A case study on Colombia: its environmental sustainable agriculture and a focus on water stress

Sofia Noemi Crobeddu
matricola: 2130389
Data Science's Master Degree - Sapienza University
The Role of International Organizations in producing Official Statistics

June 25, 2024

Abstract

This project examines the sustainability of Colombia's agriculture, with a focus on water stress. Agriculture an important part of Colombia's economy, the data analyzed want to give a quantification of the environmental aspect. This investigation was done through four proxy sub-indicators that revealed a critical overall direction. The focus on water stress, related also to the production of avocado, tries to give a prediction of this indicator in the future and in particular in 2030, the year of the SDGs Agenda.

1. First section

In the first part of the project I tried to investigate the agricultural sustainability of Colombia. This analysis was performed based on the second SDG of United Nations. The Sustainable Development Goals (SDGs) are a collection of 17 global goals created to achieve a better world and more sustainable universal conditions. They were adopted by the UN in 2015 and are part of the 2030 Agenda for Sustainable Development. In general they are interconnected between each other and deal with poverty, climate change, innovation, environmental, social and gender issues, inequalities, justice and peace.

I decided to focus on the second Goal: End hunger, achieve food security and improved nutrition, and promote sustainable agriculture and in particular on the target 2.4 regarding sustainable food production systems and agricultural practices. In general the subindicators derived from the indicator 2.4.1 of *Proportion of agricultural area under productive and sustainable agriculture*, cover three main dimensions: the Economic, Environmental and Social one. In this analysis I concentrated on the environmental aspect exploring it through some proxy sub-indicators. The decision of using proxy indicators was due to the availability and collection of data, first of all. This lack of data and information for the indicators is derived from the complexity in their calculation, but also from the few agricultural surveys done in the countries. Usually indicators are more specific metrics that provide direct information about a phenomenon, and they are designed to be as close as possible to the reality they want to measure. Proxy indicators, instead, are surrogate metrics that approximate indicators when direct data are difficult to obtain or are unavailable, and they do not measure the phenomenon directly, but are related to it and can still provide a good estimate. Moreover, in this specific case, the sub-indicators are measured at national level instead that at farm level as was supposed for indicators.

The proxy sub-indicators used for the investigation are the following:
- Fertilizer use intensity,
- Pesticide use intensity,
- Agriculture component of water stress,
- Greenhouse Gas Emissions Intensity.

1.1. Fertilizer use intensity

The Fertilizer use intensity $U_{c,t}$ measures the use of nitrogen fertilizer per area of cropland in a specific country c for a specific year t. It is defined as follow:

$$U_{c,t} = \frac{Fertilizer\ Use}{Cropland\ Area} = \frac{F_{c,t}}{Cropland\ Area_{c,t}} \quad (\frac{kg}{hectares})$$

where $F_{c,t}$ is the total agricultural use of nitrogen fertilizer in the country c and year t measured in *tons*, and $Cropland\ Area_{c,t}$ is the sum of arable land and permanent crops in the country c and year t measured in 1000 *hectares (ha)*.

1.2. Pesticide use intensity

Similarly to the previous one, the Pesticide use intensity $U_{c,t}$ measures the use of pesticides per area of cropland in a specific country c for a specific year t. It is defined as follow:

$$U_{c,t} = \frac{Pesticide\ Use}{Cropland\ Area} = \frac{P_{c,t}}{Cropland\ Area_{c,t}} \quad (\frac{kg}{hectares})$$

where $P_{c,t}$ is the total agricultural use of pesticides in the country c and year t measured in *tons*, and $Cropland\ Area_{c,t}$ is the sum of arable land and permanent crops in the country c and year t measured in 1000 *hectares (ha)*.

1.3. Agriculture component of water stress

This proxy sub-indicator refers to the level of pressure that agricultural activities place on water resources. It is defined as follow:

$$Agriculture\ component\ of\ Water\ Stress = \frac{TFWW_A}{(TRWR - EFR)} * 100 \quad (\%)$$

where $TFWW_A$ indicates the total freshwater withdrawn for agriculture and is measured in $km^3/year$, $TRWR$ identifies the total renewable freshwater resources in $km^3/year$, and finally EFR is the environmental flow requirement always measured in $km^3/year$.

This sub-indicator is a disaggregation of the SDG 6.4.2, the Level of Water Stress which comprehends also the industrial component and the service-domestic part. For this reason its interpretation is based on the indicator of water stress and on the general consideration that agricultural is approximately respondable for the 70% of water withdrawals. No stress (safe) is considered for values under 17.5%, instead values over 17.5% are considered risky and potentially problematic.

1.4. Greenhouse Gas Emissions Intensity

The last sub-indicator, Greenhouse Gas (GHG) Emissions Intensity, measures the amount of greenhouse gases emitted per unit of economic and agricultural activity. It provides a measure of how efficiently a country or a sector uses resources in relation to its emissions output. It is defined as follows:

$$GHGE\ Intensity = \frac{Farm\ gate\ emissions}{Value\ of\ Agricultural\ Production} \quad (kgCO2\ per\ constant\ 2014\text{-}2016\ USD)$$

where the *Farm gate emissions* represent the ones from drained organic soils, cultivation of histosols, a certain number of inorganic fertilizers, crop residues, manure deposited on pasture, range and paddock, manure applied to soils, manure management, enteric fermentation, prescribed burning of savanna, burning crop residues, rice cultivation, and on-farm energy use. It is measured in kg of CO_2. The *Value of Agricultural Production* represent instead the part of gross production referred to farm gate. It is measured in constant 2014-2016 USD. Regarding the interpretation of this sub-indicator, lower emissions intensity indicates a greater efficiency in resource usage and a reduction in the environmental impact (positive for climate change).

1.5. Results from the proxy sub-indicators

The data for the first three were directly taken from FAOSTAT databases, instead for the last one (GHGE Intensity) we don't have available values direct calculated, but we have data regarding the components, so, applying the formula, it is possible to reach the results.

1.5.1. FUI outputs

The availability of data was in the range of years 1961-2021 and, as shown in Figure 1.1, we have an increasing value of the intensity of fertilizers until 2015 and then in 2019 the function starts to increase again but slowly. A reason for this change in the trend of the

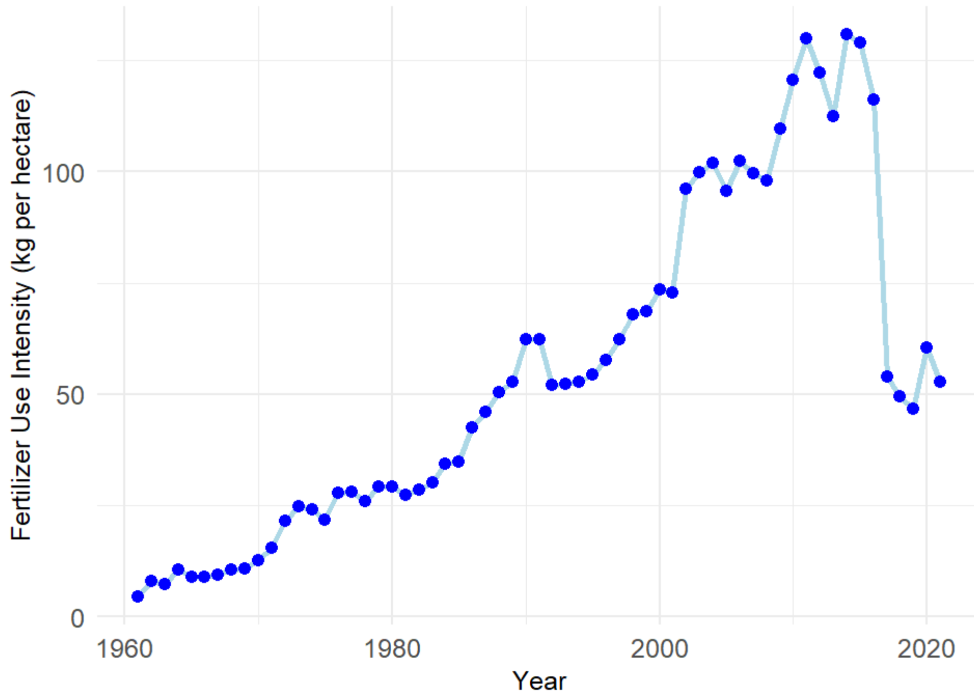

Figure 1.1: Colombia's Fertilizer Use Intensity

function can be related to the increase of the import price in the same period, as shown in Figure A.1 (Appendix) from the website $IndexBox.io$.

The data from FAOSTAT that I analyzed show also that the maximum value of FUI is registered in 2014 ($130.81\frac{kg}{ha}$), instead the minimum one in 1962. The last value registered in 2021 was $52.86\frac{kg}{ha}$, hence we have a decreasing trend in the last years.

1.5.2. PUI outputs

In this case the data available were less compared to FUI presented in the previous paragraph, infact it was possible to analyze just the years from 1990 to 2021. As shown in Figure 1.2, the function increases until 2005 and then it starts to decrease until 2008. After this year there is not a pick as in 2005, even if in 2014 and in 2019 we have quite high value compared to the others. The peak in 2005, with pesticide use reaching $32.63\frac{kg}{ha}$, can be attributed to a combination of factors, as explained in in the article [MMH22]. Firstly, government policies aimed at reducing illicit crops such as coca encouraged growers to transition to legal and more intensive agricultural crops like fruits, vegetables, and flowers, which require increased pesticide use to maintain high production standards. Secondly, the surge in agricultural exports demanded high-quality, pest-free products, driving up the demand for pesticides. Lastly, general developments in the agricultural sector, including infrastructure improvements and agricultural practices, facilitated an intensification in pesticide usage to maximize productivity and global market competitiveness. The data analyzed indicate also that the minimum value of $3.10\frac{kg}{ha}$ was registered in 1993. The last value we have is of $8.74\frac{kg}{ha}$ in 2021, then we have again a decreasing trend.

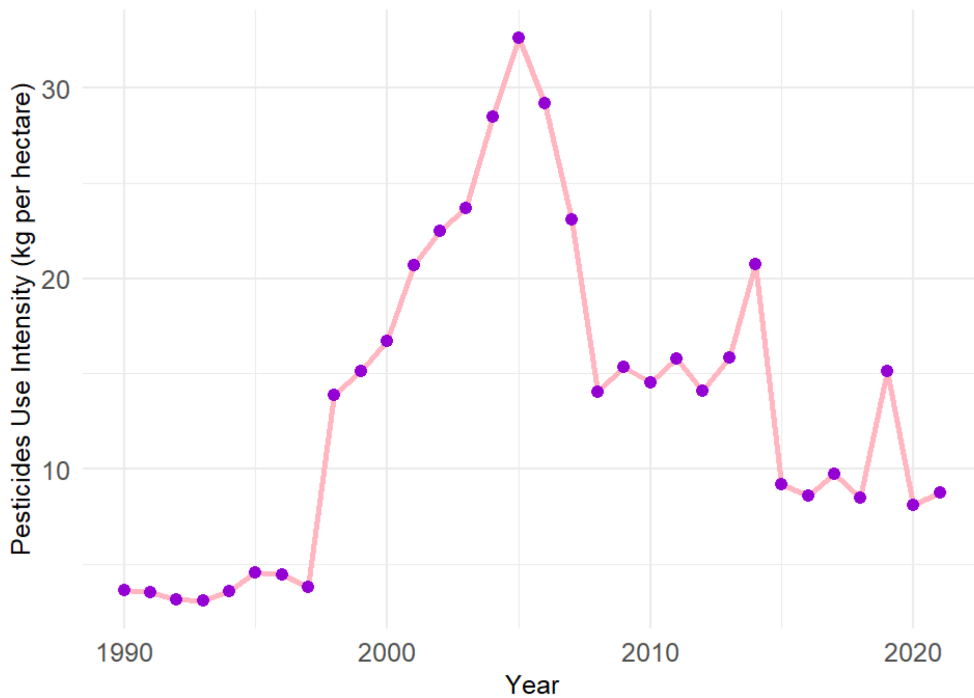

Figure 1.2: Colombia's Pesticides Use Intensity

1.5.3. Agriculture component of WS outputs

For this sub-indicator I selected the values addressed to agriculture from the indicator of Level of Water Stress. In this case it was possible to analyze data from 2000 until 2021. As it is shown in Figure 1.3, the function is strictly increasing. The maximum value of 3.8% is registered in the last available year, 2021, instead the minimum one of 0.74% is registered in 2000, the first available year. Even if the slope of the function is high from 2008 until 2016 and takes lower values after 2016, suggesting that the agricultural component of Water Stress increases slower, these data seem really worrying since the trend is still not changing among the years.

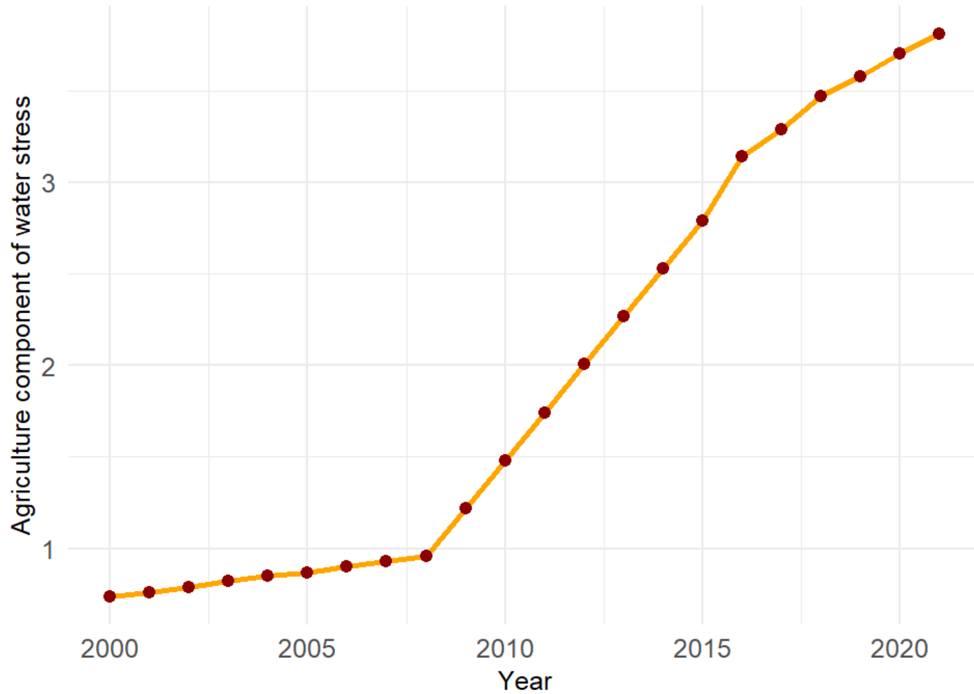

Figure 1.3: Colombia's Level of Water Stress (Agriculture Sector)

1.5.4. GHGE intensity

For this last result, as mentioned before, we don't have the direct values calculated for this specific proxy sub-indicator. To obtain it, it was made the ratio between the Farm Gate Emissions and the Value of Agricultural Production. The first ones were measured in kiloton (kt), and then converted in kilogram for the final calculation. The second one was measured in 1000 USD and after converted in USD.

The range of years considered is 1961-2021, and as it is shown in Figure 1.4, there is a pick in 1990, in fact the minimum value registered is in 1989. The maximum value of 14.39 *kg CO2 per USD* is instead registered in 1994. Considering that the last value available of 2021 is 9.91 *kg CO2 per USD*, the intensity of Greenhouse Gas Emissions tends to decrease. The pick is justified by the increase in Farm gate emissions between 1989 and 1990, as shown in Figure A.2, also derived by the increasing use of fossil fuels since the industrial revolution.

1.6. Assessing progress towards indicators

Since the indicators are inserted in the 2030 Agenda that defines 17 universal goals as explained in the First section, it is necessary to evaluate their progress to understand how

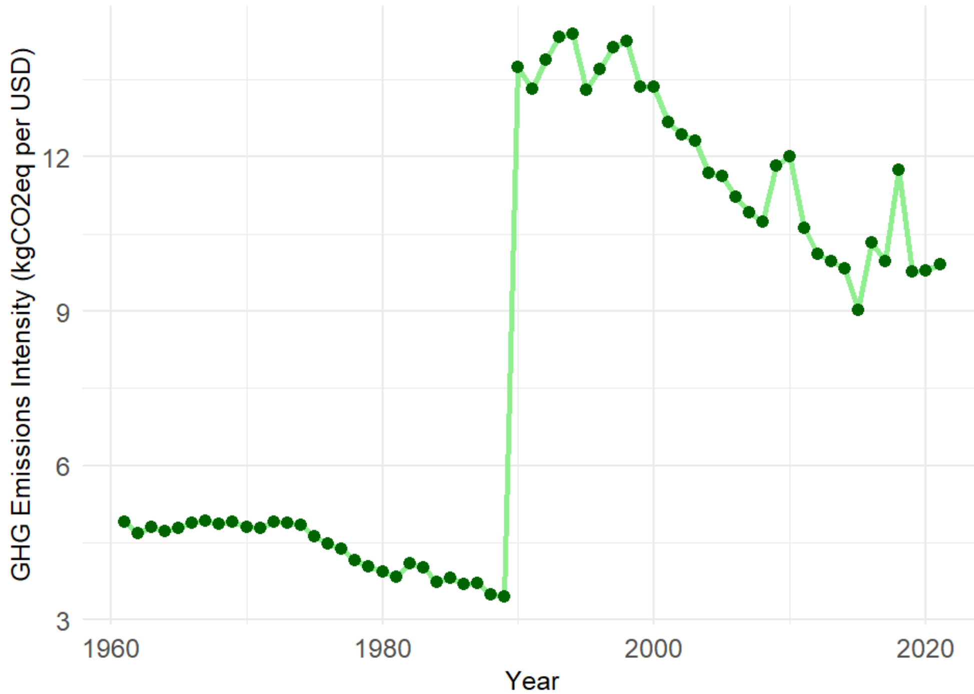

Figure 1.4: Colombia's Greenhouse Gas Emissions Intensity

much far from the final purpose we are. Some of the proxy sub-indicators have also a specific target, as for the rural poverty rate (3%). There are principally two method to evaluate the progress: the Status assessment and the Trend assessment. In presence of a specified target, the Status assessment is the distance to the target and the Trend assessment is the ratio of observed versus required compound annual growth rate (CR). In the case without a precise target, the Status assessment is the quantile distribution and the Trend assessment is measured as the observed compound annual growth rate (CAGR) compared to the normative trend.

The criteria to evaluate the Status assessment is shown in Table 1.1 where the categories 5 and 4 correspond to desirable level (green one), the category 3 to the acceptable level (yellow one) and finally the categories 1 and 2 to the unsustainable level (red one).

Table 1.1: Criteria for Status assessment

Bounds	**Category**
$q_{80} < x \leq Max$	5
$q_{60} < x \leq q_{80}$	4
$q_{40} < x \leq q_{60}$	3
$q_{20} < x \leq q_{40}$	2
$Min \leq x \leq q_{20}$	1

It is important to specify that for all the proxy sub-indicator, since the normative trends is such that to a higher value corresponds a lower sustainability (less desirable category) and the allocation of the categories has to reflect this trend, the values of the category, after the calculation, were inverted.
The results are shown in Figure 1.5, and as we can see the worst one is the Agricultural component of Water stress. In general the results don't show desirable level reached in any of the proxy sub-indicators for Colombia's environmental dimension except for the pesticide usage, then there is still a lot to be done.

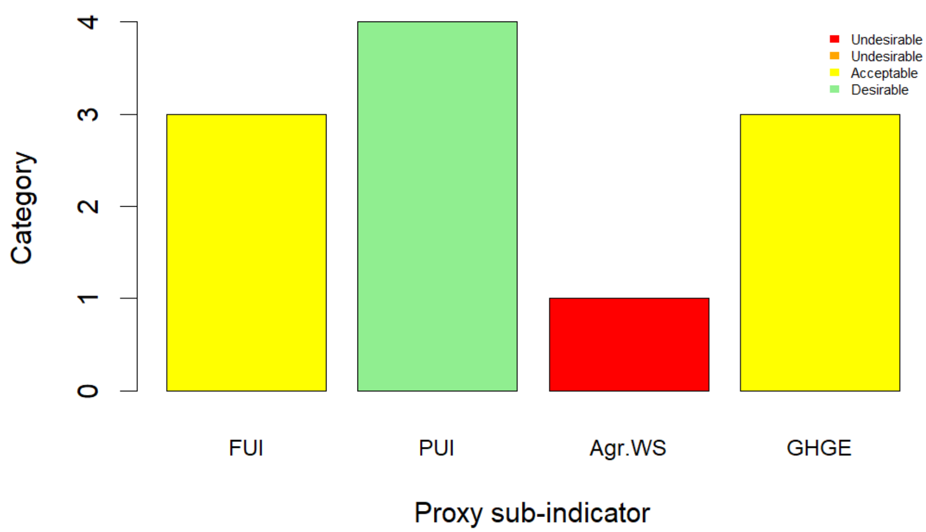

Figure 1.5: Colombia's Status Assessment

Regarding the Trend assessment, as mentioned before, it was calculated through CAGR. The criteria to interpret it is shown in Table 1.2. In particular, having as a point of reference the categorization used of Status assessment, the categories Improvement and Slight or no Improvement correspond to the the desirable category (green color), Slight deterioration defines the acceptable category (orange color) and finally Deterioration represents the undesirable category (red color). As for the calculation of the Status assessment, the values for the sub-indicators were inverted (for the same reason explained before). The results are shown in Figure 1.6, and as we can see the better ones are Fertilizer Use and Pesticide Use Intensity which have reached desirable levels, but the Agricultural component of WS and the Greenhouse Gas Emissions Intensity belongs to the category 'Deterioration'. The difference between these two is that the last one (GHGE) is increasing with a lower trend compared to Agr. WS. This output reflects the results obtained and exhibited in the First section.

Table 1.2: Criteria for Trend assessment

Bounds	**Symbol**	**Category**
$CAGR > 0.001$	$>>$	Improvement
$-0.0005 \leq CAGR \leq 0.001$	$\geq$	Slight or no Improvement
$-0.001 \leq CAGR \leq -0.0005$	$<$	Slight deterioration
$CAGR < -0.001$	$<<$	Deterioration

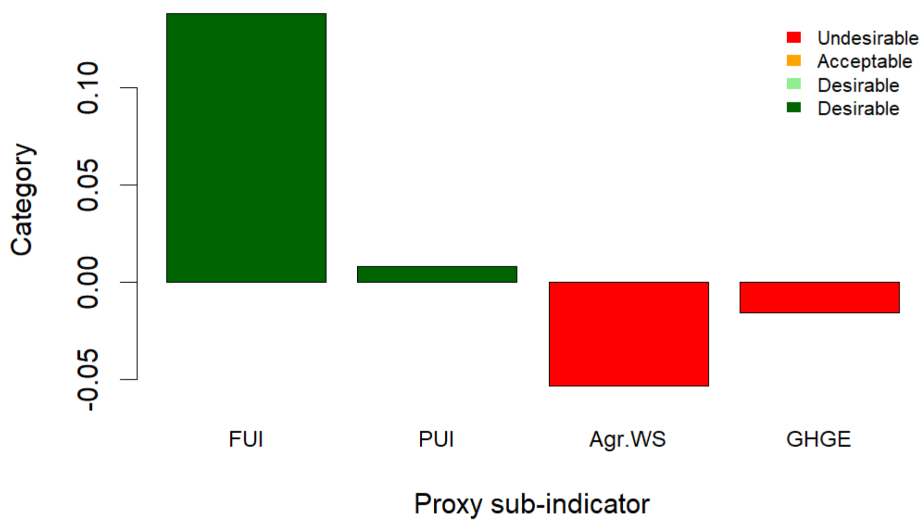

Figure 1.6: Colombia's Trend Assessment

1.7. Index of agricultural sustainability

To create the composite index of agricultural sustainability from the environmental point of view, data from the years 2000-2021 were selected, since it was the only range in common between all the proxy sub-indicators of Fertiliser Use Intensity, Pesticide Use Intensity, Agricultural component of Water Stress and Greenhouse Gas Emission Intensity, using the same data analyzed in the section 1.5. After creating a dataset with them, the data were normalized to be comparable through the min-max method. The next step, before the calculation of the final index, was to identify the weights assigned through the CAGR obtained in the previous section 1.6. The final index was constructed through the following formula:

$$I_{[2000,2021]} = Index\ Agricultural\ Sustainability = \sum_{i=1}^{n} w_i \cdot Norm(x_i)$$

where:

- n is the number of proxy sub-indicators used,
- w_i are the weights assigned to each sub-indicator,
- x_i represents the value of the i^{th} sub-indicator,
- $Norm(x_i)$ is the normalization of the value of the i^{th} sub-indicator.

The results obtained are shown in the Table A.1 in the Appendix. The index reflects a measure of the sustainability of agriculture from an environmental point of view in Colombia. In general, we have negative values from 2017 that indicates a deterioration in sustainability. The worst value is registered in 2018, instead the best one in 2011. The positive values before 2017 are really close to zero for the majority, so this suggests a slight stability. From the last years it is evident that environmental conditions in agriculture are getting worst and this is due to the increase of water stress and emissions.

2. Second section

In this section I performed an investigation on the 6.4.2 SDG Indicator, the Level of Water Stress, since based on the previous analysis, it resulted to be a very potentially problematic level for Colombia. In general, Water stress occurs when the demand for water exceeds the available supply during a certain period or when poor quality restricts its use. Its causes can be found in the agricultural sector and industrial sector, in particular, but also the household one contributes. The definition of this indicator given by FAO is the pressure level exerted by all economic sectors on a certain country's renewable freshwater resources ([Foo21]). Its formula is in fact the generalization of the formula presented in section 1.3 for the Agriculture component of water stress and it is the following:

$$Level\ of\ Water\ Stress = \frac{TFWW}{(TRWR - EFR)} * 100 \quad (\%)$$

where, in this case, $TFWW$ indicates the total freshwater withdrawn and is measured in $km^3/year$. As for the agricultural component, $TRWR$ identifies the total renewable freshwater resources in $km^3/year$, and finally EFR is the environmental flow requirement always measured in $km^3/year$.

This indicator is really important since it quantifies how much the natural freshwater resources are already being utilised and it is a helpful indicator for policy analysis for each country. High levels of water stress can have negative effects on the sustainability of natural resources and economic development, instead low levels of stress indicate that water is not a serious problem for economic development and sustainability.

2.1. Water stress and avocado's production

Remaining in the agricultural sector, one of the products that most deal with the problem of water stress is avocado. According to the World Economic Forum, one hectare of avocado with 156 trees consumes 1.6 times more compared to a forest with 677 trees per hectare ([Aya20]). Also, according to the famous international startup Greenly, the production of a single avocado needs 320 liters of water, an high value if we compare it, for example, to the production of almond milk that takes an amount of water four times smaller ([Saf23]).

This agricultural process has in general many sustainable consequences, as the excess of greenhouse gas emission and excessive carbon emissions with a negative impact on global warming. Another important consequence is also the deforestation, and Colombia is one of the most evident victim of this system. For the massive production of avocado the principal responsible are European Union and the United States, the main buyers and producers. As it is shown in Figure 2.1 from Eurostat Databases, the imports of avocado increased a lot during past years (apart from 2022 that is a little bit slower compared to 2021 and 2023), and the trend indicates that is going to grow more. According to the study conducted by the non-governmental organization Mani Tese on the impact of avocado production in Colombia ([Man23]), by 2030 the EU and the United States are expected to remain the main buyers, with 31% and 40% of overall imports.

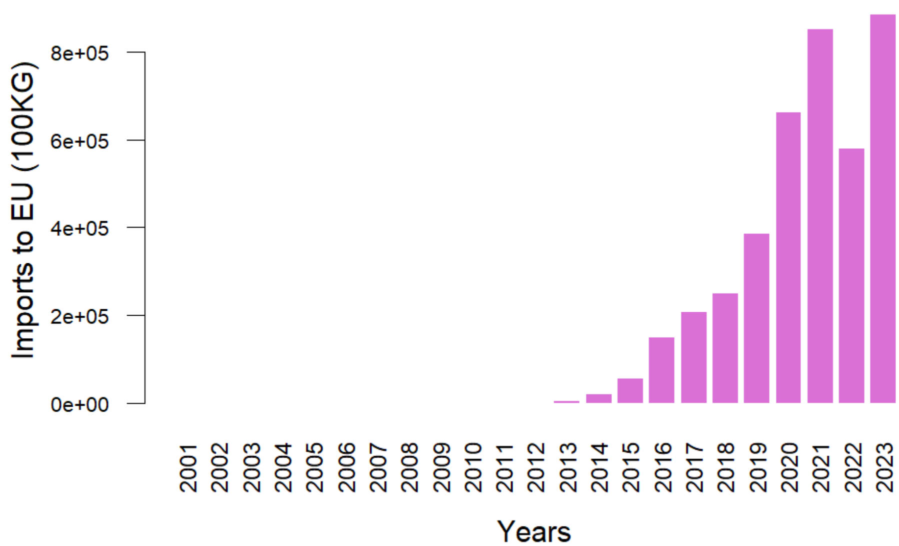

Figure 2.1: Avocado's Imports of EU from Colombia, Source: Eurostat

Analyzing also FAOSTAT data, through a linear regression model I tried to catch the relation between Water Stress Level and the quantities of import and export of avocado in and from Colombia, considering the world trade in general. As we can see from Figure 2.2 to have a preliminary idea, in Colombia the Water Stress Level increases as the variable of export quantity of avocados, instead in the same period the import quantity variable decreases. This intuition was confirmed by the results of the linear regression model that exposed a positive relation between Water stress level and Export Quantity, and a negative relation between Water stress level and Import Quantity (Figure A.3 in Appendix). The Figures 2.3 and 2.4 are plots of the linear regression line with the variable of Water Stress Level as the dependent one and the variables of Import quantity and Export Quantity as the independent ones.

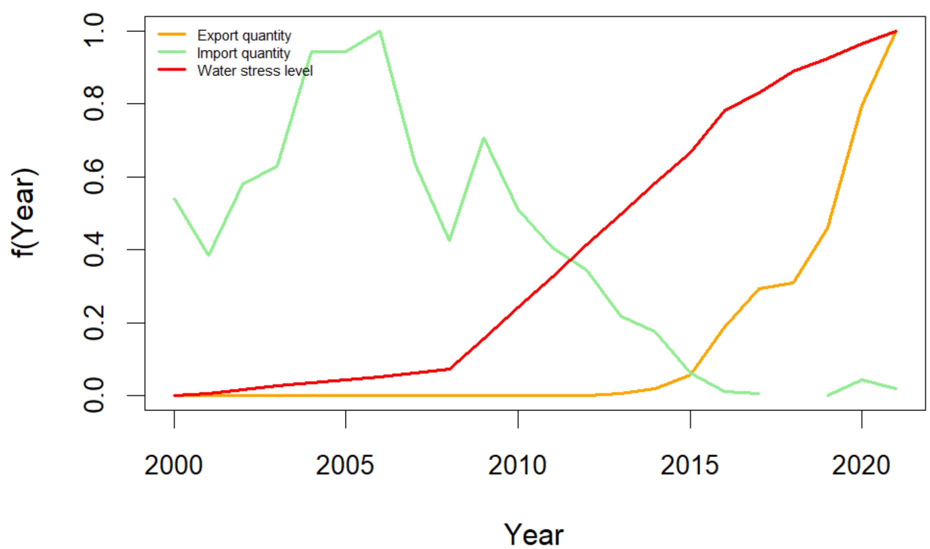

Figure 2.2: Avocado's Imports-Exports from Colombia and Water Stress Level

2.2. Prediction of Water Stress Level

After the considerations and the analysis described in the previous subsection, I decided to perform a prevision in order to obtain possible future values of Water Stress Level. The methodologies used are JAGS (Just Another Gibbs Sampler) in R and supervised machine learning through a Neural Network of Keras in python. JAGS is a software for Gibbs sampling that adopts a Bayesian model approach using Monte Carlo Markov Chain (MCMC) sampling. Explaining more deeply, Gibbs sampling allows sampling from conditional distributions, simplifying the calculation of posterior distributions of complex Bayesian models.

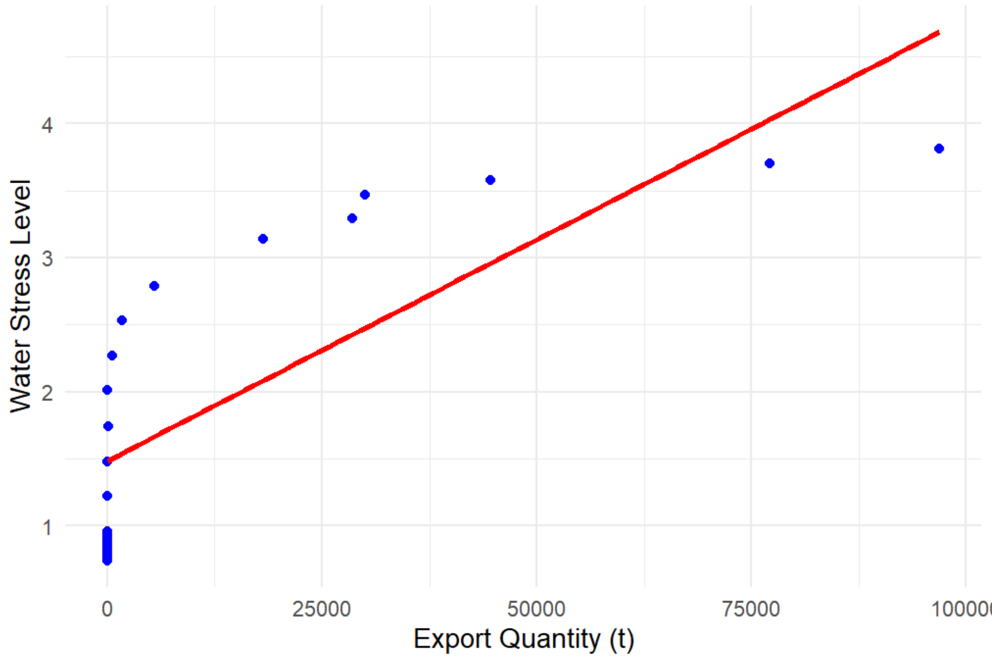

Figure 2.3: Relation between Water Stress and Export Quantity in Colombia

Regarding the second methodology, Keras is a deep learning library of python designed to create and train Neural Networks (artificial, convolutional and recurrent) and other deep learning models.

The data used are taken from the World Bank database for the range period 2000-2020, and with relevant features that potentially impact water security. The variables considered for this analysis are the following:
- Year,
- Annual freshwater withdrawals, agriculture (% of total freshwater withdrawal),
- Annual freshwater withdrawals, domestic (% of total freshwater withdrawal),
- Annual freshwater withdrawals, industry (% of total freshwater withdrawal),
- Annual freshwater withdrawals, total (billion cubic meters),
- Level of water stress: freshwater withdrawal as a proportion of available freshwater resources,
- Renewable internal freshwater resources, total (billion cubic meters),
- Average precipitation in depth (mm per year),
- Total Population,
- GDP per capita (constant 2015 US$),
- Population density (people per sq. km of land area).

The first variable Year was excluded, the dataset was cleaned making sure that there were

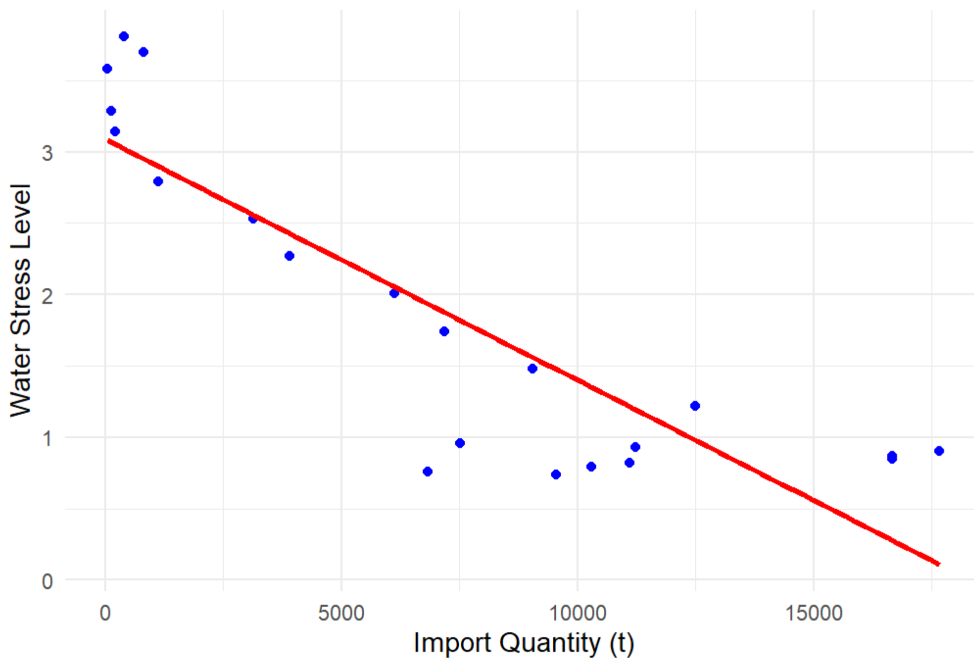

Figure 2.4: Relation between Water Stress and Import Quantity in Colombia

not NA values and the columns were renamed to simplify the interpretation of the code.

2.2.1. Prediction through JAGS

The Bayesian model was configured in order to estimate the level of water stress (variable Y) based on the selected predictor variables previous mentioned (variables X_i). The number of simulations for the MCMC was set to 20000 and the number of "burn-in" simulations (the number of initial iterations that are discarded before starting to save MCMC sampling results to reach the convergence station reducing the impact of parameter initialisation on final estimates) to 2000. The variable Y is distributed normally and its mean is represented by a linear combination of *beta* coefficients multiplied by the respective predictor variables. Similarly, *beta* coefficients are treated as normal standard variables, instead the variable of precision *tau* is interpreted through a gamma distribution. The chose of using a normal distribution was driven by the fact that the relationships between variables were complex and not perfectly known.

Regarding the expectations of the *beta* coefficients, the first three of the three dimensions (Agriculture, Industry and Domestic) of the annual freshwater withdrawals, gave negative results but very close to 0, and the agricultural and the domestic one seem to have more impact on the Water Stress Level. The variable of annual freshwater withdrawals considering the total measure, is still quite similar to the previous ones, in fact the mean of its coefficient is positive and close to 0. The coefficients $beta_5$, $beta_6$ and $beta_7$ of Renewable internal freshwater resources, Average precipitation and Total Population as respective predictors, resulted to have a positive mean approximate to 0, so this suggests an almost null effect of Y variable. $beta_8$ is similar to the last ones since it is very close to 0, but it has a negative mean. Finally, $beta_9$ has a negative mean that is quite close to 0, but its effect is bigger than the previous four coefficients. The graph of coefficients' values within the iterations is shown in the Appendix (see Figure A.4).

After this preliminary analysis, I performed the simulation of data for Water Stress Level through the JAGS model with 10000 iterations and a seed equal to 134 (arbitrary). In the output the predictions gave a minimum of -0.44485 and a maximum of 0.48008 for the variable Y, that transformed in percentage are respectively -44.49% and 48.01%. The mean is instead 0.02847 (2.85%) and the median 0.02799 (2.79%). These values suggest that the Water of Stress Level in this case has a central tendency (as shown in Figure A.5) derived, for the majority, by the use of the normal distribution. Moreover, the confidence interval at level $1 - \alpha = 0.95$ resulted in $[-0.202, 0.259]$ that confirms what explained just before.

In general, from these predictions it is difficult to have precise values of Y for the future, and probably, to eliminate this centrality in outputs, it could be interesting to make predictions using other distributions beside the normal one, or instead collecting more data for a bigger range of years. Anyway, these data show an important result: the distribution of Water Stress Level tends to be more rightward shifted in positive values. Since for this indicator higher values suggest a deterioration in the stress of the water, this confirms the analysis performed in the First section where we saw an increasing trend of this Level. Moreover, the confidence interval indicates a potential future level that can reach the 25.9% (a very high level).

2.2.2. Prediction through a Neural Network and Extrapolation

The dataset used for this second prediction is the same of the section 2.2.1. The values of the variable Level of Water Stress were converted in relative values from the original ones in percentages. In this case the Y variable gave high correlation results with the agricultural and domestic component of Freshwater withdrawals (positive and negative respectively), same with the total one and also with the variables Population total, GDP per capita and Population density (positive for all). After this control, the data, divided in X (independent variables) and Y (dependent variable, the Water Stress Level), were normalized using always the min-max method and finally transformed into an array structure. After this preliminary management, the Neural Network model was built using TensorFlow. The model was trained on an 80% training set and the performance of the model was evaluated on a separate 20% test set.

The results of the Neural Network's prediction are shown in Figure 2.5. The red line represents the perfect coincidence between the test data and the predicted data of Y. As we can see the prediction is quite good, and this impression is confirmed by the measures of the accuracy of the model: the mean squared error is around 0.0014, the mean absolute error around 0.0326 and the R^2 score around 0.9840. In particular, low values of MSE and MAE indicate a good performance of the model, instead the value of R^2 indicates that the model explains the 98.4% of the variability of data test.

The next step was about the prediction through a validation set, useful to see how the model is trained with future values (not observed). Since the SGDs are scheduled for 2030, it was meaningful to try to predict future data for this specific year through the validation dataset, with future values and estimation of the variables. The estimation of the independent variables for 2030 was performed through the technique of extrapolation, that is the process of projection of data beyond the years observed. To extrapolate the variables, the linear regression approach served to identify the tendency and the value of Y for the year 2030 was obtained. In particular, the value of Water Stress Level in 2030 retrieved is 0.2843, that is 28.43%. This result seems similar to the one obtained from the simulation through the JAGS model. Unfortunately, since the original dataset is really small, to be more realistic and to intensify the accuracy, the model should be trained on a larger one.

As a final analysis, it was performed a prediction for the range of years 2020-2035 using the extrapolation approach explained before. The results achieved are shown in Figure 2.6, where we can see an increasing trend of this indicator that in 2035 will exceed the 7% and the value for 2030 is 6.36%.

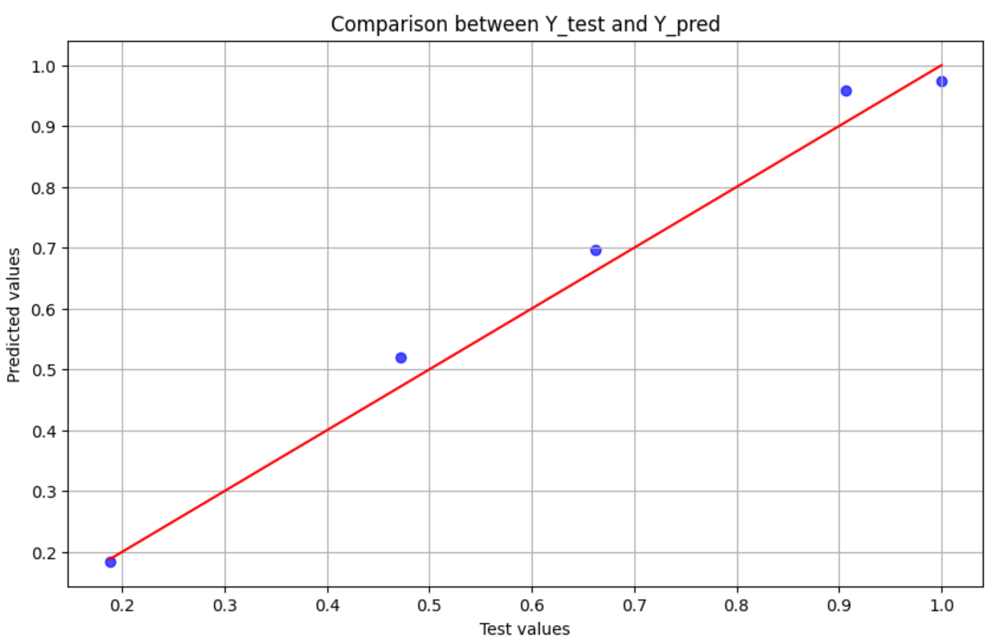

Figure 2.5: Predictions of the Neural Network

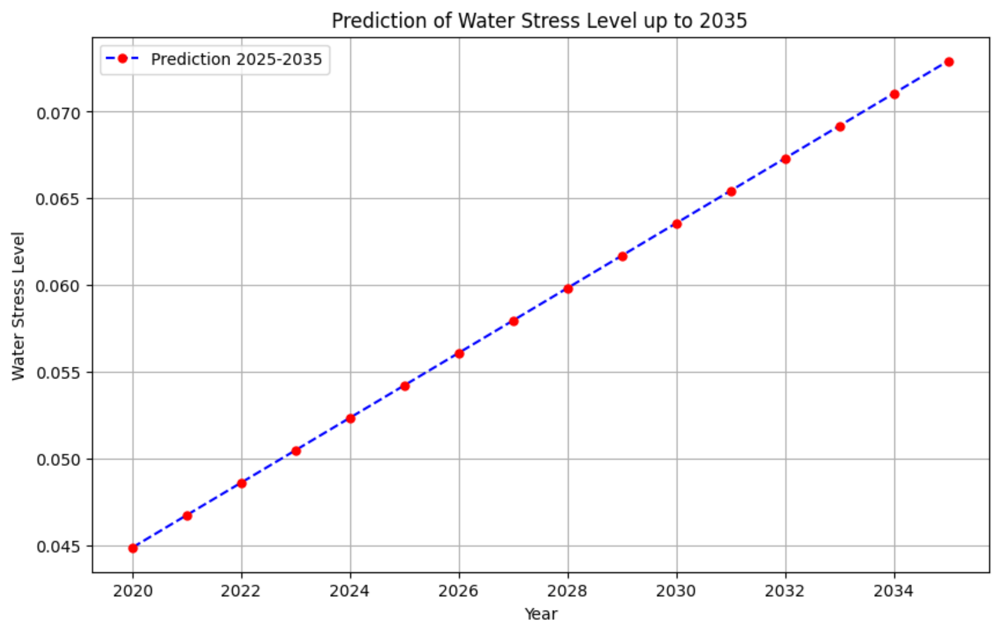

Figure 2.6: Predictions through extrapolation for the years 2020-2035

3. Conclusions

From this analysis we can conclude that Colombia has an overall negative trend in sustainable agriculture, as the index calculated shown. More specifically, the usage of fertilizers and pesticides is decreasing, same as the greenhouse gas emissions, even if the farm's gate gas emissions are increasing and in particular from 1990s. The most worried indicator is the water stress that resulted to have a strictly increasing trend from a global point of view and from the agricultural component.

The predictions of this last indicator mentioned through models seem to be not really useful in quantifying a specific value for the year 2030, since both JAGS model and the Neural Network model return a value of 26-28% (a very high value compared to the last one registered of 4.36%). This is probably due to the small amount of data we have, so it could be interesting to investigate new models taking into account the monthly value of Waters Stress Level (if available) or considering a list of countries in order to train the model on a larger amount of data (in particular for the Neural Network model). Moreover, from the analysis through JAGS, it resulted that the variables that most affect Water Stress Level are the Total Annual freshwater withdrawals and the Population Density.
The extrapolation approach that uses linear regression, seems instead the most adaptable to these data, since it returns a value over the 6% for the year 2030. This value is potentially problematic and in general confirms the increasing trend explored with the previous methods.

As mentioned, one of the product that most influence this indicator, is avocado. Data regarding this specific relations are difficult to catch, and this projects wants to be a starting point to stimulate the deepening of this case study. Colombia is really affected by its production as it was explored in this project, in fact a glaring proof is given by the department of Quindio in the centre of the Colombian Andean mountain range, where entire areas of the forest have been deforested to leave space for avocado crops. This type of ecosystems with wealthy biodiversity, have an important function of water reservoir and carbon sink to mitigate global warming. Furthermore, the capitalistic purpose behind this situation, has significant social impact. Multinationals with a high purchasing power that buy these lands, force local people to leave their place, and in the meantime, also due to the thousands of tourists that come there, the life cost increases a lot. One of the place that most attracts tourists and suffered from deforestation is the area with wax palms (unique all over the world since they grow at an altitude major than 2000 m).

A. Appendix

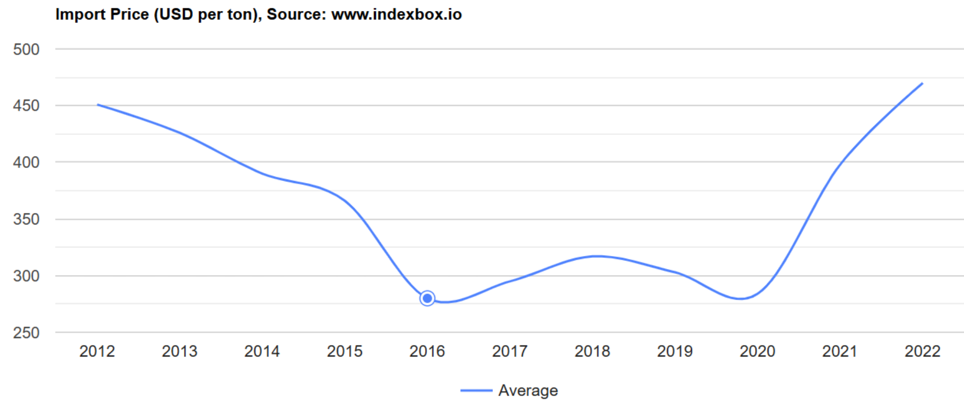

Figure A.1: Colombia's Import Price Fertilizer (Graph from *Indexbox.io*)

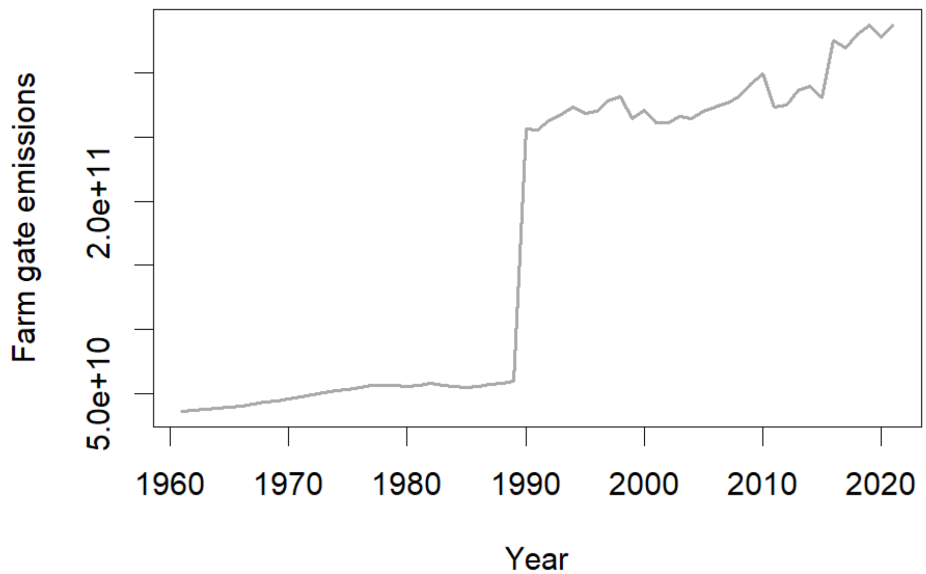

Figure A.2: Colombia's Farm gate emissions

Table A.1: Composite index of agricultural sustainability of Colombia

Year	**Index**
2000	0.03108738
2001	0.03350808
2002	0.07253525
2003	0.07911361
2004	0.08610255
2005	0.07684540
2006	0.08778861
2007	0.08168966
2008	0.07600296
2009	0.08718446
2010	0.09981381
2011	0.11595294
2012	0.09999304
2013	0.08056302
2014	0.10836140
2015	0.09984010
2016	0.06785365
2017	-0.03566313
2018	-0.05260097
2019	-0.04965287
2020	-0.03182162
2021	-0.04636018

```
> summary(mod)

Call:
lm(formula = water.df$Value ~ quantity.df$`Import Quantity` +
    quantity.df$`Export Quantity`)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9389 -0.2949  0.2031  0.3053  0.5282

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    2.536e+00  2.118e-01  11.975 5.22e-10
quantity.df$`Import Quantity` -1.225e-04  2.073e-05  -5.912 1.35e-05
quantity.df$`Export Quantity`  1.669e-05  4.388e-06   3.803   0.0013

(Intercept)                   ***
quantity.df$`Import Quantity` ***
quantity.df$`Export Quantity` **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4345 on 18 degrees of freedom
  (1 osservazione eliminata a causa di un valore mancante)
Multiple R-squared:  0.8651,    Adjusted R-squared:  0.8501
F-statistic:  57.7 on 2 and 18 DF,  p-value: 1.482e-08
```

Figure A.3: Linear Regression Model output for import-export and Water Stress variables

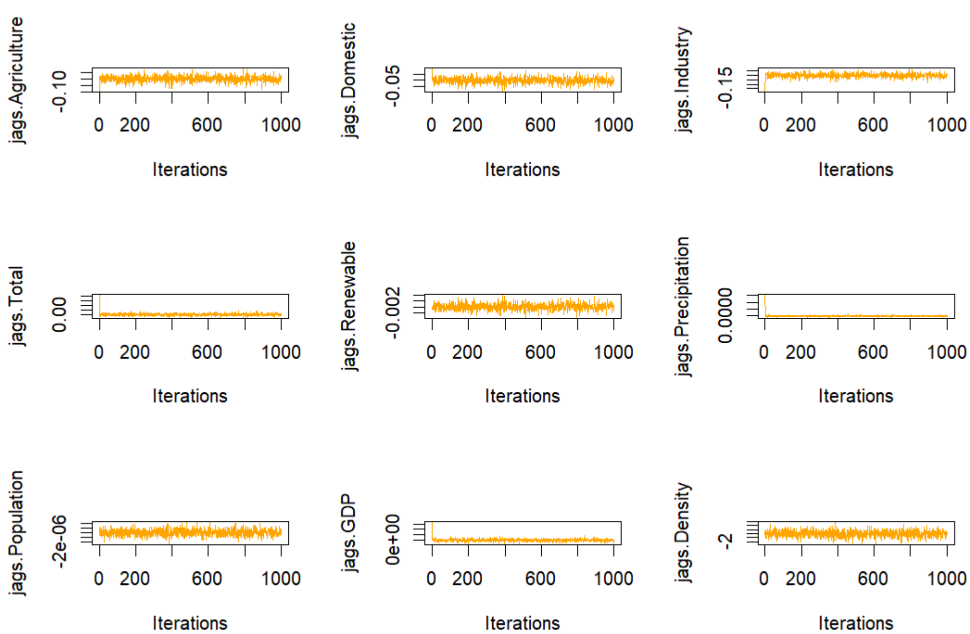

Figure A.4: Beta estimates in JAGS model within the iterations

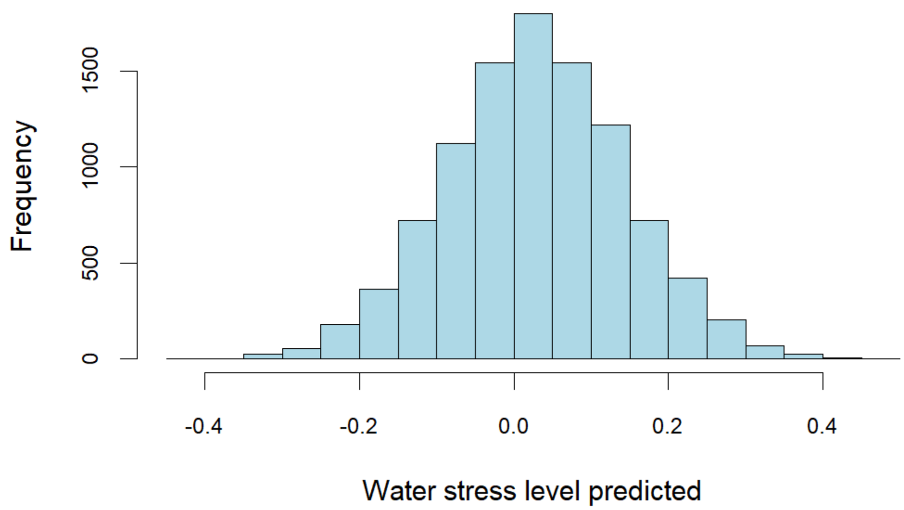

Figure A.5: Distribution of JAGS model's predictions

References

[20] *Climate Transparency Report: Colombia's Climate Action and Responses to the COVID-19 Crisis*. Climate Transparency, 2020.

[Aya20] M. O. Ayala. Avocado: the 'green gold' causing environment havoc. *World Economic Forum*, 2020. INDUSTRIES IN DEPTH.

[CVR14] C. P. Carvalho, M. A. Velásquez, and Z. V. Rooyen. Determination of the minimum dry matter index for the optimum harvest of 'hass' avocado fruits in colombia. *Agronomía Colombiana*, 32(3): 399–406, 2014.

[Foo21] Food and Agriculture Organization of the United Nations (FAO). *Progress on level of water stress: Global status and acceleration needs for SDG Indicator 6.4.2*. Tech. rep. Food and Agriculture Organization of the United Nations (FAO), 2021.

[FU23] Food and A. O. of the United Nations. *Building Responsible Global Value Chains for Sustainable Tropical Fruits: Ripe for Change: Adapting Avocado Production to a Changing Climate*. FAO, 2023.

[HPG22] T. Herrera, V. Parejo, and Á. González-Delgado. Environmental analysis of avocado (laurus persea l.) oil production in north colombia using the waste reduction algorithm. *Chemical Engineering Transactions*, 91: 235–240, 2022.

[Ker24] Keras. *Code Examples*. 2024. URL: `https://keras.io/examples/`.

[La 24] 'La Sapienza' University of Rome. *The role of international organizations in producing official statistics*. Ed. by P. Gennari. Ed. by E.-C. S. E.-F. C. Statistician. Master in Statistical Methods and Applications, Lectures 18 & 19 – Measuring Agricultural Sustainability. Rome, Italy: 'La Sapienza' University of Rome, 2024.

[Man23] Mani Tese. The (un)sustainability of the avocado production in colombia, 2023. Ed. by G. D. Salvo. Publication date: May 2023.

[MMH22] D. A. V. Martínez, N. F. Molina, and D. A. R. Herrera. A historical review of the use of pesticides in colombian agriculture. *Preprints*, 2022. Correspondence: davarela@universidadean.edu.co; Tel.: (+57) 300 737 5426.

[Saf23] S. Safdie. What is the avocado's environmental impact?. *Greenly*, 2023. Updated on July 27, 2023.

[Ten24] TensorFlow. *Regressione di base: prevedere l'efficienza del carburante*. 2024. URL: `https://www.tensorflow.org/tutorials/keras/regression?hl=it`.

[UNw18] UNwater. *Progress on Level of Water Stress: Global baseline for SDG indicator 6.4.2*. UNwater, 2018.

[UNw24] UNwater. *Progress on Level of Water Stress*. 2024. URL: `https://sdg6data.org/en/indicator/6.4.2`.