LCA database for Chocolate Production Worldwide

Preliminary results

Lisa PARUIT

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 1 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 1: Number of studies per country (a) or agriculture type (b) currently recorded in the database.

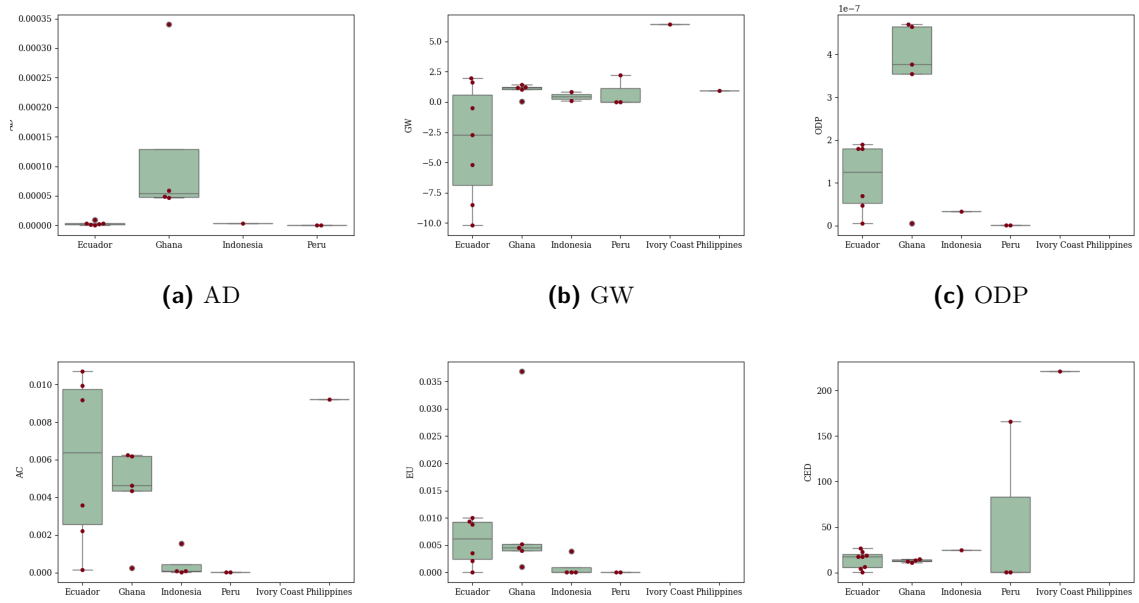
As the experimental plan for agricultural type is even less balanced than that of countries, we will only focus on the impact of the country on environmental impact potential indicators (EIP) in this *Preliminary results* section.

Statistical limitations for preliminary sample sizing

The number of studies per country and per agriculture type is not sufficient to have significant results. In additions 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

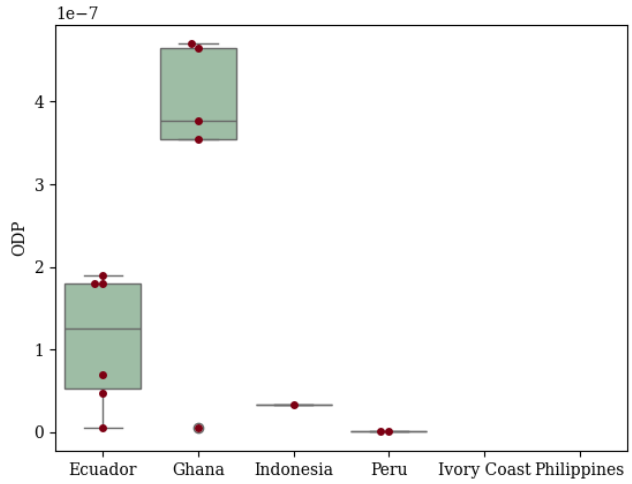
Figure 1: Boxplot of the environmental impact by country for AD (a), GW (b), ODP (c), AC (d), EU (e) and CED (f).

Figure 1 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 2 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 2: 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



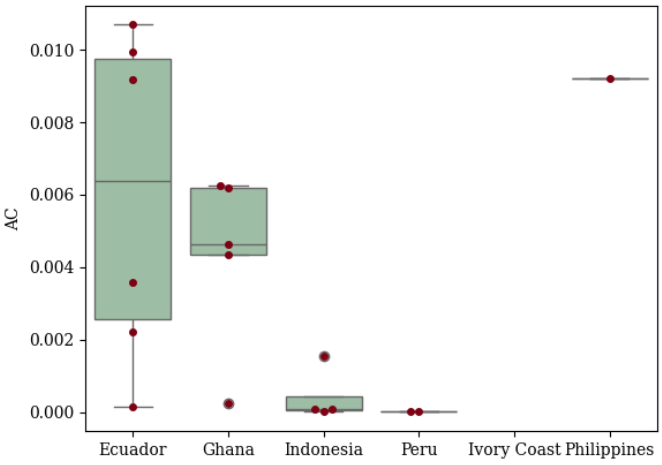
	p-value
Intercept	0.06710618701738431
Country[T.Ghana]	0.020459288019114303
Country[T.Indonesia]	0.5983357412200276
Country[T.Ivory Coast]	
Country[T.Peru]	0.3314359429092527
Country[T.Philippines]	

Table 3: Two-tailed p values for the t-stats of the parameters obtained from the OLS for ODP. *Intercept = Ecuador*

Figure 2: 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 [1]. 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 [2].
Even for Ghana, the p-values obtained when testing the coefficients for each country obtained through OLS regression are not satisfying (Table 3). 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 4: Two-tailed p values for the t-stats of the parameters obtained from the OLS for AC potential. *Intercept = Ecuador.*

Figure 3: Boxplot of the acidification potential by country.

On Table 2, 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 4 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 4.

Cumulative Energy Demand Potential

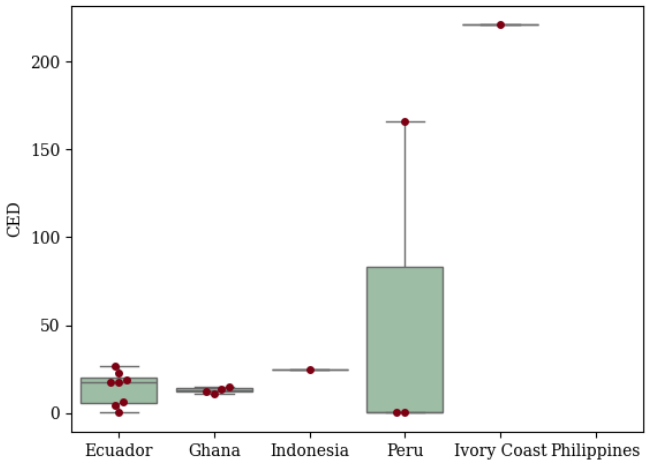


Figure 4: Boxplot of the cumulative energy demand potential by country.

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

Table 2 suggests that the variance between countries is the highest for cumulative energy demand. On Figure 4 and Table 5, 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

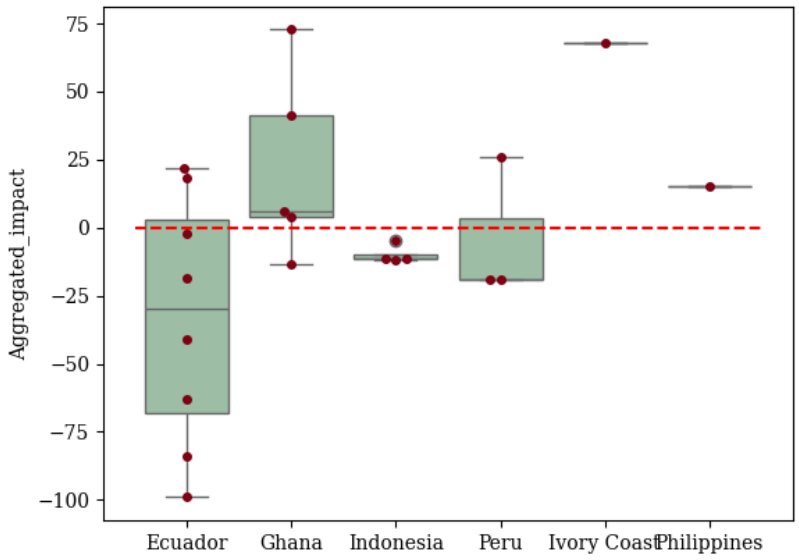


Figure 5: Addition of the relative gap to the mean of each indicator for each country.

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 [1], 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.

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 1). In order to assess the number of studies needed to get a satisfying confidence interval, sample size estimation methods exist [3]. 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 6.

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 6: 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:

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

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 [4]). 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 7, the knee point of the function (ie. the point of maximum curvature) is found using the kneedle algorithm (cf. Figure 6).

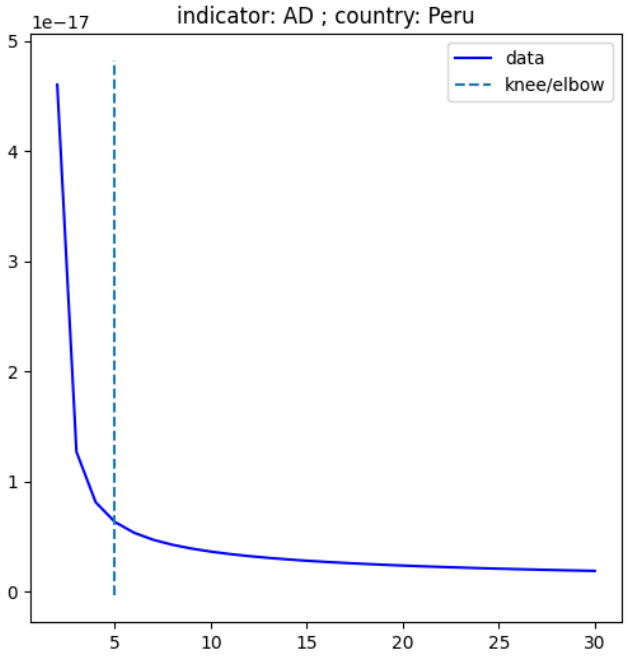


Figure 6: 1/2 CI / m vs. n Result of the kneedle algorithm for AD potential in Peru.

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 7: Coordinates for the curve of the 1/2 CI / m vs. n in Figure 6

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

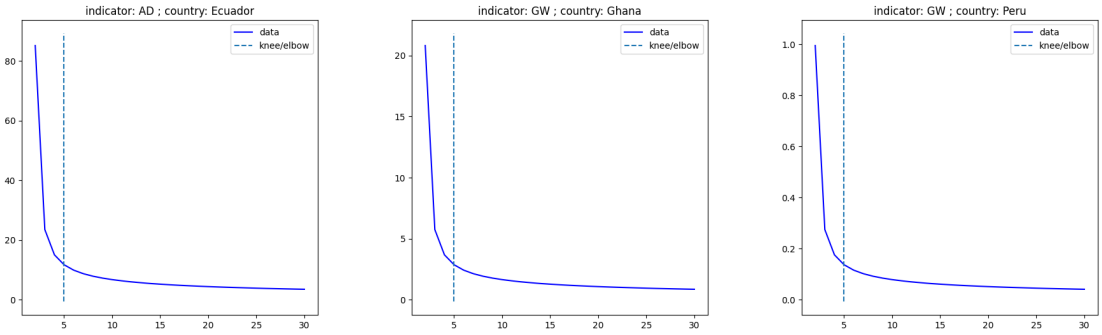
```python
import kneed as kn
import pandas as pd
import math as math
import numpy as np
import matplotlib.pyplot as plt


##### 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

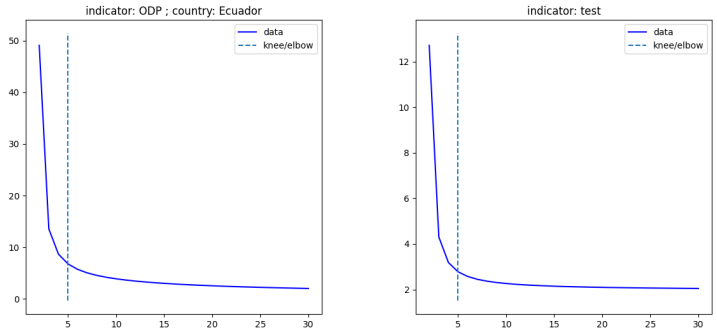
Limitations and hypothesis discarding

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 7: 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. 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.

2. 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).
3. Rousseau, K. S. `https://statinferentielle.fr/taille-dechantillon/`.
4. Wikipedia. *Law of Large numbers* 2024. `https://en.wikipedia.org/wiki/Law_of_large_numbers`.