bigger examples and we are trying to understand why. However, this program takes a long time on a very small example compare to the other ones.

### 3. CONFLICTS DETECTION

The conflict detection is used to identify conflicting constraints in order to modify them and make the problem feasible. A variable shouldn’t be getting two disjointed intervals with two constraints. In this case, our previous algorithm will notify us. But once our algorithm detect a conflict, we need to know which constraints are involved in this conflict. Indeed, the source of a conflict can be many loop before it is notified. A variable can be modify several times before encounters a constraint in which it will get a disjointed interval. Moreover, the conflict can come from another variable that affects this one. So, how can we determine which constraints should be verify ?

#### 3.1 Our Algorithm

Our first idea was to store each modification of a variable when IR is running and try to reverse the path that led to a conflict. However, it was extremely difficult to analyze each possibilities to compute the correct one and the complexity of that idea in term of time and space wasn’t satisfactory. Then, we discover the work of J.W. Chinneck [1] that Y. Puranik and N.V. Sahinidis [2] took over. In his article, J.W. Chinneck presents two algorithms to find a subset of constraint which is infeasible in a linear program. The first one called the *Deletion Filter*, consists in deleting constraints one by one until the linear program becomes feasible. Once we reach this point, we put back the last constraint dropped in the set and we keep deleting the other constraints. Here is their algorithm (Algorithm 2).

The second algorithm called *Additive Filter* works on the same idea. We add constraints to a set  $T$  until this set becomes infeasible. Then we add the last constraint to the  $I_S$  infeasible set and start over with  $T = I_S$ . Here is their algorithm (Algorithm 3).

To detect if a set is infeasible, we use our *Intervals Reduction* algorithm. When a variable gets two disjointed intervals, we have an infeasible set :

$$v_{k_{max}} \leq \text{lower-bound } v_k \text{ or } v_{k_{min}} \geq \text{upper-bound } v_k \quad (11)$$

$\implies$  infeasible set

Both algorithms end with one irreducible subset of infeasible constraints. Sometimes, our linear system could have

<sup>3</sup>according to <https://numba.pydata.org/>

---

#### Algorithm 2: Deletion Filter

---

```

1 deletion filter ( $I$ ) :
   Input :  $I$  a set of infeasible constraints
   Output: an irreducible subset of infeasible constraints
2 for each constraint  $c_i$  in  $I$  do
3   |   Temporarily drop  $c_i$ 
4   |   if the set become feasible then
5   |   |   Return  $c_i$  to the set
6   |   end
7   |   else
8   |   |   Drop the constraint permanently
9   |   end
10 end
11 return  $I$ 

```

---



---

#### Algorithm 3: Additive Filter

---

```

1 additive filter ( $I$ ) :
   Input :  $I$  a set of infeasible constraints
   Output: an irreducible subset of infeasible constraints
2  $T = I_S = \emptyset$ 
3 while  $I_S$  is feasible do
4   |    $T \leftarrow I_S$ 
5   |   for each constraint  $c_i$  in  $I$  do
6   |   |    $T \cup \{c_i\}$ 
7   |   |   if  $T$  is infeasible then
8   |   |   |    $I_S \leftarrow I_S \cup \{c_i\}$ 
9   |   |   |   GOTO while
10  |   |   end
11  |   end
12 end
13 return  $I_S$ 

```

---

more than one conflict. To include this possibility, we add a while loop around both algorithms which loop if the set  $I$  is still infeasible without the irreducible subset just found. With this modification we are able to collect all the irreducible subsets of our linear system.

Now that we have two algorithms that compute irreducible infeasible subsets of constraints, we want to imagine a new one that could give us the same result faster. Indeed, in both algorithms, constraints are added or deleted one by one and in a huge linear system that takes time. Y.Puranik and N.V. Sahinidis [2] mentioned in their paper that those algorithms can be modified by adding or deleting constraints by packs.

With this idea, we thought about a new algorithm. This algorithm consists in applying the *Additive Filter* by adding constraints by packs. Then we have  $I_S$  which is not an irreducible infeasible subset but a smaller infeasible subset. We apply the *Deletion Filter* algorithm to this subset. We finally get one irreducible subset. We repeat this until  $I$  becomes feasible. To sum up, we add constraints by pack and we drop them one by one. We called it the *Hybrid Filter*.

### 3.2 Our example

We will process our second example 2 with the algorithm *Deletion Filter*.

First,  $C_0$  is dropped from  $I$ . The set remains infeasible so this constraint is dropped permanently. Now,  $C_1$  is processed, when this constraint is dropped, the subset becomes feasible. That means this  $C_1$  is a part of the problem. So, this constraint will remain in the subset. Next,  $C_2$  will be equivalent to  $C_1$  and  $C_3$  will be permanently removed because the subset remains infeasible without them. This algorithm will now process  $C_4$  and  $C_5$  and find out that they are also a source to our problem. Eventually, this algorithm will give us the following subset :  $[C_1, C_2, C_4, C_5]$ .

### 3.3 Efficiency

The efficiency (measured with the same process details in the section 2.4) of those three algorithms is stored in this table. Like said in the section 3.1 with the expression (11), we use our IR algorithm. In order to reduce the execution time, we choose the IR algorithm with the *numba* structure which is easier to call.

	Small <sup>1</sup>	Medium <sup>2</sup>	Medium <sup>3</sup>	Huge <sup>4</sup>
Deletion Filter	0.89ms	1.2s	10s	+1h
Additive Filter	0.93ms	1.2s	9.5s	+1h
Hybrid Filter	1.60ms	2.3s	18s	+1h

<sup>1</sup> around 10 constraints and 4 unknown variables

<sup>2</sup> around 300 constraints and 150 unknown variables

<sup>3</sup> around 800 constraints and 400 unknown variables

<sup>4</sup> around 18 000 constraints and 16 500 unknown variables

For those three algorithms, an huge example takes too much time. We stop the execution once it exceeds one hour. We notice that our *Hybrid Filter* algorithm is not faster like expected. But, it was predictable. Indeed, in this algorithm, to add a pack of constraints, we are forced by our structure to add them one by one. Thus, the execution time is far more high than expected. The *Deletion Filter* and the *Additive Filter* algorithms show that on particularly large linear system, an algorithm more efficient is needed to find infeasible subsets.

## 4. ONGOING AND FUTURE WORK

We recently realize that our IR algorithm is not totally accurate. Indeed, in a particularly case, the intervals com-

puted are incorrect. When there are dependencies between unknown variables and constraints, some intervals can not be computed correctly. Here is an example, let say we have a constraint such as  $C_0 : a = b - c$  and a second one such as  $C_1 : x + b = y + c$  where  $a, b$  and  $c$  are unknown and  $x$  and  $y$  known. With some inequalities, we have :

$$\begin{cases} 0 \leq a - b + c \leq 0 \\ y - x \leq b - c \leq y - x \end{cases} \quad (12)$$

If some constraints look like those in our linear system, some intervals could be computed with infinite value, which is incorrect. Indeed, in this example, our algorithm will compute  $a$  with infinite bounds instead of  $y - x$ . One idea is to modify our matrix to a special form called *RREF* (Reduced Row Echelon Form) which is a form that we use in our variable's classification algorithm [3]. So, this tool can be easily used for our matrix here. However, this tool also takes time on large matrix. Our current and future work is to transform our current code of *RREF* into a *pybind* structure and then apply it to our matrix. Then, we will verify our assumption with tests and demonstrations. If this idea does not work, we will have to find other methods to have an IR algorithm applicable to any example.

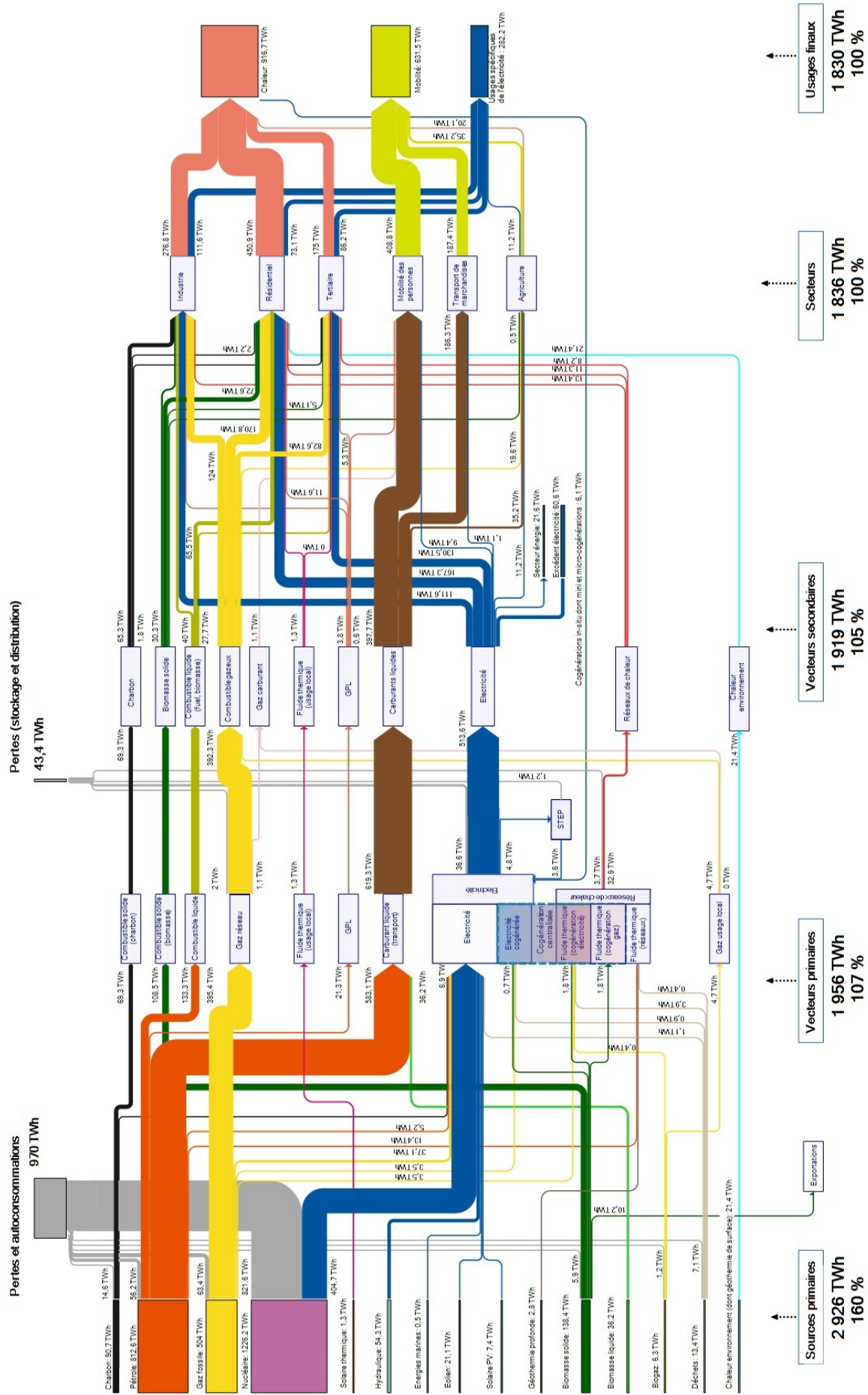
Secondly, like we said in our 3.3 section, the *Hybrid Filter* algorithm add constraints one by one because of our data's type. We could think of a new data structure which will allow us to add constraints by pack and truly exploit our idea. Potentially, we could think of new algorithm with a kind of dichotomy to approach infeasible subsets faster.

## 5. CONCLUSION

Both efficiency tables showed in this paper help us to understand where our theoretical ideas were inaccurate and where we could still gain execution time. This paper gives a new tool to the tightening intervals problem with improvements still possible. This algorithm can now be added to the STEEP project in order to increase their time efficiency on their MFA tool. Indeed, we divide here our execution time by thirty six hundred without any parallelism. This work also gives new ideas to develop a high efficiency on conflicts detection with a large number of constraints. Despite our wrong expectations on our experimental measurements with the *Hybrid Filter*, we know how to transform our data structure to increase the performance on this topic. Moreover, the classification method should now be easily improved by the ongoing work on the *RREF* algorithm.

## 6. REFERENCES

- [1] J. W. Chinneck. Finding a useful subset of constraints for analysis in an infeasible linear program. *INFORMS Journal on Computing*, 9(2):164–174, 1997.
- [2] Y. Puranik and N. V. Sahinidis. Deletion presolve for accelerating infeasibility diagnosis in optimization models. *INFORMS Journal on Computing*, 29(4):754–766, 2017.
- [3] V. Ververka and F. Madron. Material and energy balancing in the process industries, 1997.



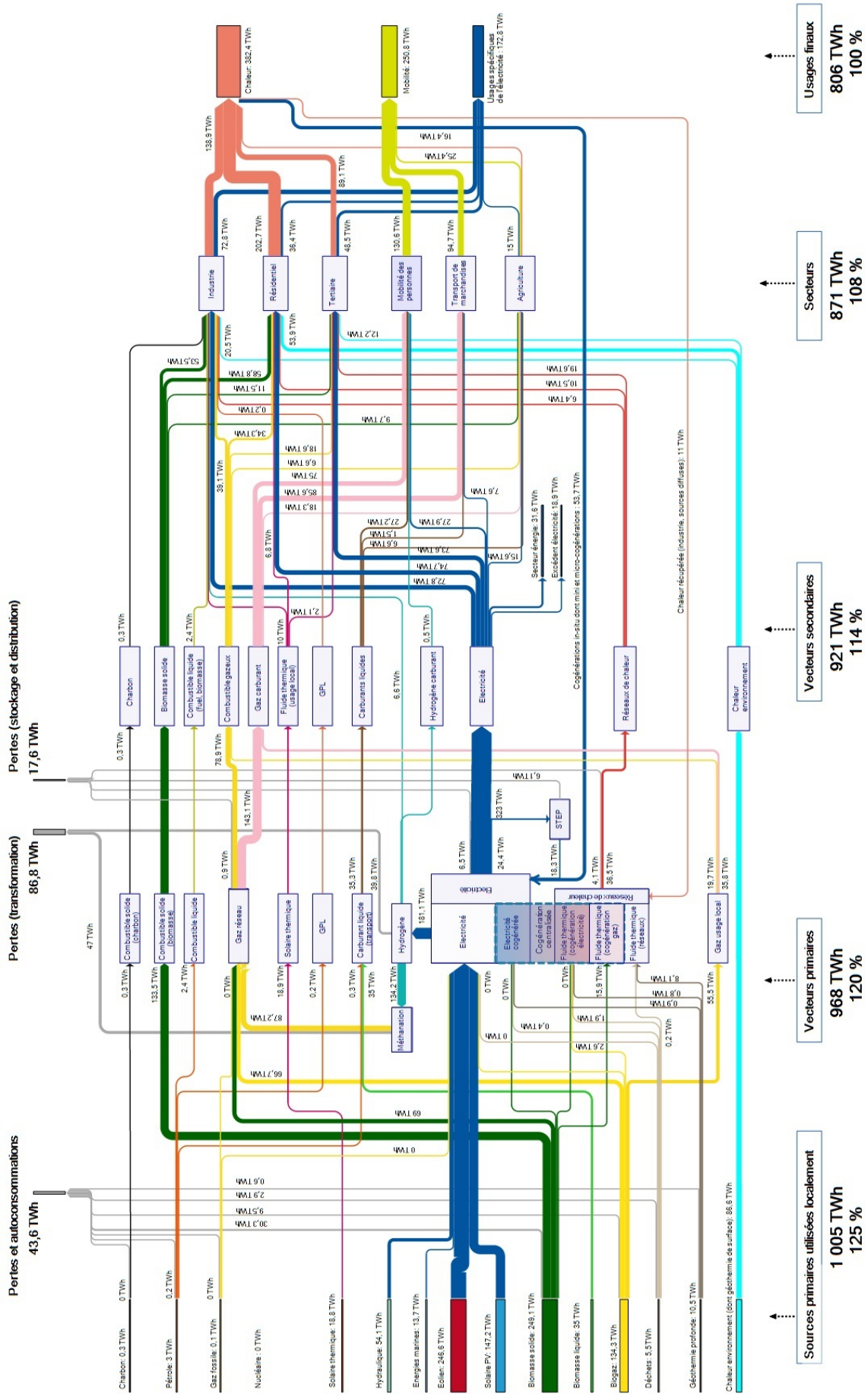
1) Ce diagramme n'est pas une représentation physique des réseaux, mais une représentation des flux et pertes liés aux différentes transformations des énergies depuis les sources primaires jusqu'aux usages finaux. Ainsi, en 2010, le biogaz n'est pas réellement injecté dans le réseau de distribution gaz, mais utilisé localement. Cela ne change rien au bilan présenté.

2) La dilution des valeurs sont arrondies à l'unité, pour ne pas alourdir le schéma. Il peut en résulter des bilans non parfaitement équilibrés.

3) JIP = matières premières

Figure 1 – Sankey 2015 (French)





Nota:

- 1) Ce diagramme n'est pas une représentation physique des réseaux, mais une représentation des flux et pertes liées aux différentes transformations des énergies depuis les sources primaires jusqu'aux usages finaux. Les pertes sont représentées par des lignes grises.
- 2) La plupart des valeurs sont arrondies à l'unité, pour ne pas alourdir le schéma. Il peut en résulter des balancés non parfaitement équilibrés.

Figure 2 – Sankey 2050 (French)

**MATRICE P**

Figure 3 – Matrix *P* of year 2015 Method AMC

**MATRICE Q**

Figure 4 – Matrix *Q* of year 2015





Matrix Z		Matrix Y		Matrix X	
Year	Category	Year	Category	Year	Category
2015	...	2015	...	2015	...
2016	...	2016	...	2016	...
2017	...	2017	...	2017	...
2018	...	2018	...	2018	...
2019	...	2019	...	2019	...
2020	...	2020	...	2020	...
2021	...	2021	...	2021	...
2022	...	2022	...	2022	...
2023	...	2023	...	2023	...
2024	...	2024	...	2024	...
2025	...	2025	...	2025	...
2026	...	2026	...	2026	...
2027	...	2027	...	2027	...
2028	...	2028	...	2028	...
2029	...	2029	...	2029	...
2030	...	2030	...	2030	...

Figure 8 – Matrix Z,Y and X of year 2015 Method IOA





