UCLouvain

École polytechnique de Louvain

The CO_2 Credit Card

CO_2 emissions tracking tool

Authors: **Merlin Guillaume, Alexis van der Plancke**
Supervisors: **Francesco Contino, Hervé Jeanmart**
Readers: **Davide Tonelli, Patricia Luis Alconera**
Academic year 2020–2021
Master [120] in Mechanical Engineering
Master [120] in Computer Science

Abstract

The objective of the thesis is the development of a tool to track the greenhouse gas emissions of Belgian citizens in order to make them aware of their individual environmental impact resulting from their behavior and actions.

At first, a methodology is established to estimate the emissions stemming from various activities or purchases. Databases are also searched to calculate the emission factors needed in our methodology.

In a second part, a mobile application is developed which will help its users to track their emissions. The computation of the emissions is based on the work accomplished in the first part and the desired features of the application are determined by a public survey. In complement to the app, another tool to scan receipts is presented as well.

We conclude with a case study of different consumption habits and the comparison of their emissions.

Acknowledgements

Throughout the work conducted on this master thesis we have received a great deal of support and assistance.

We would like to start by expressing our deepest gratitudes to our supervisors Profs F. Contino and H. Jeanmart, who supported and encouraged us throughout the year also. Their advice and opinions have helped us to develop our critical thinking skills with regard to our work on this thesis. They also helped us realize the larger dimension in which this thesis was taking place.

We would also like to thank our mentor Davide Tonelli for his never-ending support and advice throughout the year. His efforts were not limited to the follow-up of the meetings as he regularly shared with us tips or articles found on the basis of his personal approach. We feel very grateful to have benefited from his great personal involvement in the supervision of our master thesis.

We would also like to thank Pr. P. Luis Alconera for her time and efforts into assessing our work on the master thesis. We also thank her for having shared with us the needed licence to have access to the SimaPro software.

We thank Naïma Sallem, assistant at the EPL, as well for her precious time and support in answering our questions related to the usage of SimaPro software.

Finally, we would like to thank our family, close friends and relatives for their support and the sharing of their opinion on our work throughout the year.

Contents

Abstract **i**

Acknowledgements **ii**

Acronyms **v**

1 Introduction **1**

2 State of the art **2**

3 Methodology **6**
3.1 Division of the emissions by sectors 6
3.2 Selection of relevant sectors 7
3.3 Choice of a metric measure for the emissions 9
3.4 Method for computing the emissions 10
3.5 Taking the uncertainty into account 11
3.6 Search of databases 12

4 Databases **13**
4.1 The agriculture sector 13
4.1.1 The systematic review 14
4.1.2 Results of systematic literature review 16
4.1.3 Emissions factors 18
4.2 The transport sector 21
4.2.1 Emission Factors 26
4.3 The residential heating sector 28
4.3.1 Emission Factors 29

5 CO_2 tracker application **30**
5.1 Application structure 30
5.2 Online Database 31
5.3 Charts 32

5.4 Public survey . . . 33
5.4.1 Audience identification . . . 33
5.4.2 Basic usage of the application . . . 33
5.4.3 Considered features . . . 34
5.5 Database link library . . . 35
5.6 User manual of the application . . . 36
5.7 Our ideal application . . . 36

6 Receipt Scanner **41**
6.1 Image pre-processing . . . 41
6.2 Optical Character Recognition . . . 43
6.3 Pattern matching . . . 43
6.4 Improvements . . . 44

7 Case Study **45**
7.1 Users activity for one week . . . 45
7.2 Processing the users activity . . . 47
7.3 Proof of concept of the application . . . 53

8 Conclusion **54**

Appendix **56**
A.1 GWP values for the greenhouse gas . . . 56
A.2 Reference of the activities used from the ecoinvent v3 database . . . 57
A.3 Emission factors . . . 59

Acronyms

BFM Bone Free Meat.

CI Confidence of Interval.

CO_2 Carbon dioxide.

CV Coefficient of Variation.

EIO Economic Input-Output.

EPD Environmental Product Declaration.

GHG Greenhouse Gas Emissions.

GSD Geometric Standard Deviation.

GWP Global Warming Potential.

IPCC Intergovernmental Panel on Climate Change.

LCA Life Cycle Analysis.

LCI Life Cycle Inventory.

OCR Optical Character Recognition.

RDC Regional Distribution Centre.

UNFCCC United Nations Framework Convention on Climate Change.

Chapter 1

Introduction

Climate change is an important topic nowadays. It is a problem that everyone needs to be aware of and, more importantly, one that needs to be acted upon. The worldwide rise in temperature that has been observed takes its source in the emission of greenhouse gases. In order to reduce this global warming, it is therefore important to monitor and limit those emissions.

The main goal of our work is to create a tool for the purpose of evaluating the carbon footprint of individuals as well as to raise awareness on this subject. In concrete terms, this tool will take the form of a smartphone application where users will be able to assess their CO_2 emissions according to their activities and consumption habits.

A critical point in our work is to be as accurate as possible. To achieve this, the application will use several CO_2 emissions databases, which are of the utmost importance to reflect the typical Belgian emissions. A big challenge for us is to make the application the most easy to use. This is essential as we want to attract and keep users engaged in this fight against global warming.

Our work regarding this matter has been divided between two main axes. The first was the development of a scientific methodology to obtain the most accurate data regarding the CO_2 emissions. The second was the development of the tool itself to make the data found via our methodology accessible and usable for the final aim.

This report follows the same structure. It begins by detailing the methodology that was established and followed, punctuating it with the results obtained. It then describes the development of the IT tool and the various key processes associated with it. The report concludes with a case study analysis of the application.

Chapter 2

State of the art

At the start of our master thesis, the choice of tool interface (i.e. website or smartphone application) was free and the method for computing emissions as well. We have conducted a state of the art analysis to have an understanding about what was already being done. This section presents different services with the same aim as ours, tracking citizens emissions, in the industry or in the academics. It will assess individually the positive and negative points of each service before drawing a conclusion of the main trends observed.

The first similar service is the tool on which the idea of the master thesis is inspired. The DO credit card is a card that tracks the emissions of greenhouses gases of its users based upon their financial transactions. It is supported by the Åland Index[1] which provides the emission factor for the corresponding item or activity purchased.

As positive points, we can note the tracking of a significant proportion of the emissions given that nowadays, most of human activities can be traced via financial transactions. Besides, the service participates in raising awareness on the theme of climate by offering tips and tricks to help reducing the emissions. It also proposes to block the card when exceeding a certain threshold of emissions and the users can compensate their extra emissions via financial support of carbon offsetting projects. As main drawback, the Åland Index[2] on which it is based is not disclosed. One could then doubt the service as to whether it is not greenwashing.

Another drawback is that Belgian citizens can not benefit from the service as it is limited to Sweden so far.

The foodLCA smartphone application developed by Thomas le Varlet[3] was

[1] https://alandindexsolutions.com/

[2] It is a joint venture between Ålandsbanken and Doconomy

[3] https://play.google.com/store/apps/details?id=com.project.foodlca

the next tool to be encountered. It provides carbon emissions data for about more than 150 different types of food. Its aim is to help its users to quantify the carbon footprint of their food recipes. It points to visualizations and comparisons of the environmental impacts of the user's recipes through charts for a well-informed choice. It helps its users to spot the most harmful ingredients in their recipes as well. The positive points is that the origin of the emissions factors and the methodology for their calculation are transparent as it is disclosed within the application. Furthermore, it raises awareness to climate change by displaying directly the positive impact of one change of behavior (e.g. a change in the food diet). A first drawback is that it only covers the sector of food. The manual entry of the data can be time consuming and can thus be seen as a second drawback.

Within the same category of manual entry of one's activity, we encountered the CO_2 calculator proposed by the service "treedom" on their website[4]. First, it provides a CO_2 calculator to give you an idea of how much carbon you emitted throughout the course of a regular day. As in opposition to foodLCA, it covers a wider range of sectors such as: transportation, food and beverage, home, shopping and free time. Once you have assessed your CO_2 emissions of the day, you are being proposed to compensate them by financially supporting farmers in the planting of trees for their agricultural activities.

Through its aesthetic design and playful selection of activities (e.g. a night in a five star hotel room), it offers a fresh and modern approach to climate awareness. The wide range of transports could be seen as a positive side but it leads to a decrease in the number of activities per sector as well. All the activities and purchases of one person are thus not tracked but it allows to make the comparison between the different sectors to see which activity or sector is the most polluting. The defect concerns their methodology for computing the emission factors of the activity as it is not disclosed, which may call into question the credibility of the initiative.

To remain in the sector of online CO_2 calculators, the next tool discussed is proposed by the consulting firm Carbone4[5] specialized in climate change adaptation. Their tool, MyCO2[6], takes the form of a public and free workshop for the citizens willing to know more about their individual GHG emissions. They are given on a voluntary basis by ambassadors whose aim is to spread awareness regarding climate change. It lasts about two hours and it is divided in four parts: the speech of introduction, the diagnostic phase, the sharing of results between the participants and the setting of individual action steps.

[4] https://www.treedom.net/en/co2
[5] https://www.carbone4.com/
[6] https://www.myco2.fr/

The so called “diagnostic phase” is the calculation of one’s individual carbon footprint. In opposition to the entry of data previously seen, it is done via storytelling. To do so, they use a narrative form with fill-in-the-blank texts that participants can fill in by putting the corresponding activities, e.g. “ I like to eat *insert an option* for breakfast”. For this example, the participants will be proposed different answers such as; cereals with milk, fruits, eggs with sausage or a piece of bread with chocolate paste. Depending on the option chosen, myCO2 will take the corresponding emission factor and compute an average emission for the year. The fill-in activities are divided into four categories: transport, housing, eating and shopping. A fifth category of emissions called “my public services” is added and it is a constant value for everyone. It takes into account the indirect emissions resulting from public services such as the heating of public buildings (e.g. schools, municipal houses), police activities, and so on.

A positive side is the interactive and playful sides of attending a workshop. It is a good approach for raising awareness with regards to climate awareness because it tackles the big lines with a warm ambiance which is key for public acceptance. It makes it more easy to educate as they share interesting results of today’s situation. For example, participants get told the average emissions of GHG of a French person (e.g. 10 tonnes per year) and that the objective of the Paris Convention is to achieve two tonnes per year for 2050. The next added value of this service is the taking into account of the indirect emissions into the category “my public services”. It helps to remind that climate change can be addressed by behavioural changes at all levels, from citizens to politicians and the measures they put in place. The drawback of using a narrative fill-in survey is the loss in accuracy since you sum up the activity of your whole year in one scenario. The variation of the activities are not being taken into account into range of possible values for the final outcome. Furthermore, we asked to be provided with the emission factors they use but we didn’t get them. Once could assume that since it is a private service, the methodology around the computation of emission factors is not yet disclosed.

The association Avenir Climatique proposes on their webiste[7] a similar tool to MyCO2 in terms of data entry. It uses a narrative survey by providing multiple options again. Their motive is the same: allowing to evaluate your total annual individual carbon footprint and by major categories (food, transport, housing, miscellaneous, public services, digital), to situate it in relation to the climate objectives in order to take action at an individual level with personalized measures depending on our own behavior.

It is based on an analysis of existing calculators at the beginning of 2020, it is

[7] `https://avenirclimatique.org/calculer-empreinte-carbone/`

based on the MicMac model of the associations “Avenir Climatique”[8] and “Agir pour le Climat (TaCa)” [9].

It is highly similar to MyCO2 but with the positive side that it has a transparent methodology that is supported by recognized environmental agencies including ADEME (i.e. the French Environment and Energy Management Agency).

The state of the art has shown that there are many tools that allow someone to calculate their emissions. They can take different forms (e.g. website, smartphone application) and use different ways to insert data (e.g. storytelling, entry of the quantity of one activity). Among the tools review, there lies a global trend that assessing one’s GHG emissions is not the final aim. In each case, it was a tool to promote a further course of action such as the financial supporting of an environmental project or the taking of individual actions. As a follow-up to the state of the art analysis, we decided to choose as a tool a smartphone application. It seems to us to be the tool that is accessible to the majority of the population. More information about the development of the tool can be found in Chapters 5 and 6. The decision has also been made to track emissions by isolated activities and not by story telling as we want a regular tracking. The first reason is that it leads to a higher accuracy within the results as one-year projections are avoided. The second reason is to enable the users of the application to directly see the results on their emissions through behavioural change. Last but not least, the final wish is to provide a transparent and easily accessible methodology to the users of the application to gain their trust.

[8] `https://avenirclimatique.org/`

[9] `https://www.taca.asso.fr/`

Chapter 3

Methodology

It is not an easy task to assess the emissions resulting from a human behavior. The human consumption and activities responsible for the emissions are numerous and coming from different sources. Given the broad and never-ending nature of the problem we are tackling, it is essential for us to establish a methodology to provide a path and a framework for our research. We have created it by setting out in chronological order the different points:

3.1 Division of the emissions by sectors

The emissions have been divided by industrial sectors in order to simplify the understanding of the phenomenon. This is a current approach that is already done by different agencies when assessing the polluting activities of a country. Regarding the assessment of the emissions for Belgium, the Belgian Federal Public Service for the Environment considered eight different sectors[1]:

- Transport: this sector covers domestic travel by car, boat and plane. It does not cover the emission costs associated with travel for imports and exports.
- Commercial heating: this sector covers the heating of the commercial buildings.
- Residential heating: this sector cover the heating of Belgian's citizens home.
- Agriculture: this sector cover the main categories of livestock and cultivation businesses in Belgium.
- Waste: this sector covers the waste coming from industry and municipalities.
- Industry (energy): this sector covers mainly the production of public electricity and heat.

- Industry (process): this sector covers emissions from industrial activities which does not result from the combustion of fossil fuels.
- Industry (combustion): this sector covers emissions from industrial activities, resulting from the combustion of fossil fuels.
- Others: the remaining polluting activities. No further information is given on the nature of the different activities.

Those sectors are commonly used in the scientific literature. Furthermore, they have been used by the environmental service of our country, which is the geographical framework of our study. It is therefore logical to consider these sectors for the next steps of the master thesis.

3.2 Selection of relevant sectors

The choice of the different sectors had to be aligned with the main purpose of the master thesis: we had to gather the tracking of at least 80% of the Belgian citizens' emissions. To make our choice among the different sectors, we based ourselves on the following graph provided by the Belgian Federal Public Service for the Environment[2]. It represents the share of the sectors in the production of emissions resulting from Belgian activities in 2019.

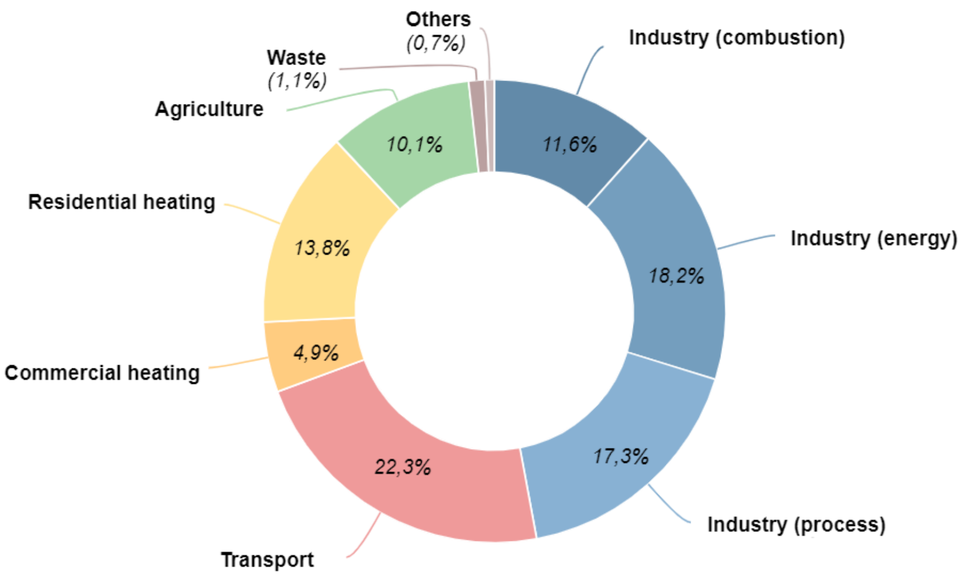


Figure 3.1: Belgium's GHG emissions gathered by sectors

We have transposed the results under the following bar chart in order to better visualize the most and least polluting sectors.

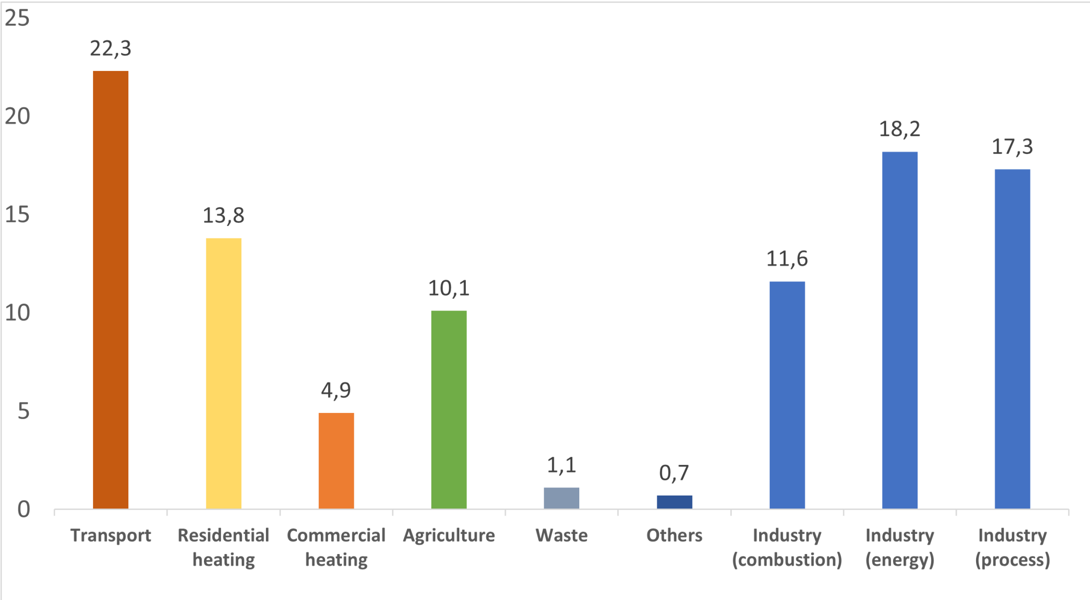


Figure 3.2: Belgium's GHG emissions gathered by sectors

Since the aim of the thesis is to track the emissions that are being directly emitted from citizens' activities, we can remove the sectors such as the industry sectors. The emissions coming from those sectors are indirect and can not be tracked by the citizens themselves using our tool.

From the resulting sectors, we have chosen the following sectors: the agriculture, the transport and the residential heating sectors based on the $> 80\%$ criteria. According to Fig. 3.2, those sectors are contributing for more than 88% of the pollution the Belgian citizens are directly responsible off. The objective of the master thesis is therefore well respected.

These three sectors present different benefits. Since the citizens have a direct influence on these sectors, it is easier for them to realize the impact of their activities. One citizen could easily see, via the usage of our application, the difference of emission resulting from a change of routine in those sectors, if he would choose for example for one week to go to work by bike rather than using a car. The same stays true if he would decide to go vegetarian for one full month. This is due to the fact that these sectors present large differences in terms of global warming potential within the range of their activities. It thus allows to directly assess the impact of a change of behavior on the individual greenhouse gasses emissions. We believe this will contribute to the aim of our tool to raise awareness towards climate change.

3.3 Choice of a metric measure for the emissions

This master thesis was conducted with having in mind a desire of tackling global warming. In order to do it properly, we have to take into account the emissions of carbon dioxide (CO_2) but not only. There are other greenhouses gas emissions that are also responsible for climate change such as methane (CH_4) or nitrous oxide (N_2O). The United Nations Intergovernmental Panel on Climate Change (IPCC) have identified more than forty greenhouse gases in their "Fourth Assessment Report (AR4)"[3], which was the fourth paper in a series designed to assess scientific, technical and socio-economic information concerning climate change and its potential effects.

To assess the impact on climate change of the different greenhouse gases, they assigned to each greenhouse gas a value called "Global Warming Potential". They defined it as the heat absorbed by any greenhouse gas in the atmosphere, as a multiple of the heat that would be absorbed by the same mass of carbon dioxide CO_2 over the same period. Carbon dioxide is thus the reference and has a GWP of 1 regardless of the three time period used. For the other greenhouse gases, the values of GWP range from 1 to 22800 depending on the gas and on one of the three time period: 20 years, 100 years or 500 years. Those values are periodically compiled, revised and published in the same serie of reports from the IPCC.

From those GWP values, can be derived the carbon dioxide equivalent (CO_2e, CO_2eq or CO_2-eq). It is defined, for a greenhouse gas, as the amount of carbon dioxide (CO_2) that would cause the same radiative forcing as that gas, i.e. that would have the same capacity to retain solar radiation. It is calculated as GWP times the amount of the gas.

$$m_{CO_{2eq}} = m_{GHG} \cdot GWP_{GHG} \tag{3.1}$$

It provides a common scale for measuring the climate effects of different gases since all the impacts of the different greenhouse gases are reduced to the same unit. Under the Kyoto Protocol, the "Conference of the Parties" decided that the GWP values calculated in the Second Assessment Report of the IPCC should be used to convert various greenhouse gas emissions into CO_2e for international reporting. This standard was updated in 2013 by the Warsaw meeting of the United Nations Framework Convention on Climate Change (UNFCCC, Decision 24/CP.19) stating that the values should be updated with the ones published by the Fourth Assessment Report of the IPCC.

The report is the largest and most detailed summary of the climate change situation ever undertaken, produced by thousands of authors, editors, and reviewers from dozens of countries, citing over 6,000 peer-reviewed scientific studies and thus was chosen as a reference based on this success. The international consensus to apply the values obtained in the AR4 is still applicable today. This was verified in our search of the scientific literature where the vast majority of the articles and databases followed the consensus. Indeed they express the emissions of an activity in terms of CO_2e per unit of the corresponding polluting action studied, e.g. CO_2e per km for transport activities or CO_2e per kg for agriculture activities for example. Those factors are expressed as followed:

$$\frac{\text{kg } CO_2e}{\text{quantity of activity}} \tag{3.2}$$

They are defined in the scientific literature as "emission factors". To obtain the emissions of CO_2e corresponding to one activity, one can multiply the quantity by the emission factor of the activity.

3.4 Method for computing the emissions

Computing the GHG emissions of one single activity is a difficult process because one needs to take into account lots of different stages of the process. For example, one should consider the emissions of transporting a passenger 10 kilometers by car. For this activity, one cannot just take into account the emissions of GHG resulting from the burning of the fuel during the ride. You have to consider into account the earlier stages of the process such as the manufacturing of the car, the building of the different warehouses and production facilities, the establishment of the roads. You might need also to consider the future stages of the process such as the recycling of the different components of this activity including the car, the roads, the warehouses.

The type of study that takes into account those different process is called a Life Cycle Analysis (LCA). It is a standardized evaluation method (ISO 14040 and 14044) that allows a multi-criteria and multi-stage environmental assessment of a system (product, service, company or process) over its entire life cycle, from the extraction of the raw materials necessary for its manufacture to its treatment at the end of its life (dumping, recycling...) or more commonly referred to as "from cradle to grave". Regarding the procedural model of a LCA, the "Code of Practice for LCA"[4] stands out currently as the most widely recognized. It divides LCA into four distinct components: (1) scoping; (2) compiling quantitative data on direct and indirect materials/energy inputs and waste emissions; (3) impact assessment; and (4) improvement assessment. For this master thesis, we only focus on the three first stages which define the climate change impact assessment. There are two fundamental types of LCA: process-based data, and environmental input-output (EIO) data. Unit process data are derived from direct surveys of companies or plants producing the product of interest, carried out at a unit process level defined by the system boundaries for the study. EIO data are based on national economic input-output data. Since you can retrieve useful information from the analyses, we decided not to limit ourselves to one of the two methodologies.

One should not forget about the drawback of using LCAs. Given the model is more complex as it takes different stages in the process, the resulting uncertainties regarding the values for the emissions of GHG for one activity are more difficult to assess. We still opted for LCA regarding GHG emissions assessment since it is the most relevant methodology as it takes into account all the emissions produced at the different stages of the process. This is the primary means to understand and assess an activity's environmental impact.

3.5 Taking the uncertainty into account

There are uncertainties around the data measured due to the fact that greenhouses gases emissions assessment is a relatively new field of research. Therefore there isn't a wide consensus over the resulting CO_2e emissions coming from the different activities. Their values vary from one database to another depending on the methodology used for assessment and the quality of the measuring equipment. For some activities, there are a large number of studies that were done, which provide databases of values. From those databases there are values that are repeated and others that are further apart. This breakdown of values can be derived into a distribution with a mean value and a standard deviation.

Among the different distribution we chose to go for the Gamma distribution since it only provides positive values. Producing negative emissions would physically not be coherent. It would have been the case for some low-GHG emitting activity such as a bike ride if we would have chosen the normal distribution for example.

With regards to this work, it was important to have a distribution and uncertainties since we know there isn't a consensus over the emission factors of the activities yet.

The procedure adopted is to accumulate the different mean values and putting confidence of interval to neglect the outlier values. By doing this, we reduce the bias in the measurement of the values and our values become more relevant. We opted for a 95% CI. A 68% CI seemed to us too narrow taking into account that by definition LCA analysis are not precise due to the lack of precision for every stages of the process of the activity. There is thus probably a large amount of correct values between the 68% and the 95% confidence of intervals.

3.6 Search of databases

Our original intention for assessing the emissions of one activity was first to search and collect all the papers or databases we could find assessing the GHG emissions. As a filter for the search, we established criteria such as the location of the study, the year it was done, the transparency of the methodology for calculating emissions, whether it was grey or white literature and whether it displayed the results in terms of CO_2e or not. From the remaining articles, the respective values of the emission factor of the activity would then be weighted according to the same criteria to obtain via a weighted arithmetic mean the corresponding mean CO_2e value and its corresponding Gamma distribution parameters.

Many obstacles were encountered with this intention. First, there is a lack of open access LCA data in the public domain as most of the databases are payable. Second, the remaining different LCA databases don't always take into account the same processes of the LCA for the same activity. It is thus complicated to weight their emission values correctly, as is the case for weighting on the basis of location or the year in which the study was conducted. We thus quickly realized that it was an exhaustive work to do by a group of two students if it had to be done for all the activities responsible for at least >80% of the GHG-emission of one citizen.

The next intention was thus to try and find papers that would already have done the work of assessing and weighting the values. The criteria for the search of databases assessed:

- The location of the studies that the databases based itself upon. The countries needed to have a similar industry as ours (i.e. occidental).
- The methodology of the database and/or the studies it based itself should be disclosed.
- The studies conducted should have been done recently ($>$ 20 years old).
- A sufficient amount of literature is taken into account.
- The results are disclosed in terms of unit measure CO_2e.
- statistical data regarding the uncertainties around the mean value are provided.

Different databases were successfully obtained for the different sectors selected: agriculture, transport and residential heating.

Chapter 4

Databases

Once the methodology was set, we started the search for databases with regards to the different sectors. We started with the sector of agriculture as we knew from previous searches that databases were quite common for this sector. As a result from the search, we were lucky to find one single academic paper containing all the information we needed. Once the sector of agriculture was covered, we proceeded again on the search for databases for the sector of transport and residential heating. This time the result was different as it led to base ourselves on one of the world's leading Life Cycle Inventory (LCI) database, ecoinvent v3. The values retrieved from those databases are displayed in the following sections for analysis and comparison. The reader can also find them in their entirety in the appendix.

4.1 The agriculture sector

There are numerous databases for the agriculture, like Agribalyse, World Food LCA Database, and so on. As being mentioned in the previous Section 3.6, it was a task too demanding to use the data from these different databases for a lot of different reasons previously detailed. Therefore we based ourselves on the paper "Systematic review of greenhouse gas emissions for different fresh food categories" who actually did the work of reviewing the databases of the agriculture sector [5].

More specifically, this article presents the results of a systematic review[1] of the literature on greenhouse gas emissions for different food categories from Life Cycle Analysis studies and databases. It was written with the aim of enabling streamline calculations that could inform dietary choice based on the GWP values of the food types consumed for individual meals, diets, catering organizations, or

[1]Systematic reviews attempt to gather all available empirical researches by using clearly defined, repeatable analytical methods to obtain answers to research questions that are broad or narrow in scope

on a nation level. The paper also provides a meta-analysis of the LCA studies for the following categories: fresh vegetables, fresh fruits, staples, dairy, non-ruminant livestock and ruminant livestock. The meta-analysis presents the median, mean, standard deviation, upper and lower quartile, minimum and maximum results for each food category.

4.1.1 The systematic review

The review protocol The systematic review was completed following the PRISMA Statement protocol[2] to minimize the risk of bias, and increase scientific validity. The systematic literature search for food LCA studies was completed in February 2015 across three types of literature: peer reviewed scholarly journal papers, conference proceedings and Environmental Product Declarations (EPD).

The initial studies reviewed identified additional studies for review by two methods: first by scanning the document text and reference list for additional studies, and second using the *cited in* function from Google scholar to identify relevant articles for review. They also identified grey literature (i.e. papers and articles coming from the industry and governments) through these two mechanisms. The inclusion of grey literature to avoid bias is viewed as a best practice.

The filtering of the studies Studies were included in the meta-analysis if they disclosed the LCA results in terms of CO_2e/mass unit for raw produce and if they disclosed the system boundary, functional unit and location of production. Studies were excluded if they included processes like cooking, air-freight or canning without disclosing the percentage that these activities accounted for as they significantly alter the results.

The setting of common boundaries The boundaries of the search can be seen on Fig 4.1.

[2]PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. It is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. It focuses on ways in which authors can ensure a transparent and complete reporting.

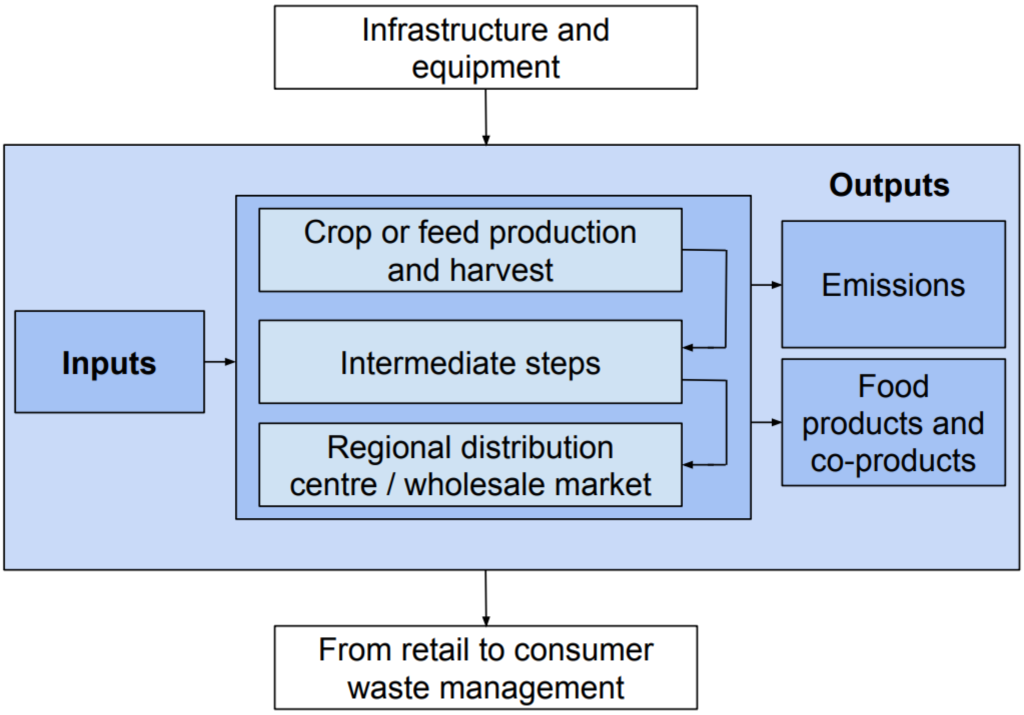


Figure 4.1: Boundaries of the LCA system[5]

The boundaries of the LCA system go from the crop or feed production and harvest to the regional distribution center/whole sale market, including the transport of the food between all the intermediate steps of production. The LCA studies thus typically analysed farm inputs from chemicals and fertilisers, fuel and energy inputs from irrigation and machinery for cultivation, harvesting and processing and transport and refrigeration to the regional distribution centre. The outputs include emissions released from fertilised soils, plants, animals in fields and the transport of the food. Human consumption, including how consumers travel to shops, store food, cook, dispose of food and packaging, and excrete were outside the scope of the study because it was only about making dietary choices, and they were excluded from entries entered into the database. To enable comparison between papers having different scope of studies, the Global Warming Potential values were converted to the system boundary of the regional distribution centre. Due to the high number of paper studied, they managed to have GWP values for the different stages such as seen on Table 4.1.

Life cycle stages post-farm gate	Mean	Standard deviation	Min	Max
Processing meats	0.66	0.14	0.54	0.87
Processing vegetables	0.07	0.04	0.01	0.13
Packaging	0.06	0.06	0.01	0.21
Transport to RDC	0.13	0.10	0.02	0.95
Retail	0.10	0.25	0.01	1.14

Table 4.1: Post farm gate emissions identified by a sample of studies

The packaging and transport median figures from Table 4.1 were for example added to studies where the system boundary finished at the farm gate.

The synthesis of results GWP values from the reviewed studies were put into a database under the following broad category headings: fresh fruits, vegetables and staples, dairy, non-ruminant livestock and ruminant livestock. In addition, data relating to the LCA method were collated including:

- Year of study
- Geographic location of study
- Original system boundary
- LCA approach used[3]
- Unique descriptors (e.g. species, feed type, farming methods etc.)

Each GWP value recorded was converted into a common functional unit and system boundary in kg CO_2e /kg bone free meat (BFM) (or produce, at the regional distribution centre.

4.1.2 Results of systematic literature review

The meta-analysis cites 369 published LCA studies that provided 1,718 GWP values for fresh produce from the year 2000 to 2015. The sources of the literature are quite diverse as they contain 192 journal papers, 80 conference papers, 64 reports (for industry and government), 29 web-based EPDs, and four theses.

The majority of GWP values (58%) were from the last five years before the paper was written (i.e. from 2010 to 2015).

[3]Three types of LCA were considered : process based, economic input-output or hybrid LCA

Regarding the location of where the studies have been conducted (see Table 4.2), the vast majority was on the Europe continent with the British Isles and Europe accounting for 68% of the utilised GWP values. The other continents were less represented. Within the European countries, Spain (n = 187), France (n = 173), Sweden (n = 153) and the Netherlands (n = 139) were the most represented.

Region	Number of recorded GWP values
Europe	930
British Isles	245
North America	167
Oceania	143
Asia	77
South America	74
Rest of the world	82
Total	1718

Table 4.2: Location of the GWP searches that were found in the database

Overview of the results The results present a large variation in emission factors between food categories. This overview indicates a clear greenhouse gas hierarchy emerging across the food categories, with grains, fruit and vegetables having the lowest impact and meat from ruminants having the highest impact. The figures can be found in the Section 4.1.3.

Results of our methodology This paper was chosen as the articles and the meta-analysis study it provides meet all the criteria of the methodology.

- The methodology is transparent and well documented.
- The majority of the studies had taken place in Europe and North America accounting together for 77% of the utilised GWP values. The dominant countries in Europe have a similar agriculture industry as ours.
- The majority of the GWP values are recent, 58% of them dating from 2010 to 2015
- The results are disclosed in terms of CO_2e per activity with the mean value and the corresponding uncertainty (i.e. the standard deviation).

The retrieval of the Gamma Distribution The meta-analysis provides statistical parameters but only for a normal distribution. We proceed on the following to obtain the parameters for the appropriate gamma distribution.

Let's note μ and σ respectively the mean and the standard deviation of the normal distribution, α and β respectively the shape and the scale parameters of the Gamma distributions. Let's consider the following theorem:
If a random variable Y has a gamma distribution with parameters α and β, then

$$\mu = E(Y) = \alpha\beta \tag{4.1}$$

$$\sigma^2 = V(Y) = \alpha\beta^2 \tag{4.2}$$

The proof of this theorem can be found in this statistics reference book[6] at page 187. Having a system of two unknowns for two equations, we can retrieve the following formulas:

$$\alpha = (\frac{\mu}{\sigma})^2 \tag{4.3}$$

$$\beta = \frac{\mu}{\alpha} \tag{4.4}$$

We thus obtain the parameters for the Gamma distribution. To obtain the 95% confidence interval of the Gamma distribution, we derive them numerically by using the Excel function `GAMMA.INV(probability, alpha, beta)`. It returns the cumulative distributive function of a Gamma distribution for the probability, alpha, and beta arguments. Thus to obtain the lower boundary we give the value 0.025 to the variable "probability" and the value of 0.975 for the upper boundary.

4.1.3 Emissions factors

For the following section, please note that the mean and standard deviation are expressed with two decimal digits due to the precision provided by the article. For greater precision, the other figures are expressed with three decimal values.

At the broadest level, the lowest median GWP values were for field-grown vegetables (0.37 kg CO_2e/kg) or field-grown fruit (0.42 kg CO_2e/kg).

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Onions	0.18	0.11	0.033	0.451	2.678	0.067
Potatoes	0.20	0.08	0.075	0.384	6.25	0.032
Carrots	0.22	0.15	0.029	0.596	2.15	0.102
Mushrooms	0.27	0.29	0.004	1.060	4.251	0.626
Cucumbers	0.33	0.32	0.010	1.19	1.063	0.31
Zucchinis	0.42	0.50	0.003	1.806	0.706	0.595
Beans	0.62	0.45	0.069	1.76	1.90	0.327
Broccolis	0.70	0.34	0.200	1.51	4.239	0.165
Peanuts	0.87	0.11	0.668	1.098	62.55	0.014
Lentils	1.10	0.45	0.403	2.14	5.975	0.184
Eggplants	1.35	0.07	1.216	1.453	371.939	3.62E-3
Pears	0.33	0.13	0.126	0.629	6.443	0.051
Oranges	0.35	0.12	0.156	0.621	8.507	0.041
Apples	0.36	0.19	0.089	0.816	3.590	0.100
Tomatoes	0.46	0.18	0.178	0.874	6.531	0.070
Cherries	0.48	0.40	0.032	1.52	1.44	0.333
Peaches	0.54	0.24	0.177	1.102	5.063	0.107
Strawberries	0.65	0.36	0.146	1.522	3.260	0.199
Melons	0.88	0.01	0.86	0.900	7744	1.13E-4

Table 4.3: Emission factors and statistics of vegetables and fruits

Rice had the highest impact of the plant based field grown crops (2.55 kg CO_2e/kg). Dairy products (cheese) and butter shared slightly higher GWP values.

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Rice	2.66	1.29	0.763	5.721	4.25	0.626
Soy milk*	0.88	0.27	0.433	1.482	10.62	0.083
Milk*	1.32	0.29	0.814	1.950	20.72	6.37E-2
Yoghurt*	1.43	0.25	0.983	1.960	32.718	0.044
Eggs	3.39	1.21	1.449	6.141	7.849	0.432
Cheese	8.86	2.07	5.278	13.354	18.32	0.484
Butter	11.52	7.37	1.859	29.810	2.443	4.715

Table 4.4: Emission factors and statistics of plant and dairy products

* For those products, the values are expressed in terms of $\frac{\text{kg } CO_2e}{\text{liter food}}$.
The fish had an average GWP value of around 3.49 kg CO_2e/kg.

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Pilchard	1.10	0.45	0.403	2.141	5.975	0.184
Herring	1.17	0.17	0.861	1.526	47.37	0.025
Mackerel	2.00	1.08	0.473	4.606	3.429	0.583
Tuna	2.60	1.45	0.577	6.117	3.215	0.809
Salmon	3.76	1.47	1.454	7.142	6.542	0.575
Mussels	7.54	4.93	1.148	19.821	2.339	3.223
Shrimp	14.85	12.37	0.981	47.027	1.441	10.304
Lobster	21.74	11.70	5.180	49.95	3.453	6.297

Table 4.5: Emission factors and statistics of fishes

Non-ruminant livestock have medium GWP values: chicken (4.12 kg CO_2e/kg) and pork (5.60 kg CO_2e/kg) while the ruminant livestock such as beef (26.05 kg CO_2e/kg) and lamb (33.84 kg CO_2e/kg) have the highest median GWP values.

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Chicken	4.12	1.72	1.469	8.112	5.737	0.718
Pork EU	5.60	1.51	1.469	8.112	3.04	8.92
Beef EU	26.05	6.78	14.50	40.92	14.76	1.7646
Lamb EU	33.84	13.06	13.29	63.83	5.737	0.718

Table 4.6: Emission factors and statistics of livestock

4.2 The transport sector

For the transport sector, the search for databases was conducted via the software SimaPro[4] instead of searching online. The motivation behind it is for time saving as SimaPro provides a transparent access to many LCI databases. Those LCI databases provide a value of the different components of an activity such as its input and output flows. Such flows include respectively inputs of water, energy, and raw materials, and releases to air, land, and water. Since the SimaPro is a LCA software package, it provides methods as well that compute the impact of the inputs and outputs flows regarding environmental and human health harms such as global warming, acidification, ozone depletion, human toxicity, etc.

Since we have access via the outputs to the amount of GHG emitted by activities, we can then proceed on assessing the climate change impact of those activities in terms of CO_2e. For this we can use the methods provided to compute the climate change impact based on emissions outputs given by the LCI. There are more than ten different methods. Some tackle only climate change and are therefore labeled as "singles issues method" like the methods "IPCC 2013 GWP 100a" or "Greenhouse Gas Protocol". Other methods tackle climate change but not only, and are therefore labeled as "multiple issues" like the methods "Impact 2002" or "ILC 2011 Midpoint+".

The original intention was to assess the climate change impact of one same activity (e.g. the transport by essence car of one passenger on one kilometer) for all the databases where this activity was included, by using one same method, the most relevant one. By going this way, we would have obtained different values for the emission factor of the activity. We could have then chosen how to assess the values using an appropriate weighting. The obstacle faced is that from all the databases included in SimaPro, only one database has data with regards to transport activities. It is ecoinvent v3.

The ecoinvent v3 database was made by researchers from the EPFL university. It is a Life Cycle Inventory that is made based on the review of lots of LCA and academic paper studies but also grey literature. It meets our various criteria: it offers Life Cycle Analyses conducted in Europe, they have been done in the last 20 years, it takes into account uncertainties around the LCA they based themselves upon and the access to their methodology is transparent. Furthermore, it gathers a wide range of transport activities ranging from a transport by plane to a transport by electric bike. We decided to keep the most common activities for the Belgian citizen: transport by plane, by car (electric or essence), by train, by bus or bicycle.

[4] SimaPro is one of the world's leading life cycle assessment software package. The Louvain School of Engineering have licenses for this software.

Further details about the activities can be found in the Appendix A.2.

The second intention was then to do it the opposite way, to assess the climate change of one activity from the database ecoinvent v3 using the different methods assessing climate change. The method of weighting would then be used again, but this time with regards to the methods according to their scientific methodology and transparency. The issue faced with this matter is that, as a proof of the wide consensus mentioned earlier in the section, they are for the vast majority returning the same results in terms of CO_2e since they are using the same values of GWP for the greenhouse gases, and the same database (ecoinvent). The slight difference in terms of results that occur with some databases comes from the fact that they use more recent versions of the GWP value, in opposition to the consensus encouraging the use of the values coming from the report of 2013[7] .

We opted then to choose one method. The method "IPCC 2013 GWP 100a" was chosen since it is the reference given it is designed by the IPCC itself, the reference institute in terms of climate change assessment.

Having the appropriate database and the appropriate method to compute the emission factors of the different activity, one last issue was faced regarding the computation of the uncertainties for one. Indeed, when methods assessing the contribution of an activity return the value in terms of CO_2e, they take into account the mean values of the emissions of GHG but don't take into account the uncertainties around the emissions of the data, and thus don't provide some. There is thus a geometrical mean value for the emissions of CO_2e but a method still needs to be designed to take into account the uncertainties defined by the geometric standard deviation. To do so, the concept of setting of the geometric standard deviation and how the value is attributed needed to be better understood.

The ecoinvent database assesses the uncertainty around the data of one activity via a pedigree matrix such as displayed on Fig 4.2.

Indicator score	1	2	3	4	5 (default)
Reliability	Verified[3] data based on measurements[4]	Verified data partly based on assumptions ***or*** non-verified data based on measurements	Non-verified data partly based on qualified estimates	Qualified estimate (e.g. by industrial expert)	Non-qualified estimate
Completeness	Representative data from all sites relevant for the market considered, over an adequate period to even out normal fluctuations	Representative data from >50% of the sites relevant for the market considered, over an adequate period to even out normal fluctuations	Representative data from only some sites (<<50%) relevant for the market considered ***or*** >50% of sites but from shorter periods	Representative data from only one site relevant for the market considered ***or*** some sites but from shorter periods	Representativeness unknown or data from a small number of sites ***and*** from shorter periods
Temporal correlation	Less than 3 years of difference to the time period of the dataset	Less than 6 years of difference to the time period of the dataset	Less than 10 years of difference to the time period of the dataset	Less than 15 years of difference to the time period of the dataset	Age of data unknown or more than 15 years of difference to the time period of the dataset
Geographical correlation	Data from area under study	Average data from larger area in which the area under study is included	Data from area with similar production conditions	Data from area with slightly similar production conditions	Data from unknown ***or*** distinctly different area (North America instead of Middle East, OECD-Europe instead of Russia)
Further technological correlation	Data from enterprises, processes and materials under study	Data from processes and materials under study (i.e. identical technology) but from different enterprises	Data from processes and materials under study but from different technology	Data on related processes or materials	Data on related processes on laboratory scale ***or*** from different technology

Figure 4.2: Pedigree matrix of the ecoinvent 3.0 database[8]

This matrix evaluates the methodology of the papers returning value for one input or output of one activity. It bases itself on different criteria and gives a number ranging from 1 to 5 for each criteria. Each input or output parameter of an activity has thus a pedigree matrix associated with it.

Depending on the value for those criteria, they will be converted into uncertainty factors using the table displayed on Table 4.7.

To determine the total uncertainty for a flow (input or output) lognormally distributed, the following formula is then applied:

$$\sigma_g^2 = \exp\sqrt{[\ln(U_1)]^2 + [\ln(U_2)]^2 + [\ln(U_3)]^2 + [\ln(U_4)]^2 + [\ln(U_5)]^2 + [\ln(U_6)]^2 + [\ln(U_b)]^2}, \quad (4.5)$$

Indicator score	1	2	3	4	5
Reliability	1.00	1.05	1.10	1.20	1.50
Completeness	1.00	1.02	1.05	1.10	1.20
Temporal correlation	1.00	1.03	1.10	1.20	1.50
Geographical correlation	1.00	1.01	1.02	1.05	1.10
Further technological correlation	1.00	1.10	1.20	1.50	2.00
Sample size	1.00	1.02	1.05	1.10	1.20

Table 4.7: Default uncertainty factors coming from criteria of pedigree matrix

with:

- U_1: uncertainty factor of reliability
- U_2: uncertainty factor of completeness
- U_3: uncertainty factor of temporal correlation
- U_4: uncertainty factor of geographic correlation
- U_5: uncertainty factor of other technological correlation
- U_6: uncertainty factor of sample size
- U_b: basic uncertainty factor

From this, we then retrieve the geometrical standard deviation of one input or output flow.

Now that the method around the computation of the geometrical standard deviation of one flow is better understood, one should know how to use this information properly. The geometrical standard deviation (GSD) cannot be used as a common standard deviation since it is dimensionless. Nevertheless the value of a normal standard deviation can be retrieved based on the GSD. For this, we based ourselves on the paper[9]. This article was written by the authors of the database ecoinvent to help convert the lognormal distribution into other more useful distributions.

Let's note μ the normal mean, μ_g the geometrical mean, σ the normal standard deviation and σ_g the geometrical standard deviation.

Under the following assumptions, they state that:

$$\sigma_g = \exp(\sqrt{\ln{(CV^2 + 1})} \tag{4.6}$$

$$CV = \sqrt{\exp(\ln^2 \sigma_g) - 1} \tag{4.7}$$

The acronym CV stands for coefficient of variation and it measures the extent to which data is separated. To obtain a normal distribution, they then state that under the following assumptions:

$$\mu = \mu_g \tag{4.8}$$

$$CV = \frac{\sigma}{\mu} \tag{4.9}$$

the following results are verified:

$$\mu = \mu_g \tag{4.10}$$

$$\sigma = \mu CV \tag{4.11}$$

Those assumptions were verified via Monte Carlo simulations[9].

Once the normal value and the normal standard deviation for the emissions of each output flow of GHG for one activity were obtained, we proceed on achieving our objective to have a normal mean and standard deviation for the activity.

For the mean, we proceed by doing the following:

$$\mu_{ghg,i} = m_{ghg,i} \times GWP_{ghg,i} \tag{4.12}$$

$$\mu_{activity} = \sum_{i}^{n} \mu_{ghg,i} \tag{4.13}$$

where n equals to the number of GHG output flows taken into account for one activity (i.e. n = 17 in our case). For the standard deviation, we proceed by doing the following:

$$CV_{ghg,i} = \sqrt{\exp(\ln^2 \sigma_{g_{ghg,i}}) - 1} \tag{4.14}$$

$$\sigma_{ghg,i} = CV_{ghg,i} \cdot \mu_{ghg,i} \tag{4.15}$$

$$\sigma^2_{activity} = \sum_{i}^{tot} \sigma^2_{ghg,i} \tag{4.16}$$

$$\sigma_{activity} = \sqrt{\sigma^2_{activity}} \tag{4.17}$$

The parameters of the gamma distribution are then obtained by applying the formulas 4.3 and 4.4 found previously in the agriculture sector.

The GWP values of the greenhouses gases provided by the database ecoinvent v3 can be found in the appendix A.1. 17 different greenhouse gas have been identified in the output flows and were taken into account in the computation of the resulting emissions of CO_2e of the transport activities.

For the activity of the transport by plane, the radiative effect needed to be taken into account. This refers to the impact on the overall energy balance of the planet due to the release at high altitude of non-CO_2 warming pollutants, such as water vapour, aerosols and nitrogen oxides. Since those greenhouses gases have a higher Global Warming Potential than if they were released on the ground, the consensus in the literature to multiply the emissions by a weighing factor of 2[10],[11] was followed.

4.2.1 Emission Factors

Activity	Mean $\frac{\text{kg } CO_2e}{\text{km}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{km}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{km}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{km}}$	α	β
Bike	0.010	0.014	1.27E-5	0.052	0.517	2.02E-2
Train	0.046	0.009	0.030	0.064	27.17	1.69E-3
Bus	0.106	0.046	0.036	0.21	5.351	1.98E-2
Electric car	0.228	0.010	0.208	0.248	491.33	4.63E-4
Plane	0.334	0.071	0.209	0.488	21.92	1.53E-2
Essence car	0.382	0.016	0.350	0.416	518.764	7.37E-4

Table 4.8: Emission factors and statistics of transports

For each transport activity, we have a gamma distribution curve such as the one for the train activity displayed on Fig 4.3.

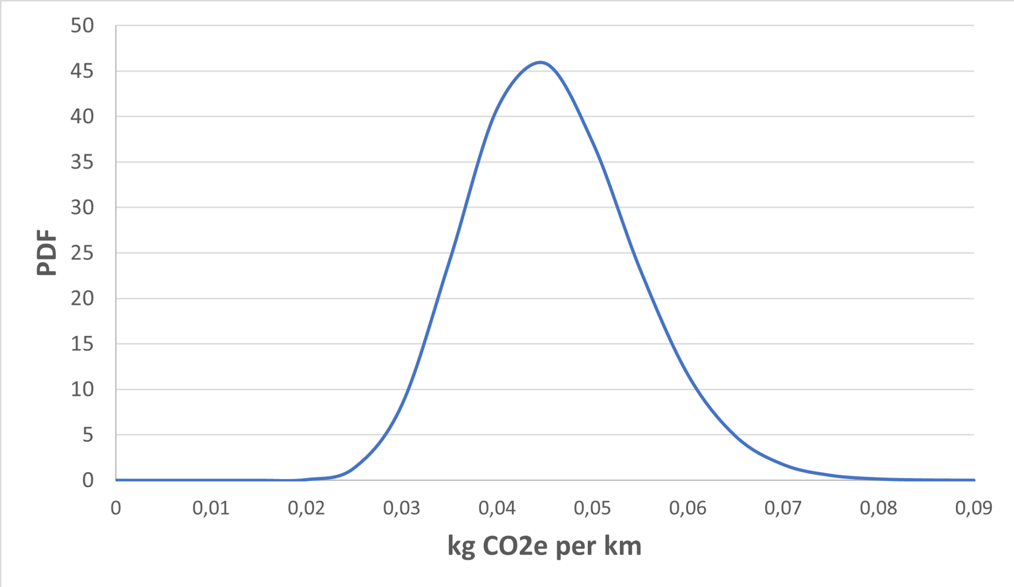


Figure 4.3: Gamma Distribution for the train activity with the parameters (α;β) set to (27,170;1,69E-3)

The 95% CI and the mean value can be displayed on the same curve such as it can be seen on Fig Fig 4.4 for the car activity.

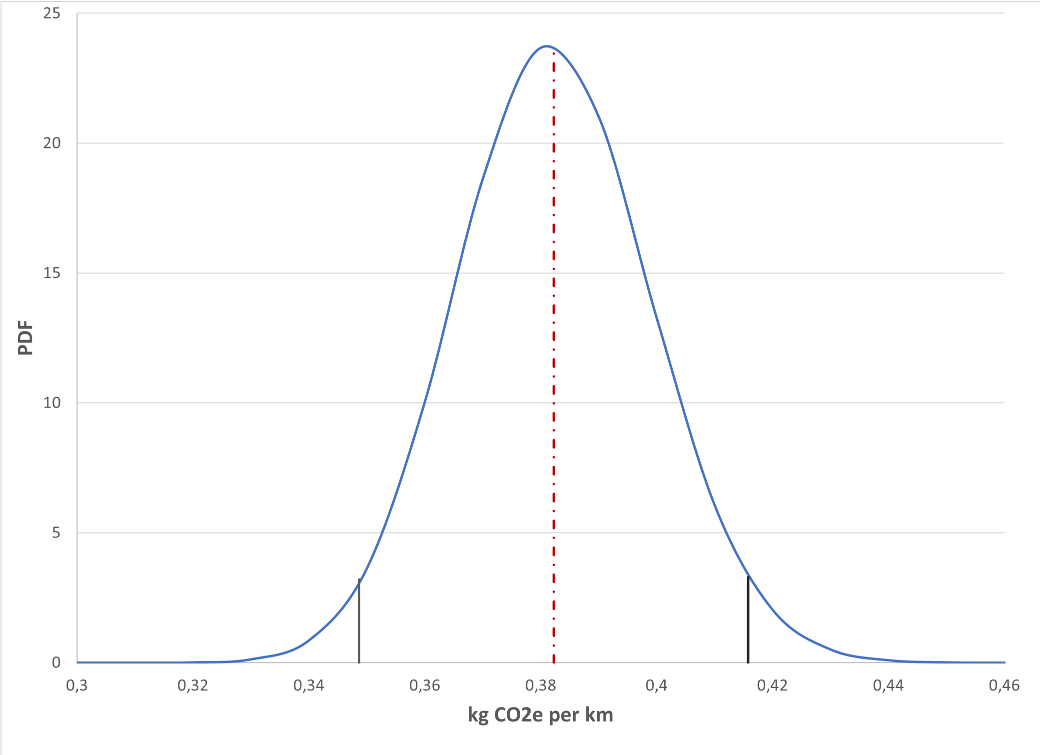


Figure 4.4: Gamma Distribution for the car activity with the parameters (α;β) set to (518,76;7,37E-4)

4.3 The residential heating sector

This sector is the smallest because it covers only a few activities as it is based on the activities in our household: electricity consumption and gas consumption, fuel or oil consumption for residential heating. The same database as the previous sector was used, ecoinvent v3. The amount of inputs/outputs were expressed in the database for all categories in terms of 1 megajoule (MJ). Since the factor must be expressed in terms of kilowatt-hour(kWh) for a better understanding of the user, the data were expressed by applying a multiplying factor of 3,6.

The computation for the emission factor has been optimized with regards to the transport sector. Collecting the information about the output mass flow of the emissions of each of the 17 GHG for one activity was a time consuming task as each value had to be input manually in our Excel file[5]. It was also observed that only four of the greenhouse gases had a real influence on the output in terms of kg

[5]This was due to the fact that SimaPro was remotely accessible on a virtual machine. One could thus not copy paste the values out of the virtual machine

CO_2e, with the CO_2 being responsible for more than 90% of the CO_2e emissions for all the activities except the transport by plane.

The decision has thus been made to only take into account the output emissions of the four following GHG: carbon dioxide (CO_2), methane (CH_4), nitrous oxide (NO_2) and Sulfur hexafluoride (SF_6).

4.3.1 Emission Factors

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kWh}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kWh}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kWh}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kWh}}$	α	β
Electricity	0.267	0.029	0.208	0.321	81.77	3.21E-3
Gas	0.255	0.033	0.196	0.327	58.46	4.44E-3
Oil	0.315	0.038	0.244	0.393	67.32	4.67E-3

Table 4.9: Emission factors and statistics of the residential sector

Chapter 5

CO_2 tracker application

In this chapter, we present a smartphone application that estimates CO_2e emissions based on the user's activities and purchases. A significant amount of work was actually to explore the possibilities we had and select the best one according to our requirements. This application operates on Google's Android operating system and the methodology adopted to compute the emission of an activity is to multiply its amount by a pre-computed emission factor.

5.1 Application structure

The tool was implemented in Kotlin, Google's new preferred language for Android app developers. This programming language is fully interoperable with Java but has a more concise syntax and permits faster compilation. We also used the Android Studio development environment which simplifies the development and testing of the application.

The application is divided in different pages that gives access to different functionalities. Starting it takes you to a login page where you have to input your email and password. After this, you are taken to the main page. This page displays the user's activities of the current month and lets you add an activity. Adding an activity allows you to choose an item from the selected category and specify an amount. This amount can be expressed in kilograms, kilometers or euros, respectively for the food, transport and home categories. The emission of the item is computed in the app when the user finalizes the operation. We multiply the amount by the emission factor of the item and then send the info to the database. We have also added a page to consult older activities as well as different graphs that will be discussed in details in Section **5.3 Charts**. More information on how to use the application is given in Section **5.6 User manual of the application**.

5.2 Online Database

To store the emission factors and the user's activities, we use a MongoDB database[1]. This database is classified as NoSQL. This kind of database notably differs from its counterpart, SQL, by its more simplistic data design. The data we store is not very complex but that will still facilitate our work. Another, and more important, factor that made us choose this database is the fact that they allow a free but limited access to their servers. This was not a problem, however, as the quotas set always were far from being met.

Although the online aspect of this database introduces a small delay to access the data, the benefits it provides are worth the inconvenience. The first advantage is that the user can keep his data when he changes or loses the device with which he accesses the application. The second one is that we can change the emission factors of items for every user more easily. This will be done sooner or later when discoveries are made about an item or about a process. We can also add an item without having to update the application.

There are two types of collections in the database. The first collection is called **Items** and contains info like the emission factor or category of a certain activity. One element of it can be seen in Table 5.1.

_id	name	category	emission_factor	alpha	lambda	user_ID
5ff[...]6f0	Onion	1	0.18	2.678	0.067	PUBLIC

Table 5.1: Sample of the **Items** collection

The `_id` field is required for the database to work. It uniquely identifies each document of the database. The category is either 1, 2 or 3 for the food, transport and home sector respectively. The emission factor is the amount of CO_2e the item releases. The `alpha` and `lambda` fields represent the Gamma distribution of the item's emissions but they are not used in our final version of the application. Finally, the user_ID is also required and specifies who can access the document. In this case, the token "PUBLIC" states that every one can have access to it.

_id	userID	itemname	category	emission	amount	date
602[...]f16	600[...]f46	Onion	1	0.36	2	2021-01-11T16:55:31

Table 5.2: Sample of the **Activities** collection

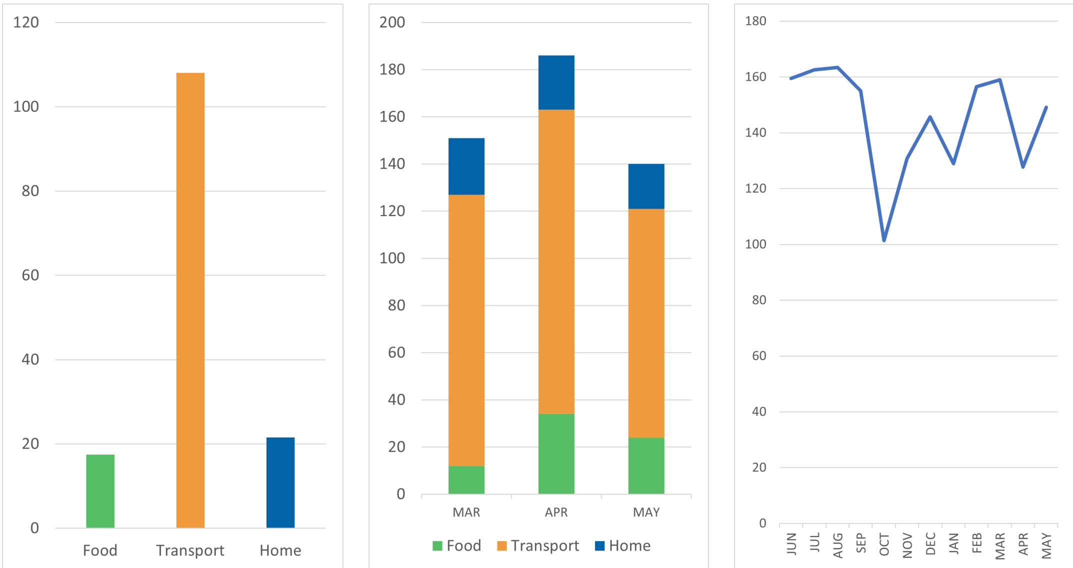
The second collection, of which one element is shown in Table 5.2, represents the activity or purchase of a certain user. Here, the `userID` is the emitter of

[1] `https://www.mongodb.com/`

the activity, identified by his ID which is created when the user registers in the application. The user info like his ID, email or password are automatically and entirely taken care of by MongoDB. Then, the `emission` field is the total amount of CO_2e emitted by the activity and `amount` contains its quantity. The last info, `date`, is the date and time at which the user added the activity.

5.3 Charts

We can build different charts to interpret the data of a user. We decided to put three that we thought were interesting in our application.



(a) Emissions of the current month by category

(b) Emissions of the three last months by category

(c) Emissions of the last year

Figure 5.1: Charts implemented in the application

The charts in the application were created with the MPAndroidChart library[2], they were reproduced to improve readability in Fig. 5.1. The first one shows a bar graph of the CO_2e emissions of the current month by category. The second one, a bar graph of the emissions of the last three months, also divided by category. The last one is a line graph displaying the total CO_2e emissions of the last year. All these charts are generated at the moment at which the user accesses the page by consulting the online database.

[2]`https://github.com/PhilJay/MPAndroidChart`

5.4 Public survey

A public survey was conducted in parallel with our early work on the application.[3] The aim of the survey was to collect the expectations of the citizens regarding the usage of our application. This is critical information with regards to our goal of increasing citizens' climate awareness. The survey was conducted in the form of a Google Forms which grouped the questions into three parts: identification of the audience, classic use of the application and specificities to be considered. It was shared publicly on the social network platforms LinkedIn[4] and Facebook[5] as well as privately through the same channels, by email and by word of mouth.

5.4.1 Audience identification

The survey was a success in terms of audience reaching as 307 people answered the survey. The relevant categories for the usage of a smartphone application were reached as 89.2% of the audience is in the age range of twenty years old to sixty years old. The audience demonstrated to have a global awareness of the climate change issue as only 0.8 % said they were not concerned by the issue of climate change.

5.4.2 Basic usage of the application

The first expectation from the audience was a daily use of the application that would be below ten minutes. When asked about the essential elements to be in the application for a regular usage of it, 89% answered they expected the application to be user-friendly and 62% are expecting that it is scientifically valid. The aesthetics of the application and its robustness against computer bugs were less of a priority.

Among the open questions we obtain as answers that the application needed to be engaging by sending notifications about the addition of data for example. Another request would be to propose via IA alternatives to polluting activities. Other ideas that were collected during the survey were:

- Having "favourites" activities (i.e. activities that users make on a daily basis) and pre-enregistered list to save time for data entry.
- Having a feature to project emissions over time in the future based on the past and current activities. This feature will enable to propose alternative

[3]All of the responses to the survey and the graphs derived from their analysis are available in the file attached to the thesis on dial.mem under the name Guillaume_73721400_vanderPlancke_70541300_2021_Appendix1.xlsx

[4]https://www.linkedin.com/

[5]https://www.facebook.com/

projections based on changes in the behavior.

- Having regular “smart” notifications that will pop up at the appropriate time. For example a notification that will occur at the usual time when you enter data at some point of the week (e.g. on Saturday morning after the grocery shopping). It will also pop up when it notices a movement of the user at a speed that will imply a transport activity rather than on foot.

Through the open questions, we also received advice from people who were sensitive and concerned about the possible drawbacks that could be unintentionally caused by certain aspects of our application. The remarks mainly concerned the respect of the use of private data. Other remarks were also collected in relation to the pressure people might put on themselves with regards to the climate impact of their behavior. It could lead to climate anxiety and other psychological distresses.

5.4.3 Considered features

The last part of the survey proposed different features for the application that we had thought about or that we identified during the state of the art. The audience was asked to rate from 1 (I am indifferent to this feature) to 5 (I find this feature essential) the following ideas:

- Offering a community service where the users can compare their results in terms of greenhouse gas emissions with those of their friends, as well as with the average of other users of the application. To help them in the comparison, they would have different tools at their disposal to make it as rich and relevant as possible.

- Having a partnership with other applications or services (e.g. Ryanair, SNCB/NMBS, Delhaize,...) to share data about the users’ activities.

- Creating a feature to help you reduce the greenhouse gas emissions of the users in the sectors where they would be above a certain threshold, such as the Belgian average for the respective sector example. It can be done by referring to articles containing tips and tricks on how to reduce your greenhouse gas emissions in these sectors.

- Providing an awareness section containing articles validated by the scientific community describing the various environmental impacts linked to excessive production of greenhouse gases and the resulting societal risks.

- Offering to the users ways of compensating their greenhouse gas emissions that they would consider excess, for example through the financing of sustainable projects to which the app would redirect.

The ideas that were the most popular in the answers of the survey were the second and third idea, with respectively 74% and 80% of the audience giving a note of 4 or 5. It confirms our intuition that the data entry is a major concern for the users. It was also a positive surprise to see that the audience is open for climate awareness and to have articles on how to reduce the impact of their actions. The first and last ideas were less popular. It denotes that the sense of community is not that much important, as much as the offsetting of the emotions.

Several lessons were learned from this survey, as expressed before. Nevertheless, they were not implemented in the application due to lack of time. They will be considered for further improvements to the application. They are enlisted in Section 5.7.

5.5 Database link library

Along with this application, we have implemented a small library to retrieve emission factors and add data to the database. This library has been designed to be a sort of link between our app and any other application by directly interacting with the database. Its purpose is to fulfill the partnership feature of the survey. For example, the Shell application, which records all your fuel purchases, could send the purchase information via this library. It will add activities to the user in the database, thus updating his data in our app. The receipt scanner we discuss about in Chapter 6 uses this library.

This library is implemented in Python and uses the PyMongo library[6] to communicate with the MongoDB database. It contains different functions:

- `get_item_category_and_factor(itemname)`: returns the category and the emission factor of the item with name `itemname`.
- `create_activity(userID, itemname, amount)`: creates and returns an activity object with format `{userID, itemname, category, emission, amount, date}`.
- `insert_activity(new_activity)`: inserts a single activity in the database. The argument `new_activity` must be generated with the `create_activity` function.
- `insert_activities(new_activities)`: inserts an array of activities in the database.

[6]https://pymongo.readthedocs.io/en/stable/

- `push_receipt_to(userID, elements)`: adds `elements` to the database of user `userID`. The parameter `elements` is an array and each of its element is of the format `{itemname, amount}`.

5.6 User manual of the application

The Android Package Kit (APK) of the application can be downloaded with this link[7]. To install it, you may need to authorize the installation of unknown apps on your phone. More info on this process can be found on internet but a link to one helpful source can be found here[8]

The first page of the application is the login page (Fig. 5.2a). If you don't have an account, you can create one by clicking on the `REGISTER` button, which will open up a dialog (Fig. 5.2b) where you have to input your email and password. After this process, the application will ask you to login (Fig. 5.2c) which will confirm your account.

You are then taken to the home page (Fig. 5.2d), containing the current month's activities as well as a button to add an activity. Logging in also gives you access to the navigation drawer (Fig. 5.2e), allowing you to navigate through the app and access the history page (Fig. 5.2f) and the charts as well as a way to log out of the application.

The addition of an activity dialog (Fig. 5.2g) allows you to select a category (Fig. 5.2h). Once this is done, the list of items will be refreshed (Fig. 5.2i). You can finalize the addition by entering the amount and clicking on the `ADD` button (Fig. 5.2j). The activities in the home page will then be updated to show the activity the user just added (Fig. 5.2k).

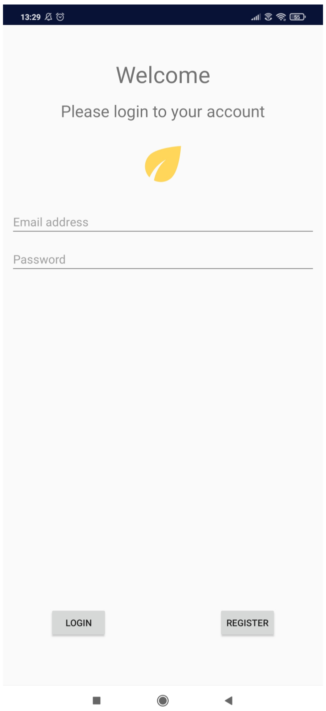
5.7 Our ideal application

How do we see the application in 5 years, if we could implement everything we want.

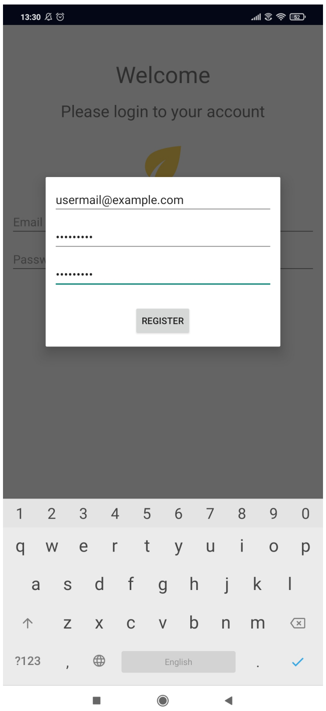
1. Proposing articles promoting the change of activities for the most polluting ones (for example promoting cycling for a regular car user).

2. AI system offering the least polluting transport or recipes based on the input of the clients or its recurrent activities: going from place A to place B, amount of calories needed per meal. For this we could implement it ourselves or have partnership with applications offering this service.

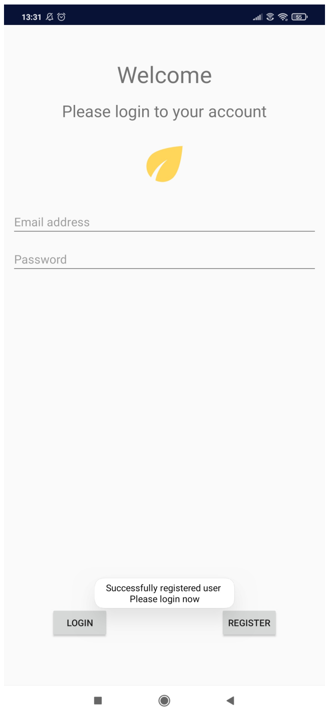
[7] `https://drive.google.com/file/d/11ttMOhQWiZRGxoN-A5Bs2LgmL48AqDdV/`
[8] `https://www.androidauthority.com/how-to-install-apks-31494/`



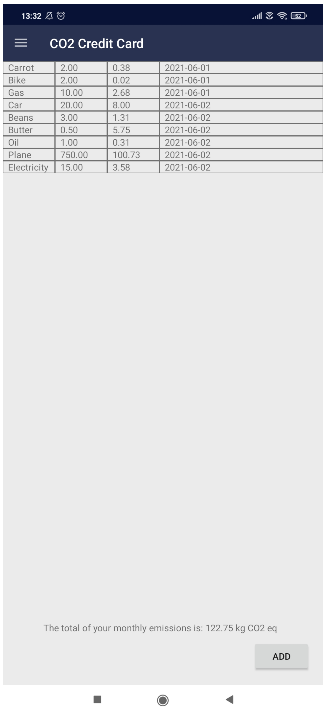
(a) Login page



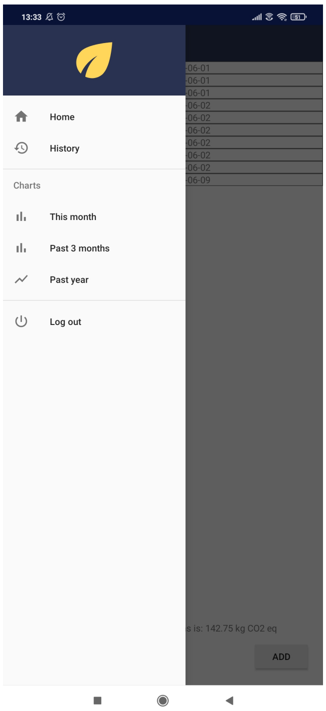
(b) Registration process



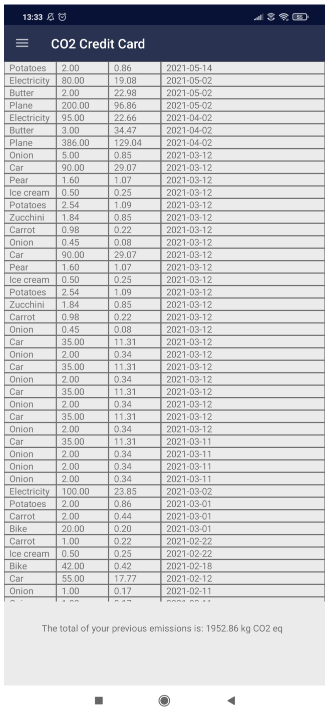
(c) Confirming registration



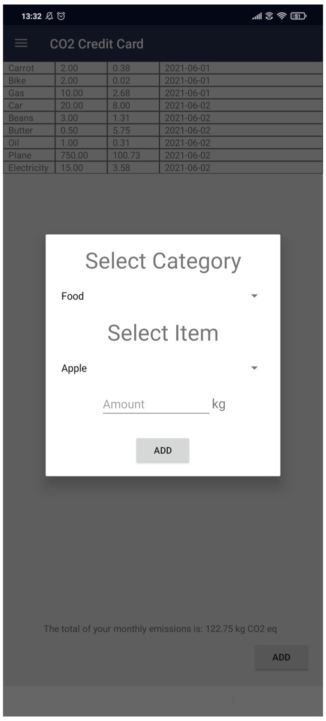
(d) Home page



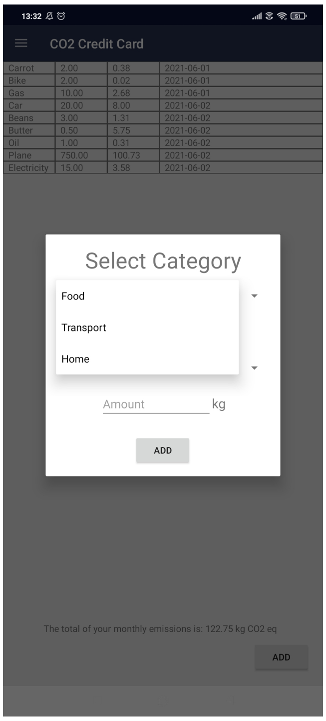
(e) Navigation drawer



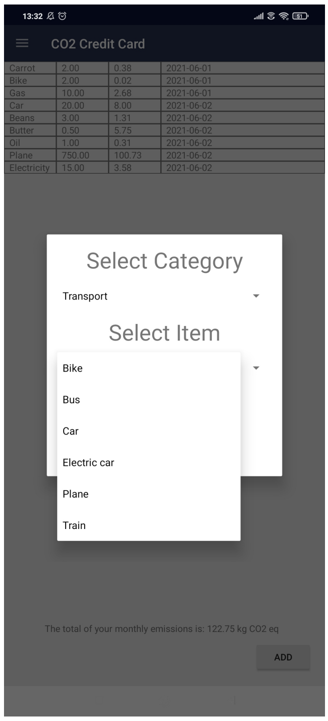
(f) History page



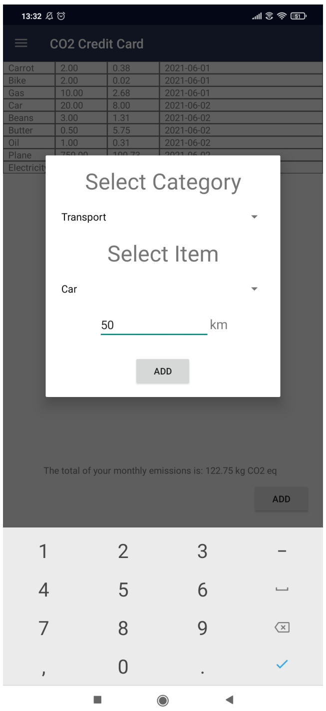
(g) Activity addition dialog



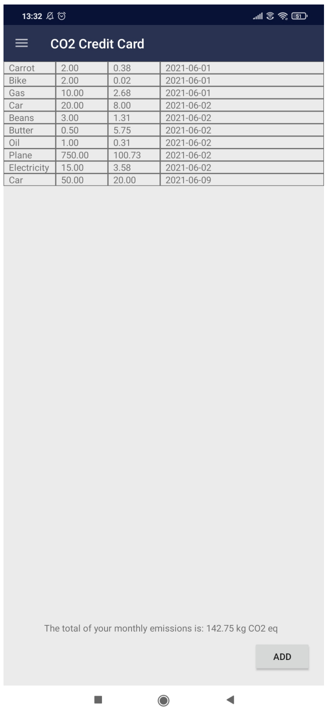
(h) Selection of category



(i) Selection of item



(j) Amount entry



(k) Update of the home page

Figure 5.2: Screenshots of the application

3. Adding more processed foods to the database (e.g. bakeries and pastries, fast food, processed meats, etc.) as they may have a significant impact on the user's emissions.

4. Notifications reminding the user to input data shortly after the application detects the end of an activity, such as a trip to the supermarket or a car ride for example. This must be done in agreement with the user of the application.

5. System of amortization of large costs in CO_2e over several months (e.g. purchase of a television, annual fuel oil delivery, annual water consumption bill, etc.).

6. Partnership with other applications (Ryanair, SNCB, Delhaize, Tesla) for a synchronized addition of data from those applications. The user doesn't have to enter the data anymore since those applications automatically compute it based on the actions of their customers, and the information is shared with us on the consent of the user.

7. Allowing the user to choose the date of the activity. At the moment, you can only add an activity at the current day.

8. In addition to an optimal ticket recognition system, we could also have partnerships with some franchises where they print a QR Code that once scanned would record all the data of the citizen's activity on our application.

9. Recognition system to reward citizens who are making an effort. We will insist that we should not give material or financial rewards because our response to climate change is not a contest or a game but a social duty. Nevertheless, any positive change in behavior must be acknowledged and rewarded with attention.

10. Since the tracking of the emissions is done individually, the entry of data can sometimes be complex for activities that are done simultaneously by a group of people (e.g. a shared meal, carpooling, residential heating of a shared flat, ...). It can be a nice feature if one user can add one activity and then includes other users on the same activity so that the emissions are shared accordingly, with a weight factor or not depending on each individual participation in the activity. This would be similar to what the application Tricount[9] is doing with regards to organizing financial group expenses.

11. Making the code of the application open source so that it can quickly be declined in other countries. It has many benefits. The first is that it will

[9] https://www.tricount.com/

allow to have more people contributing to the development of the application which will accelerate it towards the goal of making it more user friendly and less time consuming. The second benefit, and not the least, will be to expand and share our work for raising citizens' awareness on climate change to other academic spheres or to volunteers involved in the climate change.

It is beneficial since the problem of global warming is a worldwide issue and will require to be tackle globally. Sharing our work will contribute to creating synergies with regards to this fight, as they would be able to benefit from our application and we will learn in return about their approach to this problem.

People willing to decline our application will only face one constraint. It will be to adapt the database of emission factors for the activities in accordance to the reality of their country or region. A change of localisation involves change of methods of productions and thus variation in the resulting emissions of GHG by activity produced. A good example is the electricity mix of a country, which can greatly change from country to country.

Chapter 6

Receipt Scanner

This chapter introduces another tool that we have developed and whose purpose is to read a receipt in order to extract its content. The receipts we have selected to experiment with are from the Delhaize supermarket. The tool is implemented in Python and uses Google's Optical Character Recognition (OCR) engine to convert the image into machine-encoded text. This text is then parsed and formatted with the aim of adding the items to the CO_2 tracker application via the library introduced in Section **5.5 Database link library**. This has the potential to greatly reduce the time it takes for a user to input data, thus improve the ease of use and user-friendliness of the application.

6.1 Image pre-processing

First and foremost, we need to do some pre-processing of the image. Data pre-processing is a crucial step in any machine learning project. The idea is to simplify and format the image to prepare it for being given to the model. Basically, the purpose of this step is to facilitate the task of the model by emphasizing the most important parts of our data. We will use functions of the OpenCV[1] and scikit-image[2] libraries to accomplish our goal.

The first task in this operation is to distinguish the piece of paper in the image. We start by using a process called Canny Edge Detection whose purpose is pretty self explanatory; it detects edges in the image (see Fig. 6.1b). Among the edges, we then find the biggest one that forms a four-sided polygon. We have found our receipt in the image (Fig. 6.1c).

We now "flatten" the receipt and correct the angle of the photo with the function `warpPerspective` (Fig. 6.1d). The final step in our pre-processing procedure is to

[1] `https://pypi.org/project/opencv-python/`
[2] `https://scikit-image.org/`

(a) Original image

(b) Detection of edges

(c) Receipt contouring

(d) Skewness correction

(e) Thresholding

Figure 6.1: Image pre-processing of the receipt

apply a threshold mask to the image (Fig. 6.1e). Pixels under a certain value will be considered white and the others, black. This threshold also takes into consideration an area around the pixel to make it more robust to shades and folds on the receipt.

6.2 Optical Character Recognition

For this part, we used a wrapper of Google's OCR engine[3] to use it in Python. Trying to build our own model would have taken a lot of data and time and it was not the focus of this study. Therefore, the only thing we need to do is to feed our pre-processed image to the function `image_to_string`. We will also give the parameter `lang='fra'` to detect french words because its the language in which the receipts we use are written. This function returns a string containing everything that the engine was able to read. We can then move to the next step of our scanner.

6.3 Pattern matching

The text extracted in the previous part must now be analyzed to find the information that interests us. For each line, we check if it contains a price. This is done with the re module[4] that gives access to regular expression operations. A regular expression, or regex, provides rules for a certain pattern to match strings or parts of strings. In this case, the regex to match prices is the following: `([0-9]+[\.,:]1[0-9]2$)`. It matches any real number with 2 decimals at the end of a line.

If the line matches the price pattern we then scan it again to find if it contains an item in the database. In order to do this, we check each sequence of words of the line against the strings in an array that contains every name of items in the database. The sequences of words contain up to three words because the name of the items can be composed of three words. The check is done with the function `SequenceMatcher` of the difflib library[5], which gives a score of the similitude between two strings. The purpose of this is to recover from errors made by the OCR part. With the info detected, we then build an array of item names and prices

Along with this array, we also construct an object that contains the mean price per unit (expressed in euros per kilo or euros per liter) of items. This object is updated each time we encounter an entry in a receipt that also specifies an amount, which is found with the regex `(([0-9]+[\.,])?[0-9]+((KG)|([MC]L)|[KGMCL]))`.

[3] https://pypi.org/project/pytesseract/
[4] https://docs.python.org/3/library/re.html
[5] https://docs.python.org/3/library/difflib.html

We can then use this object to assemble the final receipt that contains the name and the amount of the items purchased, which can directly be added to the database of the application using the library of Section 5.5.

6.4 Improvements

Throughout our development process, we tested the scanner with 4 different receipts with a total of 35 items. The first prototype didn't have the best results because the OCR was doing a lot of mistakes. At this point, the tool had an accuracy of 57.14%. Improvements were made by fine tuning the parameters of image pre-processing, particularly the size of the neighbourhood considered in the thresholding.

We could then focus on the pattern matching and improve its robustness. A first process we introduced is the similarity check explained in the previous section. It can help and recover from the errors made in the OCR. For example, the word "`speculoos`", interpreted by the OCR as "`speculo0s`", is correctly classified because they have a similarity of 88.88%. Combining the better image pre-processing and the similarity check allowed us to increase the accuracy of our scanner up to 82.86%.

Other minor improvements were made on particular cases, such as the one shown in Fig. 6.2. Here the price is on the line after the name of the item. This kind of problem was easily fixable in our code as it is just a new rule to add to our matching.

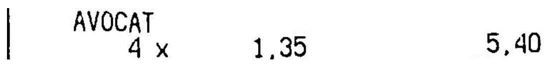


Figure 6.2: A particular case in a receipt

The tool was finally tested with the four receipts and we managed to reach an accuracy of 94.29%. The root of these errors is the OCR and we were not able to recover from them afterwards. Despite the tool being developed for Delhaize receipts, it also works with receipts that respect the same pattern.

Chapter 7

Case Study

The case study includes all the different parts of this work. It is divided into three sections. The first one is a follow-up of the activities of several users who used our application over the space of one week. The second part consists of a post processing analysis where we will take into account the same activities recorded by the users and the uncertainties obtained in the methodology part. The impact of taking into account uncertainties in the calculation of data will be studied.

The first aim of the case study is to assess if the application is returning the appropriate mean values for the registered activities. The second goal of the case study is to assess, via the post processing analysis, the impact of the uncertainties on the distribution of the mean values returned for the corresponding emission factors.

The proof of concept will be studied in the third section. It will base itself on the relevance of the emissions data returned by our tool to the actual data provided by the federal environmental agency.

7.1 Users activity for one week

Once the application was fully developed, it was shared within a close circle of people. The analysis period is one week and the participants are being asked to enter their daily activities. As the case study takes place over the space of one week, we had to adapt for the entry of activities with regards to the residential heating sector which are often based on monthly or annual invoices. Participants were thus asked to smooth their monthly or yearly input to obtain the corresponding value for a week. There was no issue for the agriculture and transport sectors.

Among the participants who took part of the case study, we have deliberately chosen to focus on two profiles according to their behavior. The first profile case is the one of Thibault[1] He admits to be slightly concerned about his environmental footprint and thus takes little account of it in his daily actions. On the other hand, there is Marie who shares a greater interest and involvement with regards to this issue. We deliberately took opposite profiles with regards to the post processing analysis. The idea is to see if the impact of uncertainties can vary according to people's different lifestyles as it involves the use of various activities.

The participants were asked to individualise the emissions that resulted from group activities. Thibault for example entered his own food consumption despite making groceries shopping for the whole family. It is important to do so because the application tracks on an individual level the emissions, and thus shouldn't take into account the one's resulting from his children food consumption. Emissions resulting from his journey to drop off the children were also divided by the number of people in the car. Thibault was asked to do the same with regards for the residential heating. The amount of CO_2e was divided by five with regards to his spouse and their three children. Marie was not confronted for the transport and residential heating sectors as she lives alone and moves by bike or public transport. She took this criterion into account when eating with her friends at the weekend.

Follow-up of Thibault's activities

Thibault follows a varied diet with few restrictions. During the weekend he likes to enjoy a piece of meat on Saturday. His work is located at Antwerp, at 55,6 kilometers for a one way trip from his house. He spent the three first days using his car to reach his office but hasn't colleagues who lives near him to do carpooling. Out of a desire to reduce his carbon footprint, he then switched to using the train for the last two days of the week. He used his car again during the weekend to drop his children at their respective occupations. Thibault uses an oil-fired boiler which is the most environmental harmful way of heating.

The next table highlights the share of Thibault's emissions for each sector over the course one week.

Agriculture	Transport	Residential heating	Total
59.09	171.62	39.65	270.36

Table 7.1: Thibault's CO_2e emissions by sector over the course of one week

[1]Names are fictitious in order to respect the anonymity of individuals.

The sector of transport has significant more importance than the other sectors. It covers 63.5% of the share of the emissions, leaving 21.8% to agriculture and 14.7 % to the heating sector. It is due to the numerous kilometers done by Thibault with his car, despite having used the train on Thursday and Friday to go to the office.

Follow-up of Marie's activities

Marie follows a vegetarian diet. She consumes dairy products in moderation. She rides her bike to work on a short distance of four kilometers one way. Her second-hand bike does not wear out quickly as she has been using it over short distance for 2 years. At the weekend she took the train to the sea with some friends. During this trip, they took the bus together to visit Bruges on Sunday, making a total round trip of 56 kilometres. Wednesday evening she also visited her friend in Leuven by taking the train. The round trip distance between her home and Leuven is 72 kilometres.

The emissions of Marie's activities by sector are provided on the next table.

Agriculture	Transport	Residential heating	Total
40.45	21.58	14.52	76.95

Table 7.2: Marie's CO_2e emissions by sector over the course of one week

The comparison is obvious in how the transport sector has a lighter influence than in Thibault's case thanks to Marie's efforts for using less polluting means of transport. The sector of transport covers 21.6% of the share of the emissions while the agriculture sectors covers 40.4%. It leaves 14.5 % to the heating sector.

7.2 Processing the users activity

For a possible future version of the application, it was thought that the values of the emission factors could vary randomly according to their distribution.

The aim of this approach is to take into account the uncertainties present in the measurement of these values. The users of the application would be informed of this. This would educate them about the measurement of the data which is complex and therefore not one hundred percent accurate. Trust is expected to be gain from the users thanks to this approach. A greater regularity in the use of the application is expected as a consequence. The purpose of this section is to check whether this approach would be consistent or not.

The different activities of Thibault and Marie and the data related to it (i.e. the identity of the activity, the quantity of the activity and the date it was performed) were retrieved from our online database within a Excel file for the post processing analysis. The file contains the emissions factors retrieved from the databases presented in Chapter 4 for each activity. The computed parameters for the gamma distribution and the lower and upper boundaries are also added.

The modus operandi conducted was the following:

1. Each activity got assigned randomly ten different emission factors based on their respective Gamma Distribution. For doing this, the Excel function `GAMMA.INV(probability, alpha, beta)` presented in Subsection 4.1.2 is used. The alpha and beta parameters are the ones computed for the activity. The randomness in the return of values is introduced by the Excel function `RAND()` which returns an evenly distributed random real number greater than or equal to 0 and less than 1. The Excel function `GAMMA.INV(probability, alpha, beta)` will thus return the cumulative distributive function of the gamma distribution for the random variable "probability".

2. Since `RAND()` can return values within the range of [0;0.025] and [0.975;1.00[, the function `GAMMA.INV(probability, alpha, beta)` might values outside of the 95% CI. Despite finding a function that can do this automatically, this was checked manually. If one value was found to be outside of the 95% CI, a new value was assigned.

3. The set of ten values for the emission factors are multiplied by the quantity of the activity to obtain the resulting mass of CO_2e emitted.

4. The set of ten values of the different activities are then summed up with one of the random value of every other activity of the day . Ten different values of the total amount of CO2 emitted for the day are thus obtained.

5. The variation of those values are displayed on box plots. As it provides a visual summary of the variation of data throughout the week, it enables also to quickly identify mean values and the dispersion of the data set.

6. In parallel of the plot, a set of ten total mean value is derived for the week by adding the daily total CO_2e emissions of one day together. We can obtain the average mean value for the week by dividing by ten. This will be compared with the value provided by the application to assess if whether or not the data are consistent.

Analysis of Thibault's activities

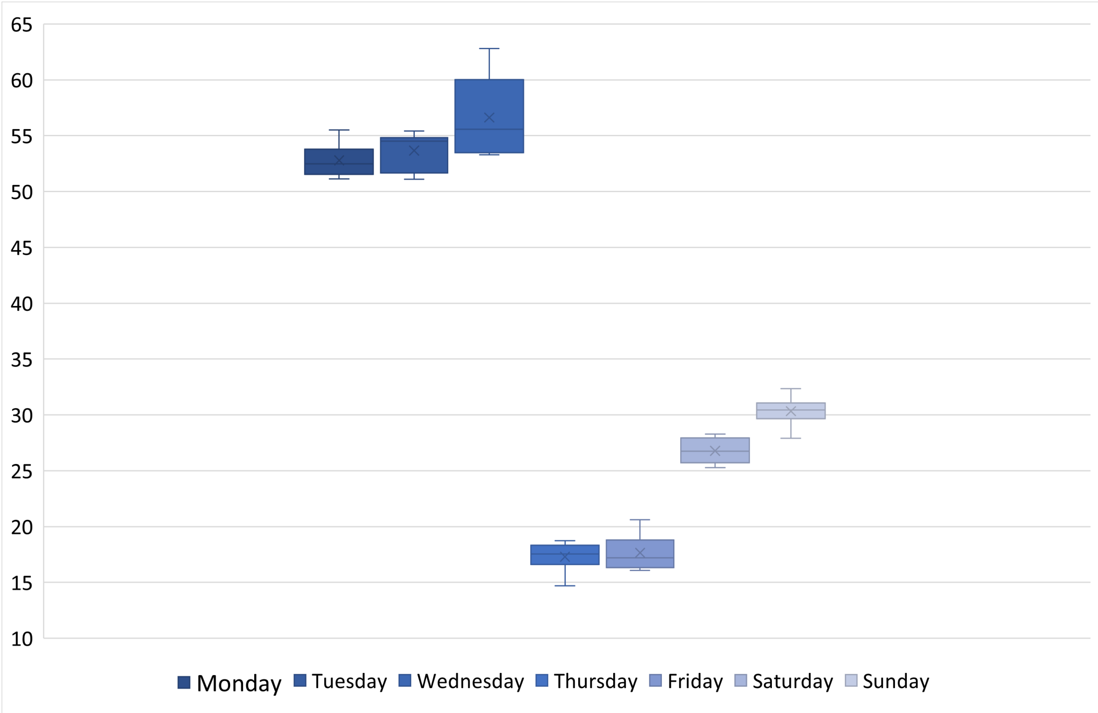


Figure 7.1: CO_2e emissions of Thibault's activities by day over the course of one week

Several lessons can be taken from the previous box plots (Fig. 7.1). First of all, one should notice the wide variation in terms of CO_2e emission from Wednesday to Thursday. This coincides with Thibault's change of behaviour from car to train.

It shows first the huge difference between transport by car and transport by train for an almost equivalent distance (112 kilometres and 108 kilometres respectively). The difference can be estimated at approximately 36 kg CO_2e emitted daily in the case of the car. Over the course of 220 days of work throughout the year, this daily difference can be evaluated to 7920 kilograms of CO_2e of difference[2]. The variation between Friday and the days of the weekend are again due to the car activities of Thibault who drops his children off at their activities. Those variations highlight the high polluting activity of the car and the importance of homeworking or carpooling.

In a second phase, it displays the huge difference between a car ride activity and the other food activities. Indeed, the food activities don't display a huge difference in emissions despite the changes in the food diet of Thibault.

[2]For the case of someone living at 55 kilometers of his office and going there everyday by car.

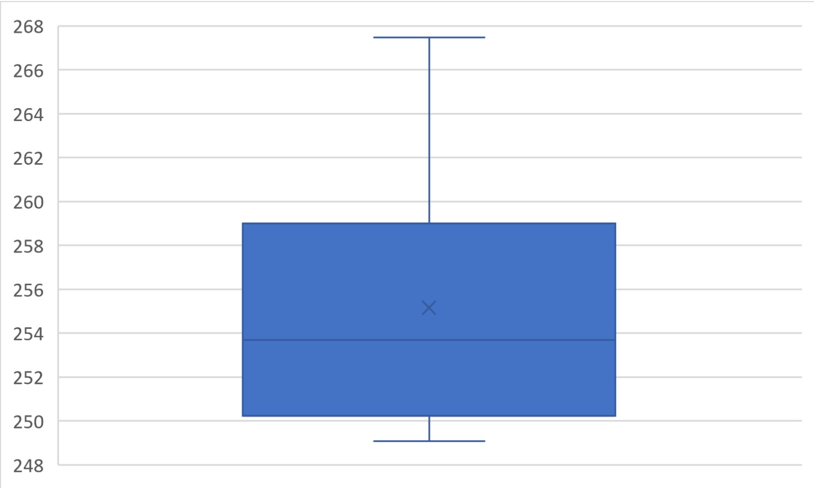


Figure 7.2: Total CO_2e emissions of Thibault's activities over the course of one week

Regarding the total of Thibault's activities on one week, the set of ten values are located between 249,7 and 258,8 kg CO_2e emitted. The resulting mean value of the post process analysis is 255,15 kg of CO_2e emitted and is represented by the cross on Fig 7.2. As displayed by the Table 7.1, the mean value returned by the app was equal to 270,36 kg of CO_2e. There is a difference of 15,21 kg of CO_2e, that is a deviation of 5,98%. In other words, the mean value of CO_2e emissions computed by taking uncertainties into account has a deviation of 5,98% from the mean value of CO_2e emissions computed without taking into account uncertainties. The deviation decreases as the number of random coefficients (e.g. equal to 10 for this case study) increase.

Analysis of Marie's activities

The plot of Marie's activities displays different results as expected, in line with the difference in their behavior.

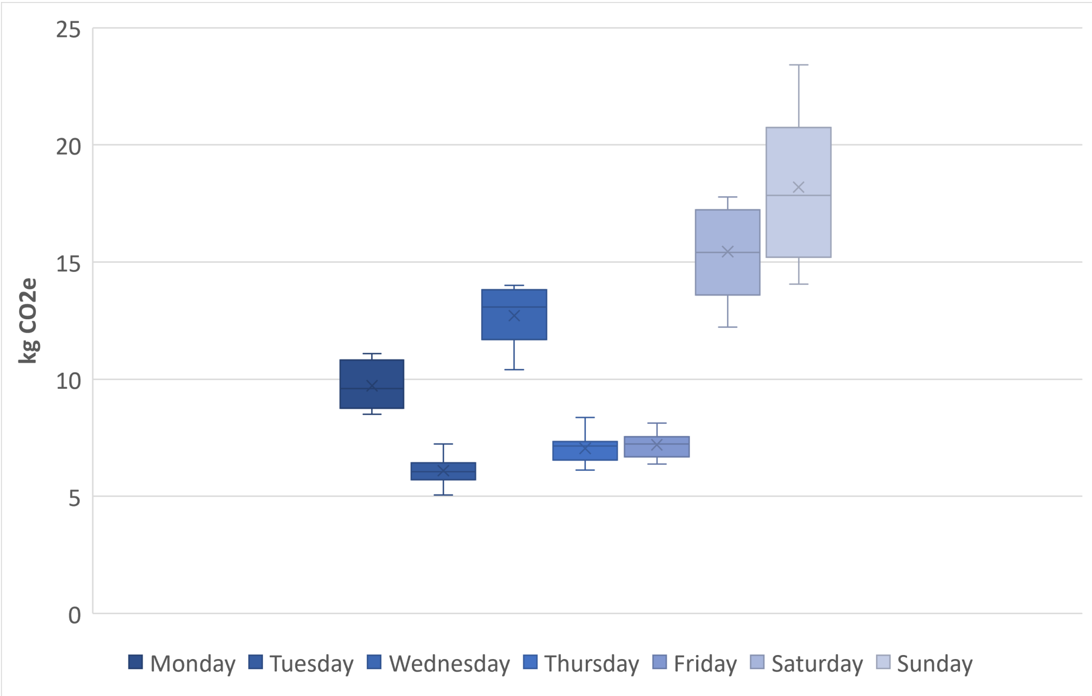


Figure 7.3: CO_2e emissions of Marie's activities by day over the course of one week

Several lessons can be again taken from Figure 7.3. First, the daily emissions of Marie in terms of CO_2e are significantly lower than Thibault's daily emissions due to their difference in behavior. The second and main lesson we can draw from this graph is the width of the data set of Sunday. The values are located between 14,06 and 23,41 kg CO_2e emitted while the mean value is 18,19 kg CO_2e. For the lower and upper values it represents respectively a deviation of 23% and 28% from the mean value. This is due to the significant uncertainties around the bus activity. As the bus trip was significant (i.e. 56 km), the activity has become a major influence on the uncertainties of the day. The uncertainties with regards to the emissions around one activity thus increases with the quantity of the same activity.

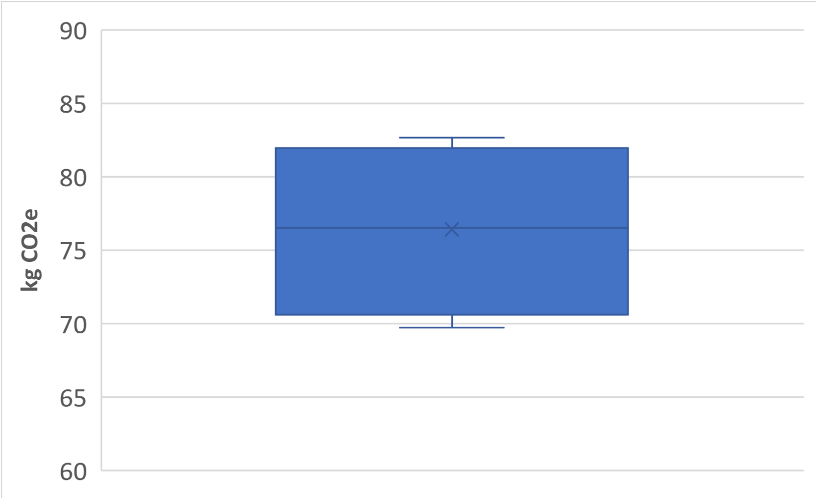


Figure 7.4: Total CO_2e emissions of Marie's activities by day over the course of one week

Regarding the total of Marie's activities on one week, the set of ten values are located between 69,74 and 82,68 kg CO_2e emitted. The resulting mean value of the post process analysis is 76,42 kg of CO_2e emitted and is represented by the cross on Fig 7.4. As displayed by the Table 7.2, the mean value returned by the app was equal to 76,95 kg of CO_2e. The difference is equal to 0,53 kg of CO_2e (i.e. a deviation of 0,6% from the mean value computed without taking uncertainties into account). This is quite an astonishing result but one shouldn't be mistaken. This result was obtained by luck as the next iterations of random values provided deviation equal to 3,4% and 4,2%.

It will be up to the people working on the next steps of the development of the application to see if they consider these deviation factors to be sufficiently small low to consider coherent the data computed and returned taking into account uncertainties. If that assumption is made, it directly implies that the values retrieved from the respective databases are accurate enough and thus one shouldn't be searching for new databases. If they consider that there is still too much variation in the data (i.e. the deviation is too high), they will have to work on the refinement of the highest uncertainties. Activities with the higher uncertainties around as it have been shown with the case of Marie and the transport by bus to have the most influence once it is used by a large amount.

7.3 Proof of concept of the application

To confirm the concept of the application we have to demonstrate consistency between the data returned by the application and reality. For this, the results obtained for Thibault's and Marie's activities need first to be aligned with the values disclosed earlier in the report at Section 3.3 in terms of the share of the sectors in the total emissions. From the Fig 3.2, we can retrieve the following share of the sectors: 48.3% for the sector of transport, 21.9% for the sector of agriculture and 29.8% for the sector of residential heating. This was done by dividing the individual percentage of each sector by the total percentage of the three sectors together, i.e. 46,2%.

The share of the sectors for Thibault's activities were 63.5%, 21.8% and 14.7% respectively for the sectors of transport, agriculture and residential heating. The share of the sectors for Marie's activities were 21.6%, 40.4% and 14.5% respectively for the sectors of transport, agriculture and residential heating. The values of the share of sectors from Thibault and Marie do not fully coincide with those attributed to the average Belgian.

One might also check the proof of concept by confronting the emissions in terms of in terms of magnitude. The CO_2e emissions of Belgium per capita in 2019 was evaluated at 10,6 tonnes[12]. As seen in the Section 3.2, the three sectors account for 46.2% of the total of the yearly GHG emissions of one Belgian. It represents a value of 4,90 tonnes with respect to the value of 10,6 tonnes.

The case of Thibault will be first considered. His total weekly emission are multiplied by the number of weeks in a year (e.g. 52), we obtain 14,06 tonnes of CO_{2e} for the sectors of agriculture, transport and residential heating. The yearly emissions of Thibault emitted from his activities are thus 2,86 times higher.

For the case of Marie, her total of emissions of CO_2e over one year is equal to 4,01 tonnes. This is significantly closer to the theoretical value. Nevertheless, Thibault's value is still too much spread for this to be considered as sufficient proof.

One could point out that Thibault's and Marie's behavior and historic of activities are not sufficient on their own for a proof of concept. This proves the limits of our master thesis in terms of resources to provide a valuable proof of concept. If one wants to do so, he should consider a larger sample gathering enough participants and data so that all the different types of behavior of Belgians can assumed to be represented. This represents a significant amount of work and might be one of the next step of the further development of this research topic.

Chapter 8

Conclusion

In a context where academics, industries and politics are committed to the fight against global warming, the subject of our master thesis could hardly be more topical.

The aim of the master thesis was to provide a tool that could track more than 80% of the greenhouse gas emissions of Belgian's activities. The intention behind this objective was to allow Belgian citizens to be aware of their individual carbon footprint so that they can make more informed choices and to change their living behavior accordingly if they wish.

A methodology was first put in place to provide a path and a framework for our research with regards to the complex task of the assessment of the greenhouse gas emissions resulting from anthropogenic activities.

On its basis, we were able to identify three sectors of activity accounting for more than 88% of the direct emissions resulting from Belgian's activities. The common unit of measurement CO_2e, a reference method to assess environmental impacts (LCA) and search criteria for databases were also defined.

Last but not least, uncertainties were considered to be taken into account in order to be close with the reality of the wide variation existing in the literature around the emissions factor of human activities. The uncertainties were chosen to be represented by gamma distributions as it is consistent with reality by not providing negative values.

Two databases were validated on the basis of the methodology and therefore retained. The first one is provided by a systematic review of more than 369 published LCA studies for the agriculture sector. For the sectors of transport and residential heating, the database ecoinvent v3 was chosen. Methods for calculating uncertainties were designed for each database accordingly .

In total, the emission factors for 47 activities from the most commonly made by humans were retrieved as well as the values of their alpha and beta parameters of their gamma distribution.

A tool in the form of a smartphone application was in parallel designed to provide citizens access to the collected data. The expectations towards the usage of our application were identified through a public survey reaching an audience of 307 people. We took them into account by making the application user friendly by working on the technology of a receipt scanner. A second popular feature in the survey has been developed enabling the possibility of future exchanges of information between our application and other ones to automate data recording.

On a short term, we can think of many ways to improve this work. Room of improvements regarding the methodological side would be to collect data from a significant large sample of users to obtain a valuable proof of concept as identified in the case study. As for the improvements for the programming part, it would be to make the application accessible on the other platforms, i.e. iOS and Windows.
On a more long term, a significant number of advanced and idealistic lines of work are available, at is has been disclosed in Section 5.7.

With an eye to the future, it will be interesting to see if synergies are created with other initiatives during the further stages of the research, in order to best contribute to the lowering of our individual carbon footprint and tackling climate change.

Appendix

A.1 GWP values for the greenhouse gas

The following values are retrieved from the Fourth Assessment Report[3] of the IPCC. It includes the 100-year time horizon global warming potentials of GHG.

Greenhouse gas	GWP $\frac{\text{kg } CO_2e}{\text{kg GHG}}$
CO_2	1
CH_4	25
NO_2	298
NF_3	17200
SF_6	22800
$CFC-11$	4750
$CFC-12$	10900
$CFC-113$	14400
$CFC-114$	10000
$HFC-23$	14800
$HFC-134a$	1430
$HFC-152a$	124
$HCFC-21$	148
$HCFC-22$	1810
$HCFC-124$	609
Halon 1211	1890
Halon 1301	7140

A.2 Reference of the activities used from the ecoinvent v3 database

The library used is the following: "Ecoinvent 3 - allocation at point of substitution - unit".

For the transport sector:

Activity	Reference name in ecoinvent	Process identifier
Bike	Transport, passenger, bicycle{RoW}	EI3ADUNI000061667507691
Train	Transport, passenger train{BE}	EI3ADUNI00006166750899
Bus	Transport, regular bus{RoW}	EI3ADUNI000061667507661
Electric car	Transport, passenger car, electric{GLO}	EI3ADUNI000061667507735
Plane	Transport, passenger, aircraft{RER} | intercontinental	EI3ADUNI000061667507488
Essence car	Transport, passenger car, medium size, petrol, EURO 4 {RER}	EI3ADUNI000061667507457

The acronyms GLO, RoW, RER and BE respectively state that the inventory is modelled for the markets of the global word, the rest of the world outside China, the region of Europe and Belgium.

To access it via the software SimpaPro:

Activity	Path to follow
Bike	Transport -> Road -> Transformation
Train	Transport -> Rail -> Transformation
Bus	Transport -> Road -> Transformation
Electric car	Transport -> Road -> Transformation
Plane	Transport -> Air -> Transformation
Essence car	Transport -> Road -> Transformation

For the residential heating sector:

Activity	Reference name in ecoinvent	Process identifier
Gas	Heat, central or small-scale, natural gas{Europe without SWIZ}| heat production, natural gas, at boiler atm, low-NOx condensing, non-modulating <100kW	EI3ADUNI000061667506378
Electricity	Electricity, low voltage{BE} electricity voltage transformation from medium to low voltage	EI3ADUNI000061667506071
Oil	Heat, central or small-scale, other than natural gas{RoW}| heat production, light fuel oil, at boiler 100kW condensing, non-modulating	EI3ADUNI000061667505920

To access it via the software SimpaPro:

Activity	Path to follow
Gas	Energy -> Heat -> Gas -> Transformation
Electricity	Energy -> Electricity country mix -> Low voltage -> Market
Oil	Energy -> Heat -> Oil -> Transformation

A.3 Emission factors

For the agriculture sector:

Vegetables and fruits

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Onions	0.18	0.11	0.033	0.451	2.678	0.067
Potatoes	0.20	0.08	0.075	0.384	6.25	0.032
Carrots	0.22	0.15	0.029	0.596	2.15	0.102
Mushrooms	0.27	0.29	0.004	1.060	4.251	0.626
Cucumbers	0.33	0.32	0.010	1.19	1.063	0.31
Zucchinis	0.42	0.50	0.003	1.806	0.706	0.595
Beans	0.62	0.45	0.069	1.76	1.90	0.327
Broccolis	0.70	0.34	0.200	1.51	4.239	0.165
Peanuts	0.87	0.11	0.668	1.098	62.55	0.014
Lentils	1.10	0.45	0.403	2.14	5.975	0.184
Eggplants	1.35	0.07	1.216	1.453	371.939	3.62E-3
Pears	0.33	0.13	0.126	0.629	6.443	0.051
Oranges	0.35	0.12	0.156	0.621	8.507	0.041
Apples	0.36	0.19	0.089	0.816	3.590	0.100
Tomatoes	0.46	0.18	0.178	0.874	6.531	0.070
Cherries	0.48	0.40	0.032	1.52	1.44	0.333
Peaches	0.54	0.24	0.177	1.102	5.063	0.107
Strawberries	0.65	0.36	0.146	1.522	3.260	0.199
Melons	0.88	0.01	0.86	0.900	7744	1.13E-4

Plant and dairy products

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Rice	2.66	1.29	0.763	5.721	4.25	0.626
Soy milk*	0.88	0.27	0.433	1.482	10.62	0.083
Milk*	1.32	0.29	0.814	1.950	20.72	6.37E-2
Yoghurt*	1.43	0.25	0.983	1.960	32.718	0.044
Eggs	3.39	1.21	1.449	6.141	7.849	0.432
Cheese	8.86	2.07	5.278	13.354	18.32	0.484
Butter	11.52	7.37	1.859	29.810	2.443	4.715

* For those products, the values are expressed in terms of $\frac{\text{kg } CO_2e}{\text{liter food}}$.

Fish

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Pilchard	1.10	0.45	0.403	2.141	5.975	0.184
Herring	1.17	0.17	0.861	1.526	47.37	0.025
Mackerel	2.00	1.08	0.473	4.606	3.429	0.583
Tuna	2.60	1.45	0.577	6.117	3.215	0.809
Salmon	3.76	1.47	1.454	7.142	6.542	0.575
Mussels	7.54	4.93	1.148	19.821	2.339	3.223
Shrimp	14.85	12.37	0.981	47.027	1.441	10.304
Lobster	21.74	11.70	5.180	49.95	3.453	6.297

Livestock

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kg food}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kg food}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kg food}}$	α	β
Chicken	4.12	1.72	1.469	8.112	5.737	0.718
Pork EU	5.60	1.51	1.469	8.112	3.04	8.92
Beef EU	26.05	6.78	14.50	40.92	14.76	1.7646
Lamb EU	33.84	13.06	13.29	63.83	5.737	0.718

For the transport sector:

Activity	Mean $\frac{\text{kg } CO_2e}{\text{km}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{km}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{km}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{km}}$	α	β
Car essence	0.382	0.016	0.350	0.416	518.764	7.37E-4
Car electric	0.2276	0.010	0.208	0.248	491.33	4.63E-4
Bus	0.106	0.046	0.036	0.21	5.351	1.98E-2
Train	0.046	0.009	0.030	0.064	27.17	1.69E-3
Bike	0.010	0.014	1.27E-5	0.052	0.517	2.02E-2
Plane	0.334	0.071	0.209	0.488	21.92	1.53E-2

For the residential heating sector:

Activity	Mean $\frac{\text{kg } CO_2e}{\text{kWh}}$	Standard deviation $\frac{\text{kg } CO_2e}{\text{kWh}}$	LB 95% $\frac{\text{kg } CO_2e}{\text{kWh}}$	UB 95% $\frac{\text{kg } CO_2e}{\text{kWh}}$	α	β
Electricity	0.267	0.029	0.208	0.321	81.77	3.21E-3
Gas	0.255	0.033	0.196	0.327	58.46	4.44E-3
Oil	0.315	0.038	0.244	0.393	67.32	4.67E-3

Bibliography

[1] F. P. H. D. Environment, "Belgium's greenhouse gas inventory (1990 - 2019)," 2021.

[2] (). "Emissions par secteur," [Online]. Available: https://climat.be/en-belgique/climat-et-emissions/emissions-des-gaz-a-effet-de-serre/emissions-par-secteur?_ga=2.26814013.1849487815.1622324715-1787315670.1622324715. (accessed: 09.06.2021).

[3] M. M. Roberto Cantoni, "Climate change 2007 : Impacts, adaptation and vulnerability," no. 261-262, p. 13, 2007.

[4] F. Consoli, "Guidelines for life-cycle assessment: A code of practice," p. 7, 1993.

[5] V. K. Clune S Crossin E, "Systematic review of greenhouse gas emissions for different fresh food categories," *Journal of Cleaner Production*, vol. 140, no. 2, pp. 766–783, 2017.

[6] R. S. Dennis Wackerly William Mendenhall, "Mathematical statistics with applications," vol. 7, p. 187, 2008.

[7] L. M. R.K. Pachauri, "Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change," *IPCC*, p. 151, 2014.

[8] W. B. P. Ciroth A. Muller S., "Refining the pedigree matrix approach in ecoinvent," pp. 5–8, 2012.

[9] S. R. Mutel C. Weidema Bo P., "The application of the pedigree approach to the distributions foreseen in ecoinvent v3," *Int J Life Cycle Assess*, vol. 1, pp. 1327–1337, 2014.

[10] H.-J. A. Brian Cox, "How to include non-co_2 climate change contributions of air travel at eth zurich," p. 8, 2019.

[11] (). "Carbon independant - aviation," [Online]. Available: https://www.carbonindependent.org/22.html. (accessed: 03.06.2020).

[12] (). "Greenhouse gas emissions per capita," [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/t2020_rd300/default/table?lang=en. (accessed: 08.06.2021).

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain
Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/epl**