

README

Reproducibility package for Evaluating Paraguay's Vulnerability to the EU Deforestation-Free Regulation

Overview

This replication package includes all the data and code required to reproduce the analysis and replicate the figures and tables presented in Borchers Arriagada, Conte Grand, Reyes, Rojas, and Yoong (2025). While every effort has been made to ensure consistency, minor differences in table formatting may occur due to manual adjustments. The “Data” folder contains the primary datasets used in the study. The “Code” folder provides the Stata and R scripts and additional files necessary to prepare, execute, and export the main figures and tables. All outputs generated by these scripts are stored in the “Results” folder. Finally, the “Paper” folder contains a copy of the manuscript.

Instructions to replicate the main results:

a) Replicating Figures 1 to 7 and Table 1:

Figures 1 to 7 and Table 1 can be reproduced by running the following Stata do-file:

.../Code/00-main-Figs17_Table1.do

Before running this file, update the working directory path in the do-file to the project root (where the Reproducibility_Package folder is located on your machine), then execute the script. This do-file will automatically replicate Table 1 and Figures 1 through 7.

The do-file saves figures in the folder **Results/Figures**, in PNG format. Table 1 is saved in **Results/Tables**, as Excel files. Each file is named according to the corresponding figure or table number.

All required data is included in this replication package. However, if you prefer to download the latest raw data, follow these steps:

1-) Download trade data for South America

- Visit: <https://wits.worldbank.org/> and log in.
- Navigate to **Advanced Query → Trade Data (UN Comtrade)**.
- Create a **New Query**, add a name and description.
- Set the following parameters:
 - **Reporters:** BOL, BRA, CHL, COL, ECU, PER, URY
 - **Products:**
 - **Nomenclature:** HS 2022
 - **Clusters:** ALL3 — Sub-Heading (all 6-digit HS codes)
 - **Partners:** All destinations except *World (WLD)*
 - **Year:** 2023

- **Trade Flow:** Exports
- Download the data as a **CSV file** and import it into Stata. Ensure these variables are stored as strings: Nomenclature, ReporterISO3, ProductCode, ReporterName, PartnerISO3, PartnerName, TradeFlow.

Importantly, be sure that all rows are downloaded. Some accounts allow only 100,000 rows per download. If this limitation applies, download one country at a time.

After obtaining data for these seven countries, create a new query for Argentina. Use the same parameters, except set “Nomenclature” to HS2017, as HS2022 data was unavailable when the paper was written.

Finally, append the Stata files for all countries and save the combined file as:
.../Data/Trade/Raw/SA_Exports/Exports_southamerica_2023.dta

These files were last downloaded on 11/12/2025. Note that they are imported into Stata assuming the file structure matches the original downloads. For replication purposes, you may use the files included in this package after verifying their contents against the latest version available on WITS.

2-) Go to

<https://unstats.un.org/unsd/classifications/Econ>

and download the following conversion tables:

- HSCodeandDescription.xlsx
- HS2022toHS1992ConversionAndCorrelationTables.xlsx
- HS2022toHS1996ConversionAndCorrelationTables.xlsx
- HS2022toHS2002ConversionAndCorrelationTables.xlsx
- HS2022toHS2007ConversionAndCorrelationTables.xlsx
- HS2022toHS2012ConversionAndCorrelationTables.xlsx
- HS2022toHS2017ConversionAndCorrelationTables.xlsx

Saved these files in the folder **.../Data/Trade/Raw/HSCodes**

These files were last downloaded on 11/27/2024. Note that they are imported into Stata assuming the file structure matches the original downloads. For replication purposes, you may use the files included in this package after verifying their contents against the latest version available at unstats.un.org.

3-) Export data for Paraguay is sourced from the country’s Customs and Tax authority (the Direccion Nacional de Tributacion). The raw data was downloaded from <https://www.dnit.gov.py/web/portal-institucional/datos-abiertos> in May 2024. The raw data is stored at “.../Trade/Raw/PRY_Exports/files/raw/ EXP_9723_raw.dta.” The do-file cleaning the data and preparing the database for the analysis is “.../Trade/Raw/PRY_Exports/code/PRY_EXP_9723_forJDReye.do.” The resulting cleaned file is stored in “.../Trade/Raw/PRY_Exports/ PRY_EXP_9723_forJDReye.dta.”

The raw data were cleaned by DECTI following the standard procedures applied to all country databases maintained in the Exporter Dynamics Database, described in [Fernandes et al. \(2016\)](#). DECTI has requested DIME not to publish the Paraguay raw data, the cleaned data, or the corresponding do-files in the public replication package. These files are provided here solely for internal replication purposes. Any questions regarding this matter should be directed to Ana Margarida Fernandes <afernandes@worldbank.org>, copying Daniel Reyes <danielreyes@worldbank.org>.

4-) Once these steps are completed, update the working directory path in the Stata do-file named:
.../Code/00-main-Figs17_Table1.do

Set it to the project root (where the “Reproducibility_Package” folder is located on your machine) and run the do-file. This script will automatically replicate Table 1 and Figures 1 through 7.

The do file saves figures in the folder “Results/Figures”, in PNG format. Table 1 is saved in the folder “Results/Tables”, as Excel files. Each file is named according to the corresponding figure or table number.

b) Replicating Figure 8, Tables 4 and 5:

Figure 8 and Tables 4 and 5 can be reproduced by running the following R script:

.../Code/00-main-Fig8_Tables45.r

This R script (**.../Code/00-main-Fig8_Tables45.r**) reads a CSV file (**df_forest_soybean_overlap_results.csv**) located in the **.../Results** folder and creates Tables 4 and 5 and 4 panels associated with Figure 8. Table 4 and 5 results will be saved in the **.../Results/Tables** folder in CSV format (**Table4_summary_overlap_scenarios.csv**, **Table5_summary_overlap_Departments.csv**), while the four panels for Figure 8 will be saved in the **.../Results/Figures** folder in PNG format (**Figure8_a.png**, **Figure8_b.png**, **Figure8_c.png**, **Figure8_d.png**).

Before running this R script, make sure the clean data and results file have been previously processed. Except for the file that estimates the overlap between soybean cover and forest cover, most of the pre-processing steps involve harmonizing datasets into the same spatial resolution (i.e., 10m spatial resolution from the JRC forest cover data). This reproducibility package already contains these (clean datasets and results file), and the following steps were taken to create them:

- 1) Process JRC forest cover tiles to create a 10m forest cover raster for Paraguay.
 - a. Make sure the following 4 JRC TIF files (tiles) are located in the **.../Data/JRC_forest_cover/Raw** folder:
 - i. JRC_GFC2020_V2_S10_W60.tif
 - ii. JRC_GFC2020_V2_S10_W70.tif
 - iii. JRC_GFC2020_V2_S20_W60.tif
 - iv. JRC_GFC2020_V2_S20_W70.tif

- b. These files were originally downloaded from the following website: https://forobs.jrc.ec.europa.eu/GFC#data_download
 - c. Load the 4 JRC TIF files (i.e., tiles) in your Geographic Information System (GIS) application (e.g., ArcGIS, QGIS, etc.).
 - d. Create a single raster file from these 4 TIF files. For example, in QGIS, you can use the **Merge** tool under **Raster>Miscellaneous>Merge**.
 - e. Load the spatial polygons for Paraguay located in the **.../Data/shapefiles/Raw/gadm41_PRY_shp** folder. This shapefile was obtained from the following website: <https://gadm.org/data.html>.
 - f. Crop the single raster created in d) with the shapefile in e). For example, in QGIS, you can use the **Clip Raster by Mask Layer** tool under **Raster>Extraction> Clip Raster by Mask Layer**.
 - g. Export the cropped raster file into a TIF file. We have exported the file **JRC_forest_cover_Paraguay.tif** to the **.../Data/JRC_forest_cover/Cleaned** folder.
- 2) Process MapBiomass land cover data to create a 10m forest cover raster for Paraguay.
 - a. Make sure the following TIF file is located in the **.../Data/MapBiomass_forest_cover/Raw** folder:
 - i. **mapbiomas_paraguay_collection2_integration_v1-classification_2020.tif**
 - b. This file was originally downloaded from the MapBiomass Paraguay website (<https://paraguay.mapbiomas.org/>), using the following link: https://storage.googleapis.com/mapbiomas-public/initiatives/paraguay/collection_2/mapbiomas_paraguay_collection2_integration_v1-classification_2023.tif
 - c. Run the **.../Code/Process_spatial_data/1_make_MapBiomass_forest_cover_and_resample_10m.r** R script. This script will process the raw TIF file in a), will create a forest cover raster, and will finally harmonize (i.e. resample) to the same 10m resolution of the JRC forest cover file. To create a forest cover raster file, the following level 2 land cover classes were assigned a value of 1: 1.1 Closed Natural Woodlands, 1.2 Open Natural Woodlands, and 1.3 Flooded Natural Woodlands. The rest level 2 land cover classes were assigned a value of 0.
 - d. The following TIF file is saved in the **.../Data/MapBiomass_forest_cover/Cleaned** folder:
 - i. **r_mapbiomas_forest_cover_2020_10m_resampled.tif**
- 3) Resample GLAD soybean cover to a 10m spatial resolution raster.
 - a. Make sure the following TIF files are located in the **.../Data/GLAD_soybean_cover/Raw** folder:
 - i. **SouthAmerica_Soybean_2020.tif**
 - ii. **SouthAmerica_Soybean_2022.tif**
 - iii. **SouthAmerica_Soybean_2023.tif**
 - b. These files were originally downloaded from the **Commodity Crop Mapping and Monitoring in South America** project from the **Global Land Analysis & Discovery**

(GLAD) website: <https://glad.umd.edu/projects/commodity-crop-mapping-and-monitoring-south-america>.

- c. Run the r script **.../Code/Process_spatial_data/2_resample_GLAD_soybean_to_10m.r**. This script will read the raw TIF files in a) and will harmonize (i.e. resample) these to the same 10m spatial resolution of the JRC forest cover file (**.../Data/JRC_forest_cover/Cleaned/JRC_forest_cover_Paraguay.tif**).
 - d. The following TIF files are saved in the **.../Data/GLAD_soybean_cover/Cleaned** folder:
 - i. Paraguay_Soybean_2020_10m_resampled.tif
 - ii. Paraguay_Soybean_2022_10m_resampled.tif
 - iii. Paraguay_Soybean_2023_10m_resampled.tif
- 4) Resample INFONA soybean cover to a 10m spatial resolution raster.
- a. Make sure the **cultivos_soja22** Shapefile is located in the **.../Data/INFONA_soybean_cover/Raw** folder. In order to work correctly, it consists of a collection of 7 files with the following extensions: .cpg, .prj, .sbx, .sbn, .shx, .dbf, .shp.
 - b. This Shapefile was originally downloaded from the following **National Forestry Institute from Paraguay (INFONA)** link:
<https://nube.infona.gov.py/index.php/s/72DaFWHGPjgAbWA>.
 - c. Run the r script **.../Code/Process_spatial_data/3_rasterize_INFONA_soybean_cover_10m.r**. This script will read the raw Shapefile in a) and will rasterize (i.e. convert from polygons to pixels) this to a raster with the same 10m spatial resolution of the JRC forest cover file (**.../Data/JRC_forest_cover/Cleaned/JRC_forest_cover_Paraguay.tif**).
 - d. The following TIF file is saved in the **.../Data/INFONA_soybean_cover/Cleaned** folder:
 - i. r_INFONA_soybean_cover_10m.tif
- 5) Assess forest and soybean cover overlap.
- a. Run the r script **.../Code/Process_spatial_data/4_assess_forest_cover_vs_soybean_scenarios.r**. This script will read the clean files produced during the previous steps, and will calculate the level of forest and soybean overlap under 12 scenarios (see paper for more detail on methods and scenarios).
 - b. The following CSV file is saved in the **.../Results** folder:
 - i. df_forest_soybean_overlap_results.csv
- 6) Once the previous steps are completed, run the following R script to create Figure 8 and Tables 4 and 5: **.../Code/00-main-Fig8_Tables45.r**

c) Replicating Figures A.3.1, A.3.2, A.3.3, and A.4.1:

Figures A.3.1, A.3.2, A.3.3, and A.4.1 were created in Excel, and can be reproduced as follows:

1) Figures A.3.1, A.3.2, A.3.3 (soybean crops):

- a. Calculate soybean area in hectares by year and Districts using the R script **.../Code/Process_spatial_data/5_calculate_GLAD_soybean_Paraguay_2001_2023.r**. This R script will read the District-level Shapefile for Paraguay located in the **.../Data/shapefiles/Raw** folder and the 30m spatial resolution soybean maps for SouthAmerica located in the **.../Data/GLAD_soybean_cover/Raw** folder, and will calculate annual soybean area in hectares by district. It will save a CSV file (df_GLAD_annual_soybean.csv) in the **.../Results** folder.
- b. Calculate soybean area in hectares by year and Districts using the R script **.../Code/Process_spatial_data/6_calculate_INFONA_soybean_Paraguay_2022.r**. This R script will read the District-level Shapefile for Paraguay located in the **.../Data/shapefiles/Raw** folder and the 10m spatial resolution soybean map for 2022 located in the **.../Data/INFONA_soybean_cover/Cleaned** folder, and will calculate soybean area for 2022 in hectares by district. It will save a CSV file (df_INFONA_soybean_2022.csv) in the **.../Results** folder.
- c. Collect statistics of soybean crops in hectares by year from CAPECO (<https://capeco.org.py/area-de-siembra-produccion-y-rendimiento/>), USDA (<https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=PA&crop=Soybean>) and INBIO (<https://inbio.org.py/superficies-de-siembra/>). These have been collected and saved in the following files:
 - i. CAPECO: **.../Data/CAPECO/Raw/CAPECO_soybean_ha_annual.csv**
 - ii. USDA: **.../Data/USDA/Raw/USDA_soybean_ha_annual.csv**
 - iii. INBIO: **.../Data/INBIO/Raw/INBIO_soybean_ha_annual.csv**
- d. Collect statistics of soybean crops in hectares for the 2022/2023 (https://inbio.org.py/wp-content/uploads/2024/01/10-Zafra-Estimacion-superficie-zafra-2022-2023_.pdf) and the 2023/2024 (https://inbio.org.py/informes/superficies-siembra/2024/INBIO_Superficies_Soja_Arroz_Maiz_2023-2024.pdf) campaigns published by INBIO at a Department level. These files were saved in the **...Data/INBIO_soybean_cover/** folder.
- e. Collect statistics of soybean crops in hectares for 2022 published by INFONA. These are available in the report (REPORTE MAPA DE CULTIVOS DE SOJA 2022 20 DE MARZO - VERSION INGLES.pdf) published together with the original Shapefile (<https://nube.infona.gov.py/index.php/s/72DaFWHGPjgAbWA>). The report has been saved in the **...Data/INFONA_soybean_cover/** folder.

2) Figure A.4.1 (forest cover):

- a. Calculate forest area in hectares by year and Districts using the R script **.../Code/Process_spatial_data/7_calculate_MapBiomass_forest_Paraguay_2020.r**. This R script will read the District-level Shapefile for Paraguay located in the **.../Data/shapefiles/Raw** folder and the 10m spatial resolution forest cover map for 2020 located in the **.../Data/MapBiomass_forest_cover/Cleaned** folder, and will calculate forest area for 2020 in hectares by district. It will save a CSV file (df_MapBiomass_forest_2020.csv) in the **.../Results** folder.

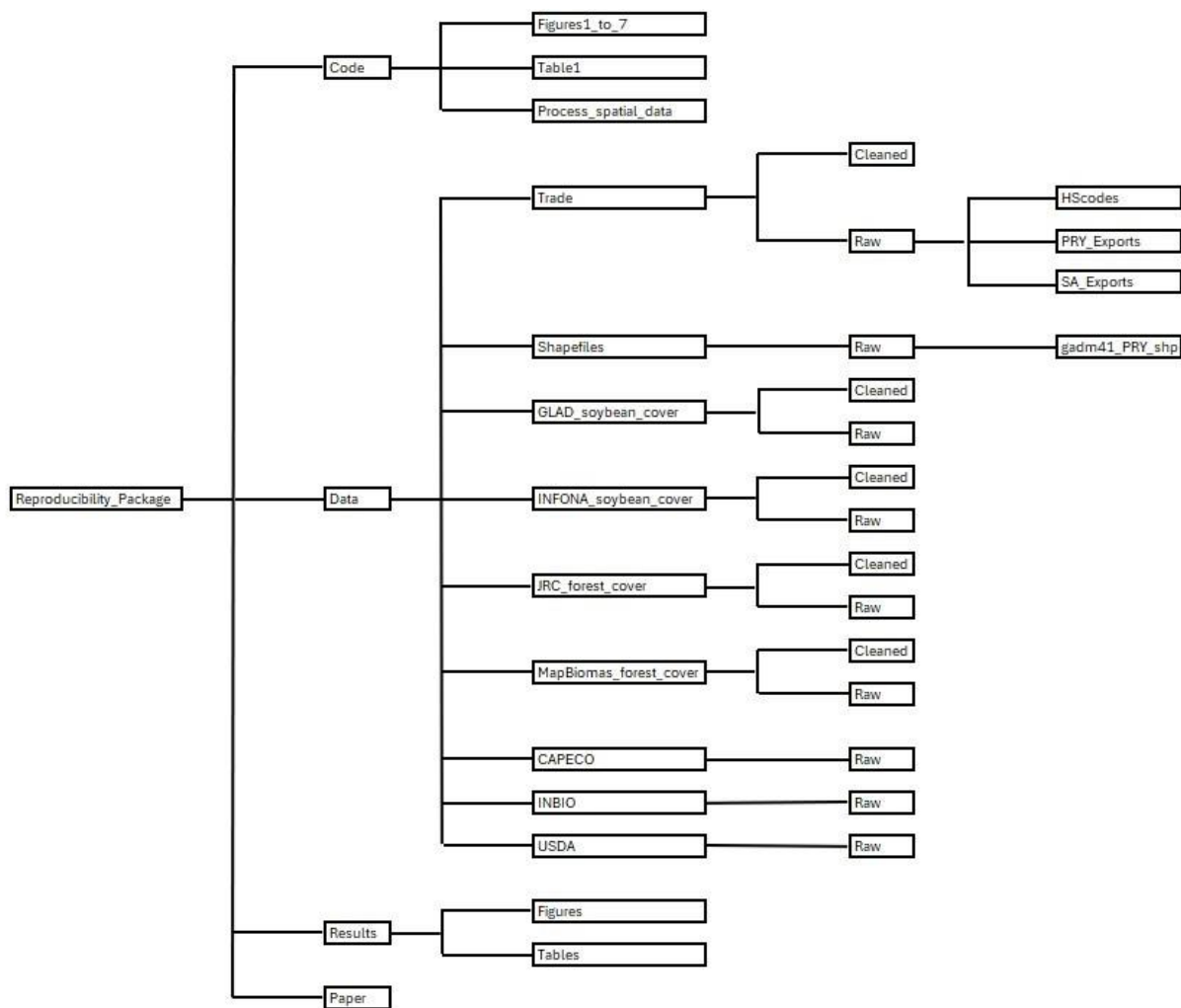
- b. Calculate forest area in hectares by year and Districts using the R script **.../Code/Process_spatial_data/8_calculate_JRC_forest_Paraguay_2020.r**. This R script will read the District-level Shapefile for Paraguay located in the **.../Data/shapefiles/Raw** folder and the 10m spatial resolution forest cover map for 2020 located in the **.../Data/JRC_forest_cover/Cleaned** folder, and will calculate forest area for 2020 in hectares by district. It will save a CSV file (df_JRC_forest_2020.csv) in the **.../Results** folder.

The data processed and collected in the previous steps (1.a–1.e, 2.a-2.b) were used to (manually) generate Figures A.3.1, A.3.2, A.3.3, and A.4.1. The file **.../Results/Figures/Figures_Annex.xlsx** contains these Figures (one figure per sheet).

Folder Structure:

The replication package includes four main folders: Code, Data, Paper, and Results. The complete structure is shown in the next figure:

Figure 1. Folder Structure



The contents of each main folder are detailed in the following sections.

A) Data

This folder contains all raw and cleaned datasets used in the analysis. The folder **Data** includes several subfolders. All subfolders, with the exception of **Data/shapefiles**, are divided into two main subfolders: **Cleaned** and **Raw**. **Data/shapefiles** only contains a **Raw** subfolder as the data was used as is.

In particular, the **Data/Trade/Raw** folder contains three additional subfolders: **HScodes**, **PRY_Exports**, and **SA_Exports** (see Figure 1).

A.1.1.1) Data/Trade/Raw/HScodes

This folder contains 16 files:

1. Seven Excel files

- One file: *HSCodeandDescription.xlsx*
- Six files: *HS2022toHSYearConversionAndCorrelationTables.xlsx*, where $Year \in \{1992, 1996, 2002, 2007, 2012, 2017\}$.
All files were downloaded from the UN Statistics Division (<https://unstats.un.org/unsd/classifications/Econ>) on **11/27/2024**. They include conversion and correlation tables across versions of the Harmonized System (HS) product classification used in international trade.

2. Seven Stata files

- *HS2022codes.dta*
- *correlations_HS2022toYear.dta*, for $Year \in \{1992, 1996, 2002, 2007, 2012, 2017\}$.
These files were generated from the Excel files above by the do-file:
Code/Figures1_to_7/aCorrelations_fromExceltoStata.do.

3. One PDF file

- *HS2022 conversion to earlier HS versions and other classifications – v.1.0.pdf*
Downloaded from <https://unstats.un.org/>. The document explains in detail how conversion and correlation tables between HS versions are constructed.

4. One Readme file

- *Readme.txt*, describing the sources of all raw data in this folder.

A.1.1.2) Data/Trade/Raw/PRY_Exports

This folder contains one file:

- *PRY_EXP_9723_forJDReyes.dta*
with the raw customs export data for Paraguay.
Variables include:
 - exporter ISO3 code (**c**)
 - year (**y**)
 - firm identifier (**Raw_data_firm_identifier**)
 - destination ISO3 code (**d**)
 - HS product code (**hs**)
 - export value (**v**)
 - export weight (**q**)

A.1.1.3) Data/Trade/Raw/SA_Exports

This folder contains:

- *Exports_southamerica_2023.dta*
Export data for the South American countries used in the analysis, retrieved from WITS, as explained above.
Variables include: HS version (**nomenclature**), reporting country ISO3 (**reporteriso3**), product code (**productcode**), reporter name (**reportername**), partner ISO3 (**partneriso3**), partner name (**partnername**), year (**year**), trade flow type and code (**tradeflowname**, **tradeflowcode**), and export value in thousands of USD (**tradevaluein1000usd**).

A.1.2) Data/Trade/Cleaned

This folder contains the cleaned datasets derived from the raw files:

1. *EUDR_products.dta*
Generated by: **Code/Figures1_to_7/bTargeted_Prds.do**
2. *PRY_EXP_9723_EUDR.dta*
Generated by: **Code/Figures1_to_7/cMerge_tradedata_EUDRprds.do**
3. *PRY_EXP_1023_EUDR.dta*
Also generated by: **Code/Figures1_to_7/cMerge_tradedata_EUDRprds.do**

A.2) Data/shapefiles/Raw

This folder contains the raw Shapefile of Paraguay by Districts (i.e., ADM2 or GID_2). This shapefile was obtained from the following website: <https://gadm.org/data.html>.

A.3.1) Data/JRC_forest_cover/Raw

This folder contains the raw files from the JRC Global Map of Forest Cover 2020, version 2 used for this analysis. It consists of the following four files, downloaded from https://forobs.jrc.ec.europa.eu/GFC#data_download:

1. *JRC_GFC2020_V2_S10_W60.tif*
2. *JRC_GFC2020_V2_S10_W70.tif*
3. *JRC_GFC2020_V2_S20_W60.tif*
4. *JRC_GFC2020_V2_S20_W70.tif*

A.3.2) Data/JRC_forest_cover/Cleaned

This folder contains the cleaned TIF file derived from the raw files.

1. *JRC_forest_cover_Paraguay.tif*

Generated using QGIS and the procedure detailed above.

A.4.1) Data/MapBiomass_forest_cover/Raw

This folder contains the raw TIF file of land cover for the year 2020 and the legend codes for the different land cover classes obtained from the MapBiomass Paraguay website (<https://paraguay.mapbiomas.org/>):

1. *mapbiomas_paraguay_collection2_integration_v1-classification_2020.tif*
2. *2.Cod_Leyenda-ES-PRY.pdf*

A.4.2) Data/MapBiomass_forest_cover/Cleaned

This folder contains the cleaned TIF file of forest cover for the year 2020 derived from the raw file:

1. *r_mapbiomas_forest_cover_2020_10m_resampled.tif*

Generated using the

.../Code/Process_spatial_data/1_make_MapBiomass_forest_cover_and_resample_10m.r R script.

A.5.1) Data/GLAD_soybean_cover/Raw

This folder contains the raw TIF files of soybean cover for the years 2021-2023 (23 files) obtained from the GLAD website (<https://glad.umd.edu/projects/commodity-crop-mapping-and-monitoring-south-america>), with the following name: *SouthAmerica_Soybean_{year}.tif*

A.5.2) Data/GLAD_soybean_cover/Cleaned

This folder contains the clean TIF files of soybean cover for the years 2020, 2022, and 2023 derived from the raw files.

1. *Paraguay_Soybean_2020_10m_resampled.tif*
2. *Paraguay_Soybean_2022_10m_resampled.tif*
3. *Paraguay_Soybean_2023_10m_resampled.tif*

Generated using the

.../Code/Process_spatial_data/2_resample_GLAD_soybean_to_10m.r R script.

A.6.1) Data/INFONA_soybean_cover/Raw

This folder contains the raw Shapefile (multi file format) of soybean cover for the year 2020 and a pdf report (in english) obtained from the following **National Forestry Institute from Paraguay (INFONA)** link: <https://nube.infona.gov.py/index.php/s/72DaFWHGPjgAbWA>:

1. *cultivos_soja22.cpg*
2. *cultivos_soja22.prj*
3. *cultivos_soja22.sbx*
4. *cultivos_soja22.sbn*
5. *cultivos_soja22.shx*
6. *cultivos_soja22.dbf*
7. *cultivos_soja22.shp*
8. *REPORTE MAPA DE CULTIVOS DE SOJA 2022 20 DE MARZO - VERSION INGLES.pdf*

A.6.2) Data/INFONA_soybean_cover/Cleaned

This folder contains the clean TIF file of soybean cover for the years 2020 derived from the raw Shapefile.

1. *r_INFONA_soybean_cover_10m.tif*

Generated using the

.../Code/Process_spatial_data/3_rasterize_INFONA_soybean_cover_10m.r R script.

A.7.1) Data/CAPECO/Raw

This folder contains a CSV file (*CAPECO_soybean_ha_annual.csv*) with the annual statistics of soybean surface in hectares published in the CAPECO website: <https://capeco.org.py/area-de-siembra-produccion-y-rendimiento/>

A.8.1) Data/USDA/Raw

This folder contains a CSV file (*USDA_soybean_ha_annual.csv*) with the annual statistics of soybean surface in hectares published in the USDA website:
<https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=PA&crop=Soybean>

A.9.1) Data/INBIO/Raw

This folder contains a CSV file (INBIO_soybean_ha_annual.csv) with the annual statistics of soybean surface in hectares published in the INBIO website: <https://inbio.org.py/superficies-de-siembra/>

B) Code

This folder contains all scripts used in the analysis. It includes three subfolders: **Figures1_to_7**, **Table 1**, and **Process_spatial_data** (see Figure 1).

- The **Figures1_to_7** folder contains the do-files that produce Figures 1–7.
- The **Table 1** folder contains the do-file that generates Table 1 of the paper.
- The **Process_spatial_data** folder contains the R scripts that are used clean the raw data, assess the forest and soybean cover overlap, and calculate forest and soybean surface area in hectares by Districts.

All scripts associated with Figures 1-7 and Table 1 can be run automatically using the main do-file **00-main-Figs17_Table1.do** located directly in the **Code** folder. Running this file will produce Figures 1–7 and Table 1 and save them in the **Results** folder.

To generate Figure 8 and Tables 4 and 5, the following four R scripts need to be run in that order:

1. **1_make_MapBiomass_forest_cover_and_resample_10m.r**
2. **2_resample_GLAD_soybean_to_10m.r**
3. **3_rasterize_INFONA_soybean_cover_10m.r**
4. **4_assess_forest_cover_vs_soybean_scenarios.r**

If cleaned spatial files have already been processed, steps 1-3 can be skipped.

To generate Annex Figures A.3.1, A.3.2, A.3.3, and A.4.1, the following x R scripts need to be run:

1. **5_calculate_GLAD_soybean_Paraguay_2001_2023.r**
2. **6_calculate_INFONA_soybean_Paraguay_2022.r**
3. **7_calculate_MapBiomass_forest_Paraguay_2020.r**
4. **8_calculate_JRC_forest_Paraguay_2020.r**

The previous R scripts will generate a series of CSV files summarising soybean and forest surface area at a district level. These CSV can be used to estimate country-level, department-level, and district-level surface areas.

C) Results

This folder stores all outputs generated by the code.

- **Figures** subfolder: contains Figures 1–8 in PNG format, and Annex Figures A3.1, A3.2, A3.3, and A4.1. Figure 8 is split into 4 figures, one for each panel. Annex Figures are contained in the following Microsoft XLSX file: **Figures_Annex.xlsx**.

- **Tables** subfolder: contains four Excel files, each corresponding to one panel of Table 1. It also contains two csv files for Tables 4 and 5.
- The following CSV files:
 - 1) df_forest_soybean_overlap_results.csv: contains the estimates of soybean and forest overlap for the 12 scenarios assessed.
 - 2) df_GLAD_annual_soybean.csv: contains the estimates of soybean surface area in hectares for the years 2001-2023 by District, using the data provided by GLAD.
 - 3) df_INFONA_soybean_2022.csv: contains the estimates of soybean surface area in hectares for the year 2022 by District, using the data provided by INFONA.
 - 4) df_MapBiomass_forest_2020.csv: contains the estimates of forest surface area in hectares for the year 2020 by District, using the data provided by MapBiomass Paraguay.
 - 5) df_JRC_forest_2020.csv: contains the estimates of forest surface area in hectares for the year 2020 by District, using the data provided by JRC.

As noted above, the outputs for Figures 1-7 and Table 1 are generated by **Code/00-main-Figs17_Table1.do**. The outputs for Figure 8 and Tables 4 and 5 are generated by running the following two R scripts: 1) **1_assess_forest_cover_vs_soybean_scenarios.r**, and 2) **2_create_fig8_and_tabs_4_5.r**. The Annex Figures (.../Figures/Figures_Annex.xlsx) is created manually using a series of outputs (estimates and statistics of soybean and forest surface area by year at a country-, department-, and district-level) previously described.

D) Paper

This folder contains a copy of the manuscript.

List of Exhibits:

Exhibit name	Output filename	Script/Source	Notes
Figure 1	Results\Figures\Figure1.png	Code\Figures_1_to_7\dFigure1.do	
Figure 2	Results\Figures\Figure2.png	Code\Figures_1_to_7\eFigure2.do	
Figure 3	Results\Figures\Figure3.png	Code\Figures_1_to_7\fFigure3.do	
Figure 4	Results\Figures\Figure4.png	Code\Figures_1_to_7\gFigure4.do	
Figure 5	Results\Figures\Figure5.png	Code\Figures_1_to_7\hFigure5.do	
Figure 6	Results\Figures\Figure6_Cattle.png Results\Figures\Figure6_Rubber.png Results\Figures\Figure6_Soy.png Results\Figures\Figure6_Wood.png	Code\Figures_1_to_7\iFigure6.do	One png file is created for each panel (i.e. commodity) of the figure.
Figure 7	Results\Figures\Figure7.png	Code\Figures_1_to_7\jFigure7.do	
Table 1	Results\Tables\Table1_Cattle.xlsx Results\Tables\Table1_Rubber.xlsx Results\Tables\Table1_Soy.xlsx Results\Tables\Table1_Wood.xlsx	Code\Table1\aTable1.do	One Excel xlsx file is created for each panel (i.e. commodity) of the table.
Table 2			
Table 3			
Table 4	Results\Tables\ Table4_summary_overlap_scenarios.csv	Code\2_create_fig8_and_tabs_4_5.r	One csv file is created for this table.
Table 5	Results\Tables\ Table5_summary_overlap_Departments.csv	Code\2_create_fig8_and_tabs_4_5.r	One csv file is created for this table.
Figure 8	Results\Figures\Figure8_a.png Results\Figures\Figure8_b.png Results\Figures\Figure8_c.png Results\Figures\Figure8_d.png	Code\2_create_fig8_and_tabs_4_5.r	Four png files are created for Figure 8, one for each panel.
Annex 1		Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023, Annex I.	Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32023R1115#d1e32-243-1 Last Visited: 11/14/2025
Annex 2			
Figure A3.1	Results\Figures\Figures_Annex.xlsx • Sheet: Figure_A.3.1_soy_Country		Created manually using the following intermediate results/outputs: • Results\df_GLAD_annual_soybean.csv

			• Data\CAPECO\Raw\CAPECO_soybean_ha_annual.csv • Data\USDA\Raw\USDA_soybean_ha_annual.csv • Data\INBIO\Raw\INBIO_soybean_ha_annual.csv
Figure A3.2	Results\Figures\Figures_Annex.xlsx Sheet: Figure_A.3.2_soy_Departments		Created manually using the following intermediate results/outputs/files: • Results\df_GLAD_annual_soybean.csv • Data\INBIO\Raw\10-Zafra-Estimacion-superficie-zafra-2022-2023_.pdf • Data\INBIO\Raw\INBIO_Superficies_Soja_Arroz_Maiz_2023-2024.pdf • Results\df_INFONA_soybean_2022.csv • Data\INFONA_soybean_cover\Raw\REPORTE MAPA DE CULTIVOS DE SOJA 2022 20 DE MARZO - VERSION INGLES.pdf
Figure A3.3	Results\Figures\Figures_Annex.xlsx Sheet: Figure_A.3.3_soy_Districts		Created manually using the following intermediate results/outputs: • Results\df_GLAD_annual_soybean.csv • Results\df_INFONA_soybean_2022.csv
Figure A4.1	Results\Figures\Figures_Annex.xlsx Sheet: Figure_A.4.1_forests		Created manually using the following intermediate results/outputs: • Results\df_MapBiomass_forest_2020.csv • Results\df_JRC_forest_2020.csv

Computational Requirements:

1-) Figures 1 to 7, and Table 1 are generated by Stata. Last run with Version 18.

2-) To download trade data from WITS (<https://wits.worldbank.org/>), you must create an account. Registration is free. Note, however, that accounts without unlimited-download privileges can only retrieve up to 100,000 rows per query. If this limitation applies to your account, it is recommended that you download the data one country at a time.

3-) Analyses and pre-processing of data for the assessment of overlap between forest and soybean crops was done using R version R-4.4.2 and QGIS version 3.34.13-Prizren. As the analysis was done on spatial datasets with a relatively high spatial resolution, we used a desktop with 64 GB of RAM.

Data Availability Statement:

The data used consists of:

Paraguay's Trade Customs Data

Use: Figures 1 to 7 and Table 1.

Reference: Datos Abiertos – Direccion Nacional de Ingresos tributarios- Gerencia General Aduanas.

Access: Data available at <https://www.dnit.gov.py/web/portal-institucional/datos-abiertos>

Last Downloaded: 05/01/2024

Comtrade Data (accessed through WITS)

Use: Figures 2 and 4.

Access: Data available at <https://wits.worldbank.org/>.

Last Visited: 11/12/2025.

List of Relevant HS product codes subject to the EUDR

Use: Figures 1 to 7 and Table 1.

Access: Available at:

<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32023R1115#d1e32-243-1>

Reference: REGULATION (EU) 2023/1115 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 31 May 2023, Annex I.

Last Visited: 11/14/2025.

Concordance and Correlation Tables across HS classification versions

Use: It was used to create the seven targeted commodities across different years.

Access: The data are available at <https://unstats.un.org/unsd/classifications/Econ>

Last visited: 01/27/2024.

Spatial data for Paraguay from the Global Administrative Areas (GADM).

Use: It was used to define the spatial boundaries for the entire country and for the Districts of Paraguay.

Access: Shapefile for Paraguay available at: <https://gadm.org/data.html>

Reference: GADM (version) (Year). Database of Global Administrative Areas [Data set]. Available at: <https://gadm.org/data.html> [Accessed 15/09/2025]

Last visited: 11/26/2025

JRC Global Map of Forest Cover 2020, version 2.

Use: It was used to define pixels (areas) that were considered as forest in the year 2020.

Access: Raw TIF files in tile format available at:
https://forobs.jrc.ec.europa.eu/GFC#data_download

Reference: Bourgoin, Clement; Verhegghen, Astrid; Amezttoy, Iban; Carboni, Silvia; Achard, Frederic; Colditz, Rene (2025): Global map of forest cover 2020 - version 3. European Commission, Joint Research Centre (JRC) [Dataset] PID:
<http://data.europa.eu/89h/8c561543-31df-4e1b-9994-e529afecaf54>

Last visited: 11/26/2025

MapBiomass Paraguay Land Cover Maps, Collection 2.

Use: It was used to define pixels (areas) that were considered as forest in the year 2020.

Access: Raw TIF file available from: <https://paraguay.mapbiomas.org/>

Reference: MapBiomass Paraguay Project - Collection 2 of the Annual Series of Land Use and Land Cover Maps, acquired on 02/07/2025 through the link:
<https://paraguay.mapbiomas.org/>

Last visited: 11/26/2025

Global Land Analysis and Discovery (GLAD) Soybean Cover Data

Use: It was used to define pixels (areas) that were considered as soybean in the years 2020, 2022, and 2023.

Access: Raw TIF files available from: <https://glad.umd.edu/projects/commodity-crop-mapping-and-monitoring-south-america>

Reference: Song, X.P., Hansen, M.C., Potapov, P. et al. Massive soybean expansion in South America since 2000 and implications for conservation. *Nat Sustain* **4**, 784–792 (2021).
<https://doi.org/10.1038/s41893-021-00729-z>

Last visited: 11/26/2025

National Forestry Institute from Paraguay (INFONA) Soybean Cover Data for 2022.

Use: It was used to define pixels (areas) that were considered as soybean in the year 2022.

Access: Raw shapefile available from:
<https://nube.infona.gov.py/index.php/s/72DaFWHGPjgAbWA>

Reference: Instituto Nacional Forestal (INFONA). 2023. *Mapa de zonas de cultivo de soja 2022*. Available at: <https://nube.infona.gov.py/index.php/s/72DaFWHGPjgAbWA> [Accessed 15/09/2025]

Last visited: 11/26/2025

Paraguayan Chamber of Exporters and Marketers of Cereals and Oilseeds, CAPECO.

Use: It was used to compare country-level annual estimates of soybean crop area.

Access: <https://capeco.org.py/area-de-siembra-produccion-y-rendimiento/>

Reference: CAPECO (2025). *Área de Siembra, Producción y Rendimiento*.
<https://capeco.org.py/area-de-siembra-produccion-y-rendimiento/>

Last visited: 11/26/2025

Paraguayan Agricultural Biotechnology Institute, INBIO.

Use: It was used to compare country- and department-level annual estimates of soybean crop area.

Access: <https://inbio.org.py/superficies-de-siembra/>; https://inbio.org.py/wp-content/uploads/2023/12/soja_maiz_zafarina2023.pdf;
https://inbio.org.py/informes/superficies-siembra/2024/INBIO_Superficies_Soja_Arroz_Maiz_2023-2024.pdf

Reference: INBIO (2025). Superficies de siembra. <https://inbio.org.py/superficies-de-siembra/>

Last visited: 11/26/2025

Foreign Agricultural Service from the U.S. Department of Agriculture, USDA.

Use: It was used to compare country-level annual estimates of soybean crop area.

Access: <https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=PA&crop=Soybean>

Reference: USDA (2025). Country summary: Paraguay Soybean Area, Yield and Production. <https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=PA&crop=Soybean>

Last visited: 11/26/2025