Internship report 2A

Research assistant at the University of Barcelona

Lisa PARUIT

Summary

Introduction I was hired as a research assistant in the agroecology department of the University of Barcelona. This was an unpaid insternship in which I was given the opportunity to work on a meta-analysis of LCA studies of cocoa and chocolate bars prodcuction.
Methods In facts, I almost lead the project on my own, from the database design to the data extraction, analysis and interpretation. As I didn't get much guidance, I took a few false directions that however allowed me to draw on knowledge and skills I have acquired throughout my studies in AgroParisTech and face the challenges of a researcher's work.
Results Since most of my time was spent in paper reading and I had to adopt a trial and error approach, which is very time consuming, I didn't have time to gather enough data to get significant results. This report however details the methodology used for analysis and the material I created for this purpose.
Conclusion Although I was disappointed that my internship wasn't as instructive in terms of scientific content and methods than I had expected due to lack of guidance, it conforted me in the path of scientific research and was a great mean to experience the day-to-day life of a researcher. I also had fun being on my own and having to adopt a genuine sientific method to conduct my project.

Résumé

Introduction J'ai été prise comme assistante de recherche au sein du département d'agroécologie de l'Université de Barcelone. Il s'agissait d'un stage non rémunéré au cours duquel j'ai eu l'occasion de travailler sur une méta-analyse des études d'Analyse de Cycle de Vie de la production de cacao et de barres chocolatées.
Méthodologie Dans les faits, j'ai presque dirigé le projet moi-même, de la conception de la base de données à l'interprétation des résultats, en passant par l'analyse des données et le développements d'outils informatiques pour les exploiter. Comme j'ai été peu encadrée, j'ai parfois suivi des pistes erronées qui m'ont toutefois permises de tirer profit des connaissances et des compétences que j'ai acquises au cours de mes études à AgroParisTech et de me confronter aux défis quotidiens du travail de chercheur.
Résultats Comme j'ai passé la plupart de mon temps sur la bibliographie et que j'ai été contrainte d'adopter une approche par essais et erreurs très chronophage, je n'ai pas eu le temps de rassembler suffisamment de données pour obtenir des résultats satisfaisants. La méthodologie utilisée pour l'analyse et les outils que j'ai créé à cette fin sont cependant détaillés dans ce rapport.
Conclusion J'ai été déçue que mon stage n'ait pas été aussi instructif que je l'attendais en termes de contenu et de méthodes scientifiques en raison d'un manque d'encadrement. Cette expérience m'a cependant confortée dans la voie de la recherche et a été un excellent moyen de m'essayer au quotidien du chercheur. J'ai également trouvé une forme de satisfaction dans le fait d'être livrée à moi-même face à un problème donné et de devoir adopter une véritable démarche scientifique pour mener à bien ce projet.

À l'intention de l'examinateur: *Ce document a été implémenté avec LaTeX. La police par défaut étant Times New Roman 10pt au lieu des 12pt recommandés par l'école, le nombre de page s'en trouve par conséquent réduit par rapport à l'ampleur du contenu. L'utilisation de LaTeX a été motivée par (i) la pertinence de maîtriser ce langage dans le monde de la recherche (ii) la facilité de mise en page permise par le logiciel. D'autre part, ce rapport a été rédigé en anglais pour des raisons de cohérence avec le stage que j'ai effectué et des différentes notes méthodologiques que j'ai pu réaliser pour l'équipe.*

Contents

Introduction

Description of the institution and personal motivation

The University of Barcelona (UB) is a public university located in Barcelona, Spain and the national leader in terms of research [1]. Interested in agronomical research with a specific focus on how to apply knowledge and methods in ecology to the development of more resilient production systems, I was attracted to this university as it hosts a research department focusing solely on agroecology. Moreover, Spain being a major farming country in Europe, I was eager to learn more about the challenges in this sector and how agroecology research was lead there to tackle them.
I was very interested in the project currently lead by the department, focusing on biodiversity restauration and understanding its involvement in agricultural systems [2]. I therefore contacted Dr. Xavier Sans Serra who was already very busy but kindly redirected me to Laura Armengot Martinez, newly arrived to the lab and who was already working on several projects.

Job description and content

After a time in the private sector, Dr. Armengot Martinez joined the UB's agroecology lab back, where she did her pHD and her first steps in research. Her focus was mainly on cocoa production and agricultural practices comparison with an emphasis on agroforestery.
During our first meeting, Laura had told me she expected data from an experimental station in Ecuador where trials on the performance of aggroforestery in cocoa production were conducted. Unfortunately, the data was not yet available yet when I arrived to the lab and we had to find another project to work on. I was therefore redirected on a meta-analysis project focusing on the environmental impact of cocoa and chocolate bar production.

Mission content

The project goal was to conduct a meta-analysis of the environmental impact of cocoa and chocolate production as a tool for companies to assess their impact and improve their practices. Laura and the paper co-leader had already selected about 40 papers to base the analysis on. I was therefore assigned to litterature review and data extraction from these papers. The data was to be stored in a database and analyzed to provide a comprehensive and accessible overview of chocolate production impact based on Life Cycle Analysis method.

Context of the study: impact studies in the food industry

Definition of a supply chain:

Agroindustrial production systems analysis requires the concept of **supply chain**. From a **cradle-to-gate approach**, supply chains are considered as a sequence of processes from the initial primary production to its end use. Each sequence is then analyzed from the three main dimensions of sustainability: economy, environment, and social balance [3]. In scientific impact studies, **potential indicators** are measured on a defined **functional unit** of end product for each subprocess then aggregated to provide a global view of the product's impact along the supply chain. This method, called **Life Cycle Analysis (LCA)**, is the most common tool for impact assessment and can be applied to both environmental and social evaluation.

Life Cycle Analysis

Life Cycle Analysis is originally designed for environmental sustainability evaluation of a value chain. Over the last few decades, the method has been normalized (ISO 14040 and ISO 14044) and allows identifying both the use/destruction of resources and the emission of substances that can create harmful pollution at different stages of the value chain [4]. The choice of the indicators is at the discretion of the scientists and is generally made with regard to what is at stake in the agroindustrial process.

In order to carry out a quantifiable analysis of the impacts resulting from the production and distribution of highly manufactured products, such as cocoa or chocolate, LCA hence seems to be a highly satisfying tool.

Relevance of environmental LCA of cocoa and chocolate production value chain:

Cocoa production value chain is the backbone of many countries' economies, encompassing a wide variety of actors across the world [5]. This supply chain has been known for showing social discrepancies among actors and environmental issues [6]. According to the FAO, conventional cocoa cultivation can have significant environmental impacts. Some of these impacts include deforestation, soil erosion, biodiversity loss, etc. that occure due to the excessive use of pesticides and inorganic fertilizers alongside with the effects of monoculture agriculture [7]. While consumers are increasingly interested in the sustainability of its end product - chocolate - and LCA studies accumulate on the topic [8], there is no broad spectrum meta-analysis gathering this data to improve its accessibility.
This project consists of creating a database to gather scientific data from LCA studies that focus on cocoa and chocolate production worldwide. This tool would allow for supply chains comparison across countries, products, and practices.

Material & methods: database design

First project: creation of a comprehensive tool for data storage

Motivations

This meta-analysis was first thought of to provide a comprehensive and accessible overview of the environmental impact of chocolate production with a emphasis on the distribution of this impact between different steps and actors of the value chain. In the direction given by the paper's leaders, the results were meant to adress to companies and eventually consumers. I therefore suggested to create an app that would easily allow the user to:

i. Get a quick overview of the each step of the value chain's impact, with filtering tools and a graphic representation of the data.

ii. Compare the impact of different practices, countries, or value chains through a graphic representation of the required data.

iii. Access raw data and the references used to extract it, to allow for a more in-depth analysis and a review of the data.

iv. Eventually allow companies and actors to input their own data on the condition it is legitimately sourced.

Implementation with SQL and Python

The app was to be built using SQL for the database manipulation and requests as well as Python for the interactive program and the graphic interface. The database was built up as shown in Figure 1.

The app was built up using tkinter and gave an interface as shown in Figure 2. Original files can be found on the Github page associated to this project (cf.Appendix A). For now and since this part of the project was given up upon, the app contains the following functionality:

i. an interactive database structure graph to visualize and modify raw database

ii. a table display function from which data can be altered by doublclicking on item

iii. a path to add data to the database.

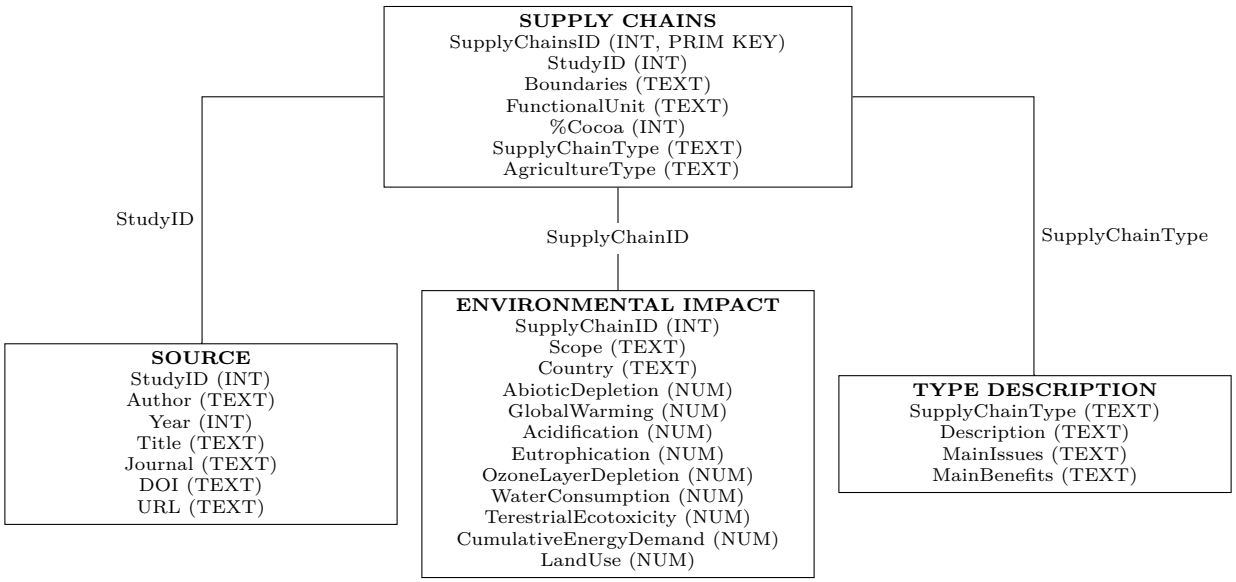
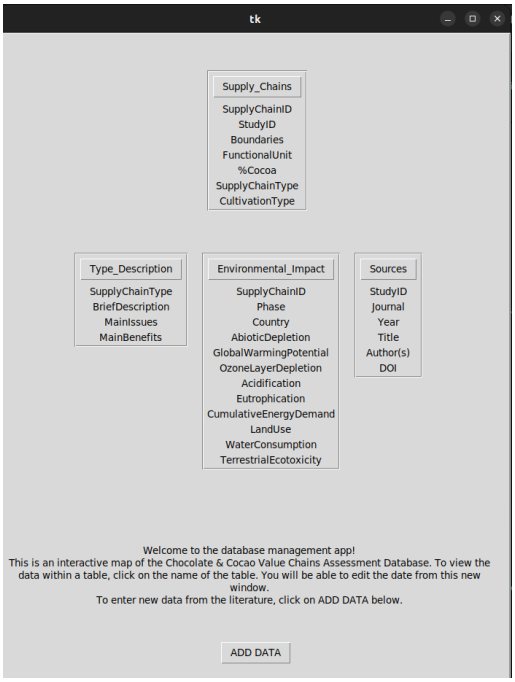
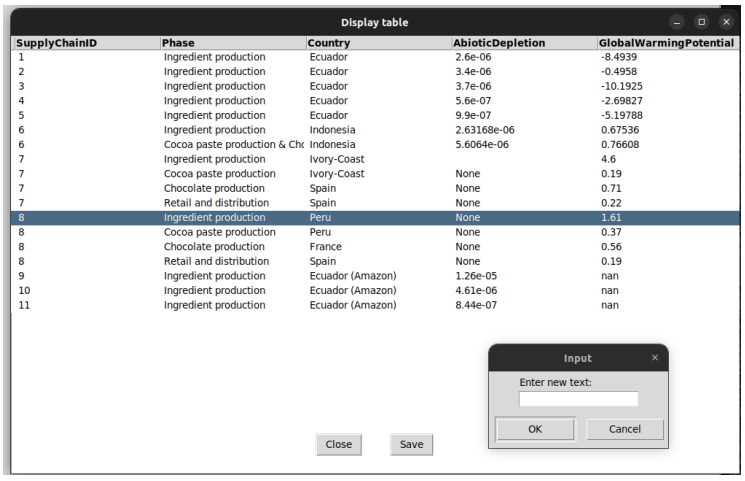


Figure 1: Graph of the SQL database design. Tables are represented as squares while nodes represent the linking item between the database's tables.



(a) Main page of the app prototype



(b) Data display and modification window that appears when double clicking on an item.

Figure 2: Image captures from the app prototype. On figure (a) is written *"Welcome to the database management app! This is an interactive map of the Chocolate & Cocoa Value Chains Assessment Database. To view the data within a table, click on the name of the table. You will be able to edit the date from this new window. To enter new data from the literature, click on ADD DATA below."*

Issues encountered and project redirection

After some time, Laura and her collegue decided the app was not necessary. The development of the code into an actual app would have required more ressources to be allocated to the project and follow up with the UB's informatic services. The app was therefore put on hold and it was decided the database would be implemented using Excel.

Final project : Excel database

Motivations

Spreadsheet softwares are widely used tools in the industry as well as in research labs and are accessible to most users. Laura and her collegue were not familiar with SQL nor Python and were thus affraid not to be able to take over the project when my internship would come to an end. Hence, it was decided to implement the database using Excel, although the goal of the app was to design an ergonomic tool.

Description of the database

Table 1 shows a description of the different columns and their description. The methodology used for design and filling out is detailed throughout the folowing subsections.

Supply chain type

Conventional = This is the default supply chain type. The primary objective of supply chain management is to fulfill customer demands through the most efficient use of resources, including distribution capacity, inventory, and labor [9]. A supply chain seeks to match demand with supply and do so with minimal inventory. This is usually done through economies of scale on volumes. Therefore, conventional supply chain is structured around collection centers of collectors/brokers supplied by conventional producers [10]. Since farm types vary from one country to another, we chose not to focus on farm characteristics but to consider «conventional» farms as representative entities of the producing country (cf. Table 2).

Semi-processed = This sub-chain is structured around a small group of primary, industrial processors, which use cocoa blends to produce semi-processed products (i.e. liquor, butter, powder) mainly for the international market.[10]

Quality = This type of supply chain is structured around private or corporate collection centres. It focuses on moderate volumes of quality cocoa to be exported as beans by national agro-exporters. Quality supply chains produce smaller quantities of semi-processed products. [10]

Premium = It is structured around producers who produce very high-quality cocoa, in very small volumes, traded (after careful post-harvesting) at very high prices on the international market.[10] Producers usualy work with artisan chocolate makers who select their beans carefully for a high quality end product.

Fair trade = Fairtrade certification tries to ensure global social welfare and address the environmental responsibilities of consumers. For now, the products most involved with fairtrade are coffee, cocoa, bananas and cane sugar. In 2021, 10% of the world's cocoa was fairetrade [6], showing the uprising of an alternative sub-chain that is nothing but marginal on the international market.

Organic = This sub-chain can be identified through the involvement of a handful of associative collection centres or with cooperative statutes working with organic certified producers only. This sub-chain represents a very low weight in terms of volume and value, not due to a lack of demand, but to supply capacity (as the organic price differential does not compensate for the certification costs). It is linked in many cases to Fairtrade certification.[10]

Table 1: Overview of the database structure.
(NB: Columns are heredisplayed as rows for formatting purposes.)

Reference	Title	title of the article / scientific reference used
	Author(s)	Main autor(s) of the study
	Journal	Journal where the study was published
	Year	Year of publication
	DOI	Digital Object Identifier with hyperlink towards the online version
Supply Chain Characteristics	Supply Chain Type	Conventional, Quality, Organic, Fair trade, etc.
	Agriculture Type	Conventional, Organic, Agroforestry, etc.
	Boundaries	Cradle-to-gate, Cradle-to-retailer, cradle-to-grave, etc.
	Functional unit	Unit of product used for the study
	FU type	Raw material for Cocoa or End product for any chocolate of confectionnery type
	% Cocoa	Percentage of cocoa in the final product
	Cocoa mass	Cocoa mass in the final product
	Country	Country where the SC last sub-process within the boundaries of the study was conducted (eg. country of retail when using a cradle-to-retailer approach)
Environmental Assesment	AD (kg Sb eq.)	Abiotic Depletion potential
	GW (kg CO2 eq.)	Global Warming potential
	ODP (kg CFC-11 eq.)	Ozone Depletion potential
	AC (kg SO2 eq.)	Acidification potential
	EU (kg PO4 eq.)	Eutrophication potential
	CED (MJ)	Cumulative Energy Demand
	TE (kg 1,4-DB eq.)	Terrestrial Ecotoxicity
Environmental Assesment *per kg ofcocoa*	Idem	In each LCA study, impact potential indicators are calculated for the functional unit. Values of these columns are defined as: $\frac{EA}{kgCocoa} = \frac{EA}{FU} \times \frac{MassCocoa}{FU}$. This allows for a comparison across studies that don't have the same functional unit.

Table 2: Average farm characteristics in several cocoa producing countries, defining «conventional» agriculture type. *(source: [12])*

Country	**Farm size**	**Density**	**Inorganic fertilisat°**	**Yield** (FAO)	**Labour**
Ivory Coast	2.9 ha	975 trees/ha	10% farms	500-600 kg/ha	73 day/ha total from household (94% farms) and seasonal workers (hired)
Ghana	2.1 ha	1245 trees/ha	80% farms	400- 500 kg/ha	120 day/ha total from houshold and farming communities (hired & communal)
Indonesia	0.8 ha	890 trees/ha	80% farms	700-800 kg/ha	—
Ecuador	3.7 ha	625 trees/ha	33% farms	800-900 kg/ha	39 day/ha from household (100%) and daily contractors
Brazil	5.5 ha	980 trees/ha	30% farms	500-600 kg/ha	household and seasonal hired labour depanding on farm size
Nicaragua	0.9 ha	650 trees/ha	33% farms	500-600 kg/ha	91 day/ha from household (80% farms) and hired labour
Costa Rica	7.4 ha	450 trees/ha	400-500 kg/ha	—	

Agriculture type

Conventional = This is the default agriculture type. Farms are considered conventional when they match the most common agricultural practices for cocoa production in the country.
Table 2 shows the diversity of «conventional» farming across the world's cocoa producing countries.

Organic = Organic farming consists of a system of agricultural practices based on standards through which it is possible to ensure food production without using certain chemical products that are harmful for the environment and can cause damage to farmers' health while promoting a reduced use of resources. Inputs are regulated by the corresponding regional organic certification laws, which have previously analysed the noxious impacts of such additives and their organic origin. [6] However, regulations can change from one country to the other.

Technified = This type of agriculture is characterized by the use of technology to build up highly efficient agricultural processes (technified irrigation, high fertilization doses, mechanisation, etc.) [11] resulting in very high production yields. This model can only be applied in oversized farms with high economic input and therefore often found within higly integrated supply chains (ie. upstream, core and downstream phases are managed by a single decisionary actor). This cannot be found in every country.

Agroforestry = Agroforestry is defined as a dynamic, ecologically based, natural resource management system that integrates trees in farmland for increased social, economic and environmental benefits [13]. However, the integration of trees in agricultural system can look different depending on the biome and the land characteristics (cf.Table 3) shows the diversity of agroforestry practices across cocoa producing countries.

Table 3: Common agroforestry practices in several cocoa producing countries. *(source: [12])*

Country	Agroforestry practices
Ivory Coast	- In the first 2-3 years of establishment, cocoa trees may be intercropped with food crops (maize, plantain, cassava and vegetables). Poultry and livestock can also be breeded on farm. - Cocoa is generaly cultivted under the shade of a selectively thinned forest.
Ghana	- Cocoa trees are planted in association with other crops (banana, plantain, cassava, etc.) and sometimes with livestock. It concerns 53.9% of cocoa farmers. - Cocoa trees often constitutes the canopy as 66% of cocoa plantations have little or no shade.
Indonesia	Coconut, banana, durian, rambutan, avocado, robusta coffee, spice, ginger and other fruit crops are commonly grown on cocoa farms and sometimes associated with livestock (eg.tree prunings and pod husk may be used to feed goats)
Ecuador	- Around 57% of cocoa farms are shaded with timber, fruit and shade trees. - Smaller cocoa farms often have a larger area devoted to subsistence crops (cassava, rice, sweet potato, beans, tomatoes, bananas) corn, passion fruit.
Brazil	- In the Atlantic forest region of Southern Bahia, cocoa trees are planted under the canopy of the cabrucas. - Intercropping with rubber is sometimes encountered and provides shading to the cocoa trees. Otherwise, food crops and livestock remain exceptional in cocoa farms.
Costa Rica	Much of the cocoa is produced in smallholder farms with diverse integrated agroforestry and high levels of shade.

Scope & boundaries

LCA methodology uses a cradle-to-approach that requires well defined boundaries between each step of the overall production process. Cocoa supply chain flowchart is shown on Figure 3. Boundaries considered in our methodology are defined as indicated by the dotted lines on the graph.
Cradle-to-farm gate = As post-harveting processes are usually carried out on farm, we chose to consider the upstream phase overall in this scope.

Cradle-to-cocoa mass manufacturer's gate = This scope is defined as the last sub-process of the cocoa processing phase but the first step of the core phase. It usually takes place in the cocoa producing country.

Cradle-to-chocolate manufacturer's gate = This scope is defined as the second step of the core phase. It can take place in a completely different country than that of where the raw material is produced. Transport is therefore taken into account within this boundary. Packaging is also included when considered in the study.

Cradle-to-retailer = This scope is defined as the fist sub-process of the downstream phase. Distribution can take place in many different countries.

Cradle-to-grave = This scope is rarely used as end-of-life processing are hard to assess. It is however the most holistic approach of LCA studies.

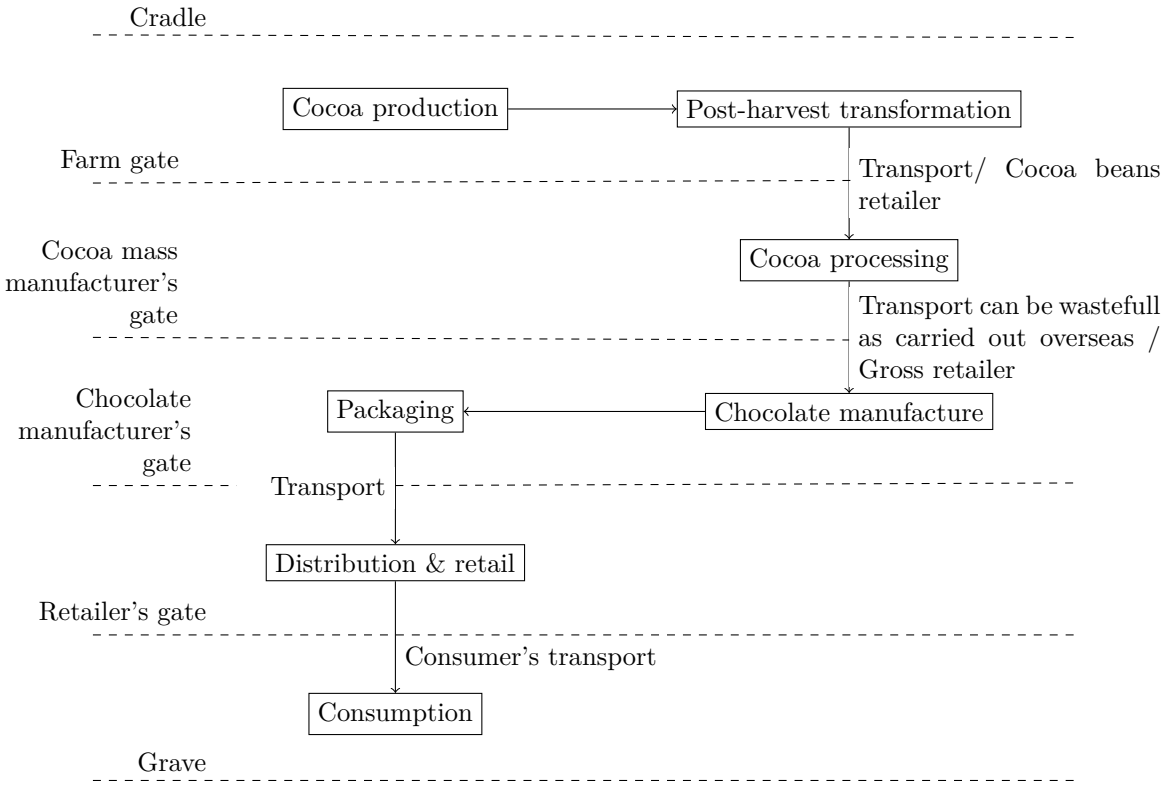


Figure 3: Chocolate supply chain with boundaries as defined in LCA studies and considered in this meta analysis. *NB:* Transport is allocated to the latest phase.

% Cocoa and Cocoa mass

In order to compare studies that do not have the same functional unit and provide a comparison depending on product type based on studies that do not necessarily have the same scope, environmental indicators are divided by the total cocoa mass in the final product. Therefore, studies that focus only on cocoa production can be treated as studies that focus on chocolate or other cocoa end-products. Cocoa percentage is defined as the percentage of cococa solids (cocoa powder + cocoa mass) in the final product [14]. Cocoa mass is then calculated as $FU_{endproduct} \times \%Cocoa$.

Environmental assessment

In product base LCA studies, the environmental impact of the product is assessed through **impact potential indicators**. These indicators measure the potential impact of the process on the environment using inference data.

Abiotic Depletion potential (AD) = This indicator represents the depletion of non-renewable resources(abiotic, non-living (fossil fuels, metals, minerals)). It is based on concentration reserves and the de-accumulation rates. It is expressed in kg antimony equivalents.

Global warming potential (GW) = This represents the potential change in climate attributable to increased concentrations of CO2, CH4, and other GHG emissions that trap heat. It leads to increased droughts, floods, losses of polar ice caps, sea-level rising, soil moisture losses, forest losses, changes in wind and ocean patterns, and changes in agricultural production. It is expressed in CO2 equivalents usually for time horizon 100 y.

Ozone depletion potential (ODP) = This is the potential for the reduction in the protective stratospheric ozone layer. The ozone-depleting substances are freons, chlorofluorocarbons, carbon tetrachloride, and methyl chloroform.

Acidification potential (AC) = This indicator is based on the potential of acidifying pollutants (SO2,NOx, HCl, NH3, HF) to form H1 ions. It leads to damage to plants, animals, and mineral structures.

Eutrophication potential (EU) = Eutrophication leads to an increase in aquatic plant growth attributable of nutrients left by over-fertilization of water and soil, such as nitrogen and phosphorus. EU potential measures nutrient enrichment that may cause fish death, declining water quality, decreased biodiversity, and foul odors or tastes.

Terrestrial Ecotoxicity (TE) = This potential focuses on the emissions of toxic substances into the air, water, and soil. It includes the fates, exposures, and effects of toxic substances.

(source: [15])

Preliminary results

WARNING! *The following results are based on the data collected so far. Data is still lacking to have significant results on this cross-methodology study. Table 4 shows the number of studies per country and potential impact indicators*

Country	Tot	AD	GW	ODP	AC	EU	CED
Ecuador	9	7	9	7	7	7	9
Ghana	5	5	5	5	5	5	5
Indonesia	4	3	4	3	4	3	2
Ivory Coast	1	1	1	1	1	1	1
Peru	3	2	3	2	3	3	2
Philippines	1	0	1	0	1	1	0

(a) Number of studies per country

Country	Tot	AD	GW	ODP	AC	EU	CED
Technified	2	0	2	0	2	2	0
Conventional	15	10	15	10	11	10	13
Organic	3	2	3	2	3	3	2
Agroforestry	4	3	4	3	4	4	3

(b) Number of studies per agriculture type

Table 4: Number of studies per country (a) or agriculture type (b) currently recorded in the database.

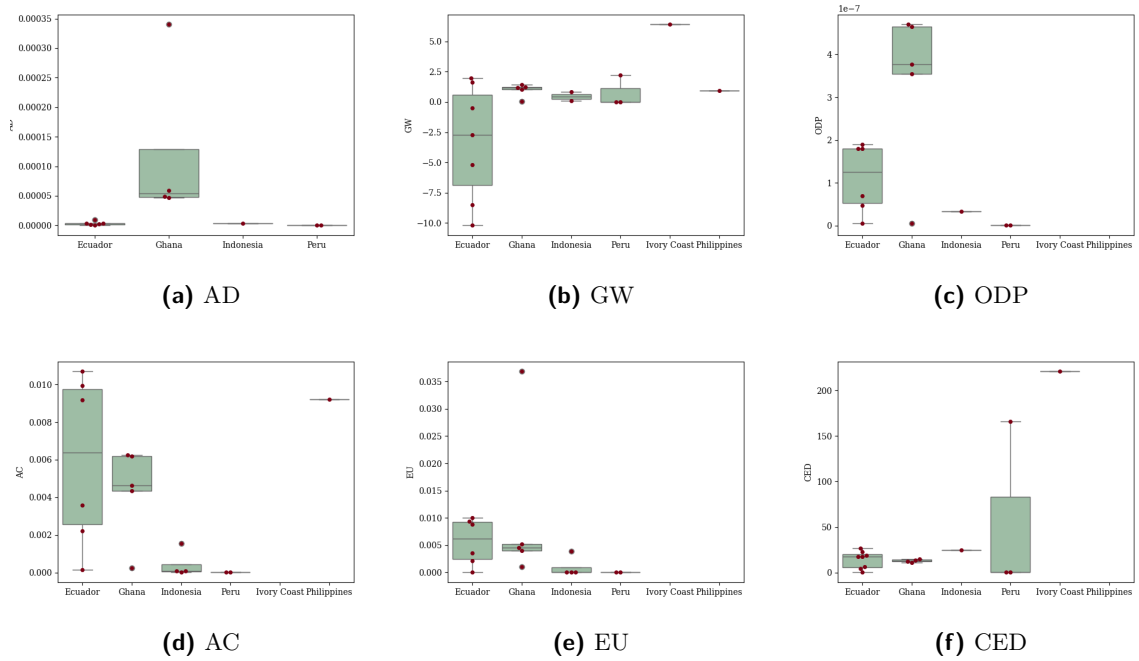
As the experimental plan for agricultural type is even less balanced than that of countries, we will only focus on the impact of the country on environmental impact potential indicators (EIP) in this *Preliminary results* section.

Statistical limitations for preliminary sample sizing

The number of studies per country and per agriculture type is not sufficient to have significant results. In addition to showing this limitation in our preliminary results using statistical methods, I also wanted to consider the number of studies needed to have a significant mean for each country and each indicator. Appendix B details why, after large mathematical considerations, this was impossible to estimate to my knowledge.
Statistical analysis will therefore be limited to analysis of variance (ANOVA) and ordinary least square regression (OLS) for t-testing of country-relative effect on global modelling.

Environmental impact per country

NB: Graphs and tabs presented in this section are generated using the code in Appendix C



(a) AD **(b)** GW **(c)** ODP

(d) AC **(e)** EU **(f)** CED

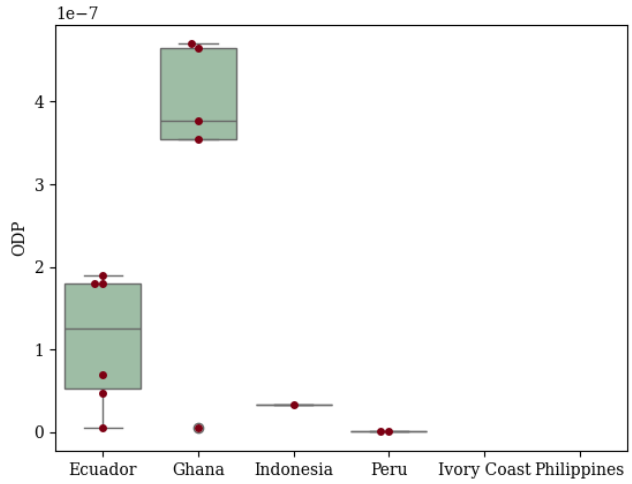
Figure 4: Boxplot of the environmental impact by country for AD (a), GW (b), ODP (c), AC (d), EU (e) and CED (f).

Figure 4 shows that boxes often overlap between countries, meaning that significant differences between countries are rare, especially with this type of poor sampling. This is confirmed by Table 5 where one can see that the variance analysis shows low significance levels for most indicators except maybe for cumulative energy demand. Indicators with highest significance levels to this test will be discussed in the following subsections.

EIP		**sum_sq**	**df**	**F**	**PR(>F)**	
AD	Country	4.0504463781035557e-08	3.0	1.9422622189446004	0.19338108232561607	
AD	Residual	6.256281472083386e-08	9.0			
GW	Country	124.06964513902007	5.0	2.2551498730945423	0.11028809818202166	
GW	Residual	143.04196861151527	13.0			
ODP	Country	3.8586514921702223e-13	5.0	4.323336998496465	0.033755789246336404	*
ODP	Residual	1.7850338724518362e-13	10.0			
AC	Country	0.00016548826738867182	5.0	3.385865383820226	0.04169112467390858	*
AC	Residual	0.00012707814589045462	13.0			
EU	Country	0.0004313173474729028	5.0	1.1319763909740337	0.3724091953540076	
EU	Residual	0.00099067888020579	13.0			
CED	Country	51412.015159260685	5.0	6.542172031985661	0.004938260254344846	**
CED	Residual	18860.530689037085	12.0			

Table 5: Results of ANOVA type II test with $\alpha = 0.05$ using stastmodels in Python. *Significance codes: 0 *** 0.001 ** 0.01 ** 0.05 * 0.1 ' ' 1*

Ozone Layer Depletion Potential



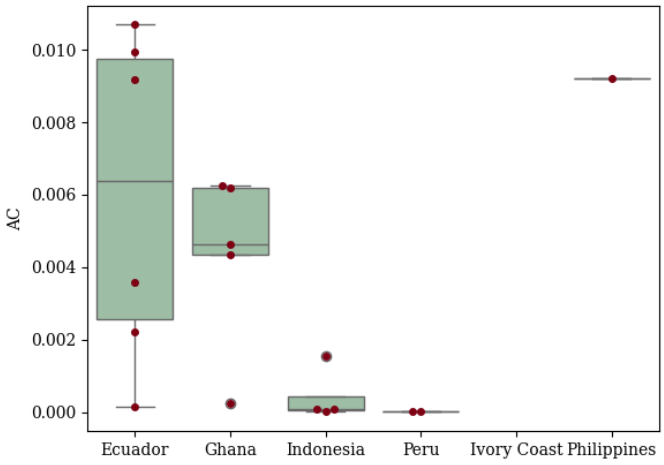
	p-value
Intercept	0.06710618701738431
Country[T.Ghana]	0.020459288019114303
Country[T.Indonesia]	0.5983357412200276
Country[T.Ivory Coast]	
Country[T.Peru]	0.3314359429092527
Country[T.Philippines]	

Table 6: Two-tailed p values for the t-stats of the parameters obtained from the OLS for ODP. *Intercept = Ecuador*

Figure 5: Boxplot of the abiotic depletion potential by country.

One can see that the potential impact of Ghana on the ozone layer seems to stand out from the other countries. The prediction is significant enough as the p-value is below 0.05. This result is not surprising as Ghana is the second largest producer of cocoa in the world and is known for its extensive use of pesticide [12]. Indeed, The production of pesticides used in cocoa farming can involve the emission of halogens and CFCs, which contribute to ozone layer depletion. This is particularly relevant during the manufacturing process of these chemicals [16].
Even for Ghana, the p-values obtained when testing the coefficients for each country obtained through OLS regression are not satisfying (Table 6). This is due to the poor sample sizes and the lack of representativity of the available data.

Acidification Potential



	p-value
Intercept	0.0004406732579534725
Country[T.Ghana]	0.40829562393258445
Country[T.Indonesia]	0.017099871702539658
Country[T.Ivory Coast]	0.016876817226944035
Country[T.Peru]	0.03689435548390101
Country[T.Philippines]	0.35165822991591933

Table 7: Two-tailed p values for the t-stats of the parameters obtained from the OLS for AC potential. *Intercept = Ecuador.*

Figure 6: Boxplot of the acidification potential by country.

On Table 5, the acidification potential shows a p-value just below 5% for the ANOVA test. This means that the variance between countries exists but needs a stronger validation to be considered significant. However, the OLS coefficients tested on Table 7 show a significant impact of Ecuador on acidification potential prediction ($p < 0.001$). This may be due to the high variance of the sample obtained so far for this country with regards to others, which might explain the low p-value obtained for the ANOVA test. However, one can see we need more data to confirm the results extracted for other countries as suggested by the relatively high p-values on Table 7.

Cumulative Energy Demand Potential

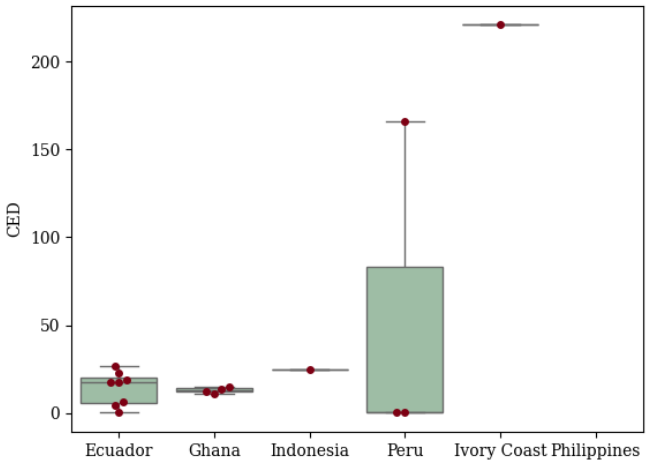


Figure 7: Boxplot of the cumulative energy demand potential by country.

	p-value
Intercept	0.3198676334701379
Country[T.Ghana]	0.9544876898748276
Country[T.Indonesia]	0.813458743757204
Country[T.Ivory Coast]	0.0003637803978548956
Country[T.Peru]	0.15179957810887446
Country[T.Philippines]	

Table 8: Two-tailed p values for the t-stats of the parameters obtained from the OLS for AD. *Intercept = Ecuador.*

Table 5 suggests that the variance between countries is the highest for cumulative energy demand. On Figure 7 and Table 8, one can see that the impact of Ivory Coast on this indicator is extremely significant as it is the only country to present CED potential values above 200 MJ. Nevertheless, this datum is based on a single LCA study for the entire country. This result could thus be considered as an outlier and needs to be confirmed with more data.

Aggregated impact and conclusion

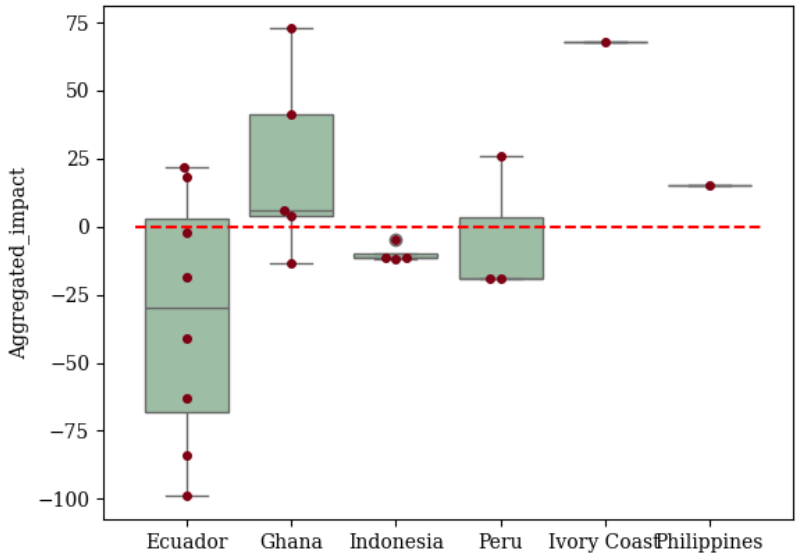


Figure 8: Addition of the relative gap to the mean of each indicator for each country.

The aggregated impact is calculated as follows:

$$Aggregatedimpact_j = \sum_{i=1}^{6} \frac{I_i j - Mean(I_i)}{m}$$

with

- $I_i j$ = the value of the indicator i for the country j,
- $Mean(I_i)$ = the global mean of the indicator i and
  - m = the order of magnitude of the mean.

Overall, Ghana, Ivory Coast and Phillippines seem to have the greatest impact. While this number is not surprising from the two major world's cocoa producers, Ghana and Ivory Coast, representing alltogether 63.5% of global production [12], it is rather intriguing when it comes to Phillipines. This result might be due to the lack of representativity of the available data. Indeed this preliminary result is based on a single LCA for the entire country. It is yet to be verified with a wider literature review.
On the other hand, the apparently low impact of Ecuador, which is a major producing country, can be explained by the fact that Ecuadorian cocoa production relies on small producer who practice agroforestry and low imput agriculture more readily than on the African West coast.

In conclusion, the preliminary results show that the impact of the country on the environmental potential indicators can be significant but we still need more data to validate the assertions made in this section and realise an effective statistical analysis.

Conclusion and personal insights

A disappointing turn of events...

This internship was an opportunity for me to discover the world of research in agroecology in a way that I was expecting to be thorough. The initial project was to analyse data acquired over 10 years of experimental

agroforestry surveys. As I want to specialise more in statistical modelling and applied mathematics, I see agroecology studies as statistical exploration of an environment shaped by ecological interactions and natural hazards. The change of project was therefore a great disappointment as I was expecting to focus on data analysis and statistical modelling.

...and a lack of guidance...

I also found that my supervisor, like the rest of the researchers in the department, seemed poorly organised and rather involved in their teaching activities than in their research work. Laura, although reachable, was very busy during my time in Barcelona and seemed unsure of what I could bring to the table. I have the feeling that I have been put on a project out of spite rather than necessity for the research team and that my added value as an M1 engineering student was far from being fully used. I would certainly have benefited more from this internship in terms of knowledge deepening and new skills development if I had had more guidance on this project.

...which lead to a great learning experience!

As I was instructed to start the project without any clear guidelines, I had to be very proactive, suggest and support ideas that I was then able to test by myself in a scientific approach. In a way, I had a free rein on this project, which meant I faced a number of pitfalls that I might have avoided if I had been better supervised. Although I would have preferred to get more out of my time in the lab by being supported more, I appreciated this freedom, which also allowed me to confront with no concessions the day-to-day decision making and challenges of an actual researcher.

Conclusion

I will take something very positive from this experience: I love research and its conundrums. Although most of my time was spent doing literature review and filling out the database, I very much enjoyed the more creative moments that were also part of my work. The need to stand back, zoom in and out during both the reflection and the realisation as well as the need to think the project globally were a pleasure and very instructive. I think I managed to draw on the skills and knowledge I had acquired during my studies whenever I could and to be critical in the light of this knowledge and the objectives that were set.

A. Github repository

General Github commit link:

https://github.com/lisaparuit/LCA-for-cocoa-and-chocolate-production-database

Other links:

Prototype app folder: https://github.com/lisaparuit/LCA-for-cocoa-and-chocolate-production-database/blob/lisaparuit-prototype-app/New_app
Associated SQL database: https://github.com/lisaparuit/LCA-for-cocoa-and-chocolate-production-database/blob/lisaparuit-prototype-app/ProjectUB.db
Final Excel database: https://github.com/lisaparuit/LCA-for-cocoa-and-chocolate-production-database/blob/main/Chocolate%20LCA.xlsx

All the other programms/files mentioned in this report can be found on the general github link.

B. Impossibility of sample size estimation

Motivations

Sample sizes are not equal for each group (eg. country, agriculture type, product, etc.) and are not big enough to give out statistically satisfying results (cf. Table 4). In order to assess the number of studies needed to get a satisfying confidence interval, sample size estimation methods exist [17]. However, these methods are not applicable to our case study for the following reasons:

(i) Difficulty to find expected values: This study is a meta-analysis that aims to assess and compare cocoa and chocolate production processes around the world. No meta-analysis of this spectrum yet exist. It is therefore very difficult to obtain expected variables to compare the empirical variables obtained from partial data to.

(ii) Values close to zero: Some indicators have values that are very close to zero (eg. AD, ODP, etc.). Under the hypothesis that the empirical mean is the true mean, we want to find the sample size needed to validate this, we compare this mean to 0. However, when values are very small, this makes the sample size estimation very high with regards to other countries and/or indicators.

$$n \geq \frac{2\sigma^2}{(\mu_1 - \mu_2)^2} \times (Z_\alpha + Z_{2\beta})$$

In this equation to estimate the sample size (n), $\mu_1 = \bar{x}_{country,indicator}$ in both cases. However, the other variable change under the hypothesis made in each of the cases detailed above:

(i) $\mu_2 = \mu_{country,indicator}$ and $\sigma = \sigma_{country,indicator}$ from the literature.

(ii) $\mu_2 = 0$ and $\sigma = \bar{s}_{country,indicator}$, the empirical standard deviation.

Testing of hypothsis (ii)

Using the programm written below Listing 1, we can test the hypothesis (ii). The results are shown in Table 9.

Country	AD	GW	ODP	AC	EU	CED
Ecuador	5.14e+00	1.38e+01	2.93e+00	3.21e+00	3.17e+00	2.38e+00
Ghana	2.31e+01	1.67e+00	1.82e+00	1.78e+00	1.17e+01	7.94e-02
Indonesia	nan	1.30e+01	nan	1.59e+01	2.24e+01	nan
Peru	7.51e-01	1.67e+01	7.88e-01	7.85e-01	7.92e-01	1.65e+01
Ivory Coast	nda	nan	nda	nda	nda	nan
Philippines	nda	nan	nda	nan	nda	nda

Table 9: Output of Listing 1 that tests for hypothesis (ii). nan = values are empty or not enough in the database and could not be calculated ; nad = data isn't available

```python
import pandas as pd
import numpy as np
import math as math

# Import the data
raw_data = pd.read_csv('CSVfiles/Chocolate LCA - Main.csv', header=1)
data = raw_data[['Country***', 'Agriculture type*', 'AD (kg Sb eq) .1', 'GW (kg CO2 eq) .1', 'ODP (kg CF11 eq).1', 'AC (kg SO2 eq).1', 'EU (kg PO4 eq).1', 'CED (MJ).1']].loc[(raw_data['Boundaries / production phase ******'] == 'Cradle to farm gate')]
data = data.replace(0, float('nan')) # Replace 0 values with NaN

def my_funky_function(indicator,country):
    dico = {'AD': 'AD (kg Sb eq) .1', 'GW': 'GW (kg CO2 eq) .1', 'ODP': 'ODP (kg CF11 eq).1', 'AC': 'AC (kg SO2 eq).1', 'EU': 'EU (kg PO4 eq).1', 'CED': 'CED (MJ).1'}
    # Extract the data for the given country and the given indicator
    data_indicator_country = data[dico[indicator]].loc[data['Country***'] == country]
    # Calculate variables
    mean_indicator_country = np.nanmean(data_indicator_country)
    if math.isnan(mean_indicator_country) or mean_indicator_country == 0:
        return 'nda'
    std_indicator = np.nanstd(data_indicator_country,ddof = 1)

    #results
    S = (2*(std_indicator)**2/mean_indicator_country**2 )*(1.96+0.842)
    return S = "{:.2e}".format(S)

# Test the function
df = pd.DataFrame(columns=['AD', 'GW', 'ODP', 'AC', 'EU', 'CED'], index=data['Country***'].unique())
for indicator in ['AD', 'GW', 'ODP', 'AC', 'EU', 'CED']:
    for country in data['Country***'].unique():
        df.loc[country, indicator] = my_funky_function(indicator, country)
```

Listing 1: Python code to test hypothesis (ii)

Overview of the suggested method

In order to circumvent these issues, I suggested a way to assess for the minimal sample size to get 95% confidence interval (CI) with an acceptable size.
Instead of assessing the p-value of a test with our uneffective sample, let's turn the problem upside down and rather find the sample size for which the 95% CI is *small enough* to be useful to the interpretation. No explicit criteria could be used to define the CI's *smallness*. However, we can fairly suppose that the value of an indicator k follows a gaussian law of parameters μ_k and σ_k with "country" as an additional qualitative parameter. Therefore, the curve CI vs.n can be plotted using the following formula:

$$CI = \left[\bar{X}_{i,k} \pm t_{\alpha=0.025, n_{i,k}-1} \times \frac{\sigma_k^2}{\sqrt{n_{i,k}}}\right] \tag{1}$$

with:

CI = the 95% confidence interval

$\bar{X}_{i,k}$ = the sample mean for the country i and the indicator k

$t_{\alpha=0.025,n-1}$ = the 97.2th quantile of Student's law for $\alpha = 0.025$ and $n-1$ degrees of freedom

σ_k = the standard deviation of the sample for the indicator k, all countries combined. We can make the hypothesis this standard deviation is that of the population for the indicator k.

$n_{i,k}$ = the size of the sample

This curve is positive, decreasing and asymptotic to a limit (when $n \to \infty$) being, in our case, the best mean estimator (cf. weak law of large numbers [18]). An acceptable CI size could therefore be that of the curves' inflection point as it flattens on the asymptote line. The following exemple illustrates the method step by step.

Eg. Abiotic Depletion potential in Peru

i. Select the sample group. In this example, we have chosen to work on **Peru** and to focus on **abiotic depletion**.

ii. Calculate the order of magnitude (m) of the indicator's values using:

$$m = 10^{\lceil \log_{10}(\sigma^2_{AD}) \rceil} \tag{2}$$

with σ^2_{AD} the variance of the AD indicator **for the entire set of studies** sampled so far

iii. Calculate the standard deviation of the indicator for studies that have been conducted in Peru (incomplete sample). This is carried on using the embedded `STDEV()` function in Excel.

iv. Derived from Equation 1 and Equation 2, use the following formula to calculate the CI variation indicator ($1/2CI/m$) :

$$1/2CI/m = \frac{t_{\alpha=0.025,n-1} \times \sigma^2}{\sqrt{n} \times m} \tag{3}$$

v. Using a set of x and y coordinates as shown on Table 10, the knee point of the function (ie. the point of maximum curvature) is found using the kneedle algorithm (cf. Figure 9).

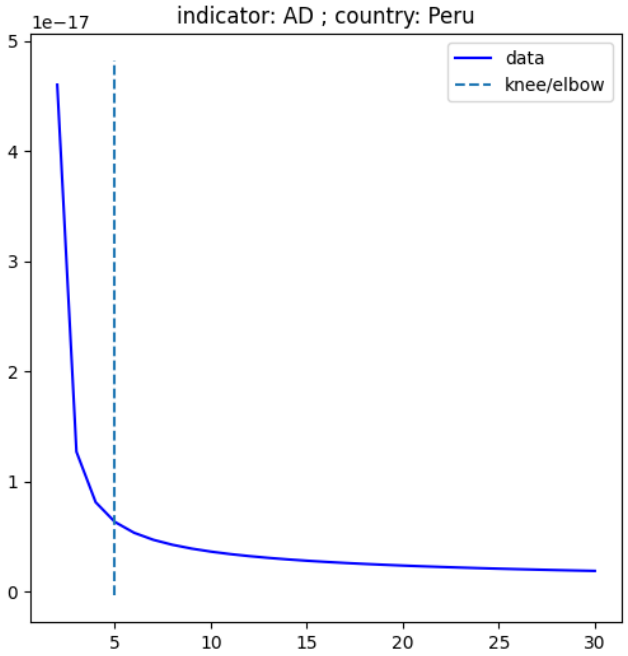


Figure 9: 1/2 CI / m vs. n Result of the kneedle algorithm for AD potential in Peru.

n	1/2 CI / m
2.0	4.6015115207070834e-17
3.0	1.271981162661356e-17
4.0	8.145920000000003e-18
5.0	6.356300498472364e-18
6.0	5.373984536486873e-18
7.0	4.7353808151223274e-18
8.0	4.2811072960158345e-18
9.0	3.935573333333335e-18
10.0	3.6623728984580484e-18
11.0	3.4394484516826e-18
12.0	3.2531147327650996e-18
13.0	3.0942508225914238e-18
14.0	2.9556955265578855e-18
15.0	2.8356434867592233e-18
16.0	2.727680000000001e-18
17.0	2.6325786883943738e-18
18.0	2.546338659504838e-18
19.0	2.4678525791336983e-18
20.0	2.3962062217764154e-18
21.0	2.3306368974452753e-18
22.0	2.2705023489007097e-18
23.0	2.214189541142391e-18
24.0	2.162344225202516e-18
25.0	2.113536000000001e-18
26.0	2.068476100529456e-18
27.0	2.0258682085612732e-18
28.0	1.9854927324542328e-18
29.0	1.9471567492434044e-18
30.0	1.911624753673647e-18

Table 10: Coordinates for the curve of the 1/2 CI / m vs. n in Figure 9

In practice, the method was implemented in Python using the following code:

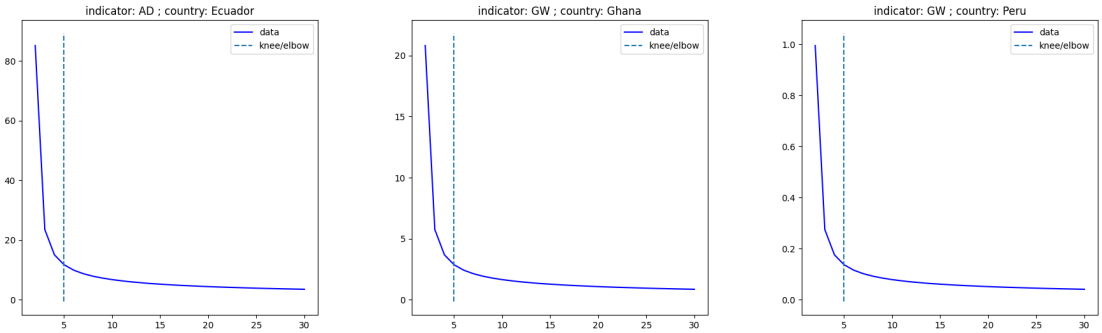
```python
import kneed as kn
import pandas as pd
import math as math
import numpy as np
import matplotlib.pyplot as plt


##### DATA #####

# Import the data
raw_data = pd.read_csv('Chocolate LCA - Main.csv', header=1)
data = raw_data[['Country***', 'Agriculture type*', 'AD (kg Sb eq) .1', 'GW (kg CO2 eq) .1',
    'ODP (kg CF11 eq).1', 'AC (kg SO2 eq).1', 'EU (kg PO4 eq).1', 'CED (MJ).1']].loc[(raw_data['Boundaries / production phase ******'] == 'Cradle to farm gate')]
data = data.replace(0, float('nan')) # Replace 0 values with NaN

# List of the student quantile for n = 2 to 30 and alpha = 0.025
t_arr = np.array([12.71, 4.303, 3.182, 2.776, 2.571, 2.447, 2.365, 2.306, 2.262,
    2.228, 2.201, 2.179, 2.16, 2.145, 2.131, 2.12, 2.11, 2.101, 2.093, 2.086, 2.08,
    2.074, 2.069, 2.064, 2.06, 2.056, 2.052, 2.048, 2.045])
n_arr = np.arange(2, 31)

def my_funky_function(indicator, country):
    dico = {'AD': 'AD (kg Sb eq) .1', 'GW': 'GW (kg CO2 eq) .1', 'ODP': 'ODP (kg CF11 eq).1', 'AC': 'AC (kg SO2 eq).1', 'EU': 'EU (kg PO4 eq).1', 'CED': 'CED (MJ).1'}
    # Extract the data for the given indicator
    data_indicator = data[dico[indicator]]
    var_indicator = np.nanmean(data_indicator) # variance of the indicator values forthe entire population

    # Extract the data for the given country and the given indicator
    data_indicator_country = data[dico[indicator]].loc[data['Country***'] == country]
    var = np.nanvar(data_indicator_country, ddof = 1) # variance for the indicator values of the partial sample NB: ddof = 1 for sample std deviation
    if math.isnan(var) or var == 0:
        return 'nda'

    # Calculates the order of magnitude for the EIP indicator
    var = var_list[-1]; sign = -1 if var < 1 else 1
    m = 10**(sign*math.ceil(abs(math.log10(var))))

    # Calculates the IC values
    IC = (t_arr * (var / np.sqrt(n_arr)) )/ m

    # Knee location
    knee = kn.KneeLocator(n_arr, IC, S= 1, curve='convex', direction='decreasing', interp_method='interp1d')
    knee.plot_knee()
    knee_value = knee.knee
    table_data = np.concatenate([n_arr, IC], axis=0).reshape(2, 29).T
    table = pd.DataFrame(table_data, columns=['n', 'IC'])

    return knee_value, table
```

Listing 2: Python code to find the knee of a curve

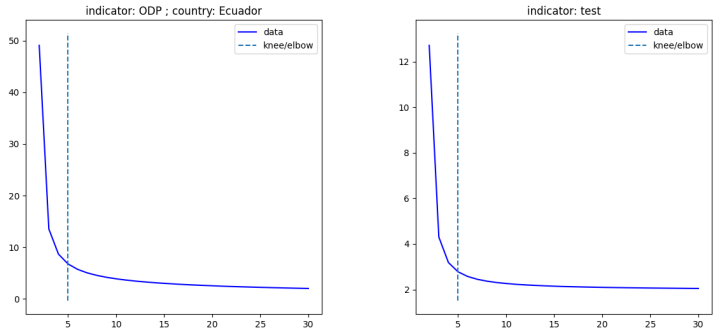
Limitations and hypothesis discarding

The method has been tested on several indicators and country. The results are shown in **??** and show that is always is $n = 5$. I therefore plotted the evolution of the Student's quantile over the sample size and noticed that the shape of the curve is similar to that of the other graphs. One could have expected this as $\sigma_{i,k}$ variation is insignificant with regards to that of the denominator. The knee of the curve is therefore always at $n = 5$

and the method is not as satisfying as expected.



(a) AD, Ecuador **(b)** GW, Ghana **(c)** GW, Peru



(d) ODP, Ecuador **(e)** student's quantile

Figure 10: Output of Listing 2 for different couples (indicator,country) and the Student's quantile (e)

C. Code for results extraction and statistical analysis

Statistical analysis has been carried out under Python 3.10.12 using the satsmodels module instead of Rstudio for technical reasons. The module however reproduces R embeded functions that are used in the following code.

```python
import pandas as pd
import numpy as np
import seaborn as sns
import math as math
import statsmodels.api as sm
from statsmodels.formula.api import ols
import matplotlib.pyplot as plt


# Load data
data = pd.read_csv('CSVfiles/Chocolate LCA - Main.csv', header=1)
data = data.loc[data['Boundaries / production phase ******'] == 'Cradle to farm gate',
                ['Agriculture type*', 'Country***', 'AD (kg Sb eq) .1', 'GW (kg CO2 eq) .1', 'ODP (kg CF11 eq).1', 'AC (kg SO2 eq).1', 'EU (kg PO4 eq).1', 'CED (MJ).1']].reset_index(drop=True)
data = data.replace(0, np.nan)
data.columns = ['AgricultureType', 'Country', 'AD', 'GW', 'ODP', 'AC', 'EU', 'CED']

def results(indicator,data=data):
    data = data[data['Country'] != 'Unknown']
    if indicator == 'GW':
        data = data[data['AgricultureType'] != 'Agroforestry']
        data = data[data['AgricultureType'] != 'Organic & Agroforestry']

    elif indicator == 'AD':
        data = data[data['AD'] <= 0.005]

    # Anova model
    model = ols(formula = indicator+' ~ Country', data=data).fit()
    anova_table = sm.stats.anova_lm(model, typ=2)
    anova_table.to_csv('CSVfiles/anova_results_'+indicator+'.csv')

    # Pairwise t-tests
    pvalues = model.pvalues
    pvalues.to_csv('CSVfiles/pvalues_'+indicator+'.csv')

    # Plot
    print(data)
    ax = sns.boxplot(data = data, x='Country', y=indicator, color='#99c2a2')
    ax = sns.swarmplot(data= data, x="Country", y=indicator, color='#7d0013')
    plt.rcParams['font.family'] = 'serif'
    plt.xlabel('')
    plt.savefig('Images/boxplot_'+indicator+'.png')

for indicator in ['AD', 'GW', 'ODP', 'AC', 'EU', 'CED']:
    results('AD')
```

Listing 3: Python code for result extraction and statistical analysis of the collected data (1/2)

```python
import pandas as pd
import numpy as np
import seaborn as sns
import math as math
import statsmodels.api as sm
from statsmodels.formula.api import ols
import matplotlib.pyplot as plt


# Load data
data = pd.read_csv('CSVfiles/Chocolate LCA - Main.csv', header=1)
data = data.loc[data['Boundaries / production phase ******'] == 'Cradle to farm gate',
                ['Agriculture type*', 'Country***', 'AD (kg Sb eq) .1', 'GW (kg CO2 eq) .1', 'ODP (kg CF11 eq).1', 'AC (kg SO2 eq).1', 'EU (kg PO4 eq).1', 'CED (MJ).1']].reset_index(drop=True)
data = data.replace(0, np.nan)
data.columns = ['AgricultureType', 'Country', 'AD', 'GW', 'ODP', 'AC', 'EU', 'CED']

def aggregated_results(data = data):
    #res = pd.DataFrame() # empty dataframe
    res = data.copy().drop('AgricultureType', axis=1)
    res.set_index('Country', inplace=True)
    row = pd.DataFrame(index = ['Order of magnitude'], columns = ['AD', 'GW', 'ODP', 'AC', 'EU', 'CED'])
    res = pd.concat([row, res])

    # mean for each indicator and each country
    for indicator in ['AD', 'GW', 'ODP', 'AC', 'EU', 'CED']:
        global_mean = data[indicator].mean()
        res.loc[:,indicator] = res.loc[:,indicator] - global_mean

        # calculate the order of magnitude of the indicator
        sign = -1 if global_mean < 1 else 1
        m = 10**(sign*math.ceil(abs(math.log10(abs(global_mean)))))
        res.loc['Order of magnitude', indicator] = m

    #aggregate results
    res = res/res.loc['Order of magnitude']
    if res.loc['Order of magnitude'].sum() == 6:
        res = res.sum(axis=1).drop('Order of magnitude')
        res = res.drop('Unknown')


        # Convert Series to DataFrame with index as first column
        res = res.reset_index()
        res.columns = ['Country', 'Aggregated_impact']

        # Plot
        ax = sns.boxplot(data = res, x = 'Country', y= 'Aggregated_impact', color='#99c2a2')
        ax = sns.swarmplot(data= res, x = 'Country', y= 'Aggregated_impact', color='#7d0013')
        plt.hlines(0, -0.5, 5.5, colors='red', linestyles='dashed')
        plt.rcParams['font.family'] = 'serif'
        plt.xlabel('')
        plt.savefig('Images/boxplot_aggregated_impact.png')

aggregated_results()
```

Listing 4: Python code for result extraction and statistical analysis of the collected data (2/2)

References

1. University of Barcelona. *The university* `https://web.ub.edu/en/the-university`.
2. University of Barcelona, agroecology department. *Current projects* `http://www.ub.edu/agroecologia/projectes/`.

3. *166th EAAE Seminar Sustainability in the Agri-Food Sector. Bridging research and policy: evidence based indicators on agricultural value chains to inform decision-makers on inclusiveness and sustainability* NUIG Galway (2018). https://agritrop.cirad.fr/590600/1/Full%20paper%20Dabat%20et%20al%20EAAE%20Paper%20August%202018.pdf.
4. Meinrenken, C. *et al.* Agroecology as a means to improve energy metabolism and economic management in smallholder cocoa farmers in the Ecuadorian Amazon. en. *Scientific Reports* **10,** 6184. https://www.nature.com/articles/s41598-020-62030-x (Apr. 2020).
5. Nur, T., Hidayatno, A., Setiawan, A. D. & Suzianti, A. Environmental Impact Analysis to Achieve Sustainability for Artisan Chocolate Products Supply Chain. en. *Sustainability* **15,** 13527. ISSN: 2071-1050. https://www.mdpi.com/2071-1050/15/18/13527 (2024) (Sept. 2023).
6. López-Del-Amo, B. & Akizu-Gardoki, O. Derived Environmental Impacts of Organic Fairtrade Cocoa (Peru) Compared to Its Conventional Equivalent (Ivory Coast) through Life-Cycle Assessment in the Basque Country. en. *Sustainability* **16,** 493. ISSN: 2071-1050. https://www.mdpi.com/2071-1050/16/2/493 (2024) (Jan. 2024).
7. Food and Agriculture Organization of the United Nations (FAO). *Zero-deforestation cocoa sweetens World Food Day* https://www.fao.org/gcf/news-and-events/news-detail/http-www.fao.org-climate-change-news-detail-en-c-1314699/en. 2020.
8. Caicedo-Vargas, C., Pérez-Neira, D., Abad-González, J. & Gallar, D. Agroecology as a means to improve energy metabolism and economic management in smallholder cocoa farmers in the Ecuadorian Amazon. en. *Sustainable Production and Consumption* **41,** 201–212. ISSN: 23525509. https://linkinghub.elsevier.com/retrieve/pii/S2352550923001926 (2024) (Oct. 2023).
9. Wikipedia. *Supply Chain Management* https://en.wikipedia.org/wiki/Supply_chain#Management.
10. Avadí, A. Environmental assessment of the Ecuadorian cocoa value chain with statistics-based LCA. en. *The International Journal of Life Cycle Assessment* **28,** 1495–1515. ISSN: 0948-3349, 1614-7502. https://link.springer.com/10.1007/s11367-023-02142-4 (2024) (Nov. 2023).
11. Pérez Neira, D. Energy sustainability of Ecuadorian cacao export and its contribution to climate change. A case study through product life cycle assessment. *Journal of Cleaner Production* **112,** 2560–2568. ISSN: 0959-6526. https://www.sciencedirect.com/science/article/pii/S0959652615016108 (2016).
12. Daymond, A., Giraldo-Mendez, D., Hadley, P. & Bastide, P. *Global Review of Cocoa Farming Systems* Report (2021). https://www.icco.org/wp-content/uploads/Global-Review-of-Cocoa-Farming-Systems_Final.pdf.
13. Leakey, R. Definition of agroforestry revisited. *Agroforestry Today* **1,** 5–7 (1996).
14. Santos, I. *et al.* NIR and MIR spectroscopy for quick detection of the adulteration of cocoa content in chocolates. *Food chemistry* **349,** 90–95 (2021).
15. Čuček, L., Klemeš, J. J. & Kravanja, Z. in *Assessing and Measuring Environmental Impact and Sustainability* (ed Klemeš, J. J.) 131–193 (Butterworth-Heinemann, Oxford, 2015). ISBN: 978-0-12-799968-5. https://www.sciencedirect.com/science/article/pii/B9780127999685000051.
16. Ntiamoah, A. & Afrane, G. Environmental impacts of cocoa production and processing in Ghana: life cycle assessment approach. *Journal of Cleaner Production* **16,** 1735–1740. ISSN: 0959-6526 (2008).
17. Rousseau, K. S. https://statinferentielle.fr/taille-dechantillon/.
18. Wikipedia. *Law of Large numbers* 2024. https://en.wikipedia.org/wiki/Law_of_large_numbers.