

Guide to using codes and data for the reproducibility of the results reported in the manuscript: “*Eco-morphodynamic carbon pumping by the largest rivers in the Neotropics*” by Salerno, Vezza, Perona, Camporeale

Prerequisites: Google Earth Engine account and Matlab® software are required for the analysis. To get access to Earth Engine, fill out the form at signup.earthengine.google.com. You will receive an email titled "Welcome to Google Earth Engine" with instructions for getting started.

The outcomes presented in this submission (Supplementary Dataset S6 and S7) were obtained through the combined use of the GEE Java script (provided here as a text file *MainCode.txt*) and *Matlab*® scripts (see *Result1.m*, *Results1E.m*, *Result2.m*, *Results2E.m*, *Result3.m*, *Results3E.m*, *Result4.m*, *Results4E.m*, *Result5.m*, *Result6.m*, *Results6E.m*, *Results7.m*, *Results7E.m*, *ResultsTot.m*, *ResultsTotE.m*, *Analysis.m* and *Classification.m*).

1) USE OF THE GEE SCRIPT

- i) Copy and paste the script *MainCode.txt* into the code Editor section of the GEE platform;
- ii) Run the code.
- iii) In order to export and visualize the results, access the “Task Manager” section in the right windows of the web page. Click the run button next to the exported task in the Tasks tab to start the export. A configuration dialog will appear that allows you to choose the format (CSV, GeoJSON, KML, or KMZ), filename, and drive folder for tables (we suggest maintaining the current setting, in order to avoid issues in the following post-elaboration steps).

The GEE script aims to process the Global datasets (see Supplementary Table S2) and obtain the raw results (i.e., EMCE of each ROI, Data to uncertainty analysis, Data to signature classification). The code requires as inputs: the ROIs geometries and datasets about forest loss, carbon density, anthropic impact, free-flowing rivers (for further details see the complete list of the loaded global datasets reported in the Supplementary Table S1).

Due to the amount of data analyzed, in order to avoid memory saturation in the GEE platform, the ROIs must be grouped into ten FeatureCollection (America1-America6, AmericaE, the E refers to collections containing ROIs do not covered by Baccini Data) and the analysis must be split into seven sections (Analysis1 to Analysis7). The outputs of the code are 70 comma-separated files (.csv) that will be saved into 7 folders (Results1 to Results7, one for each section) of the user’s Google Drive account. The running time of each part ranges from 10 minutes to some hours (depending on the number and dimensions of ROIs in the FeatureCollection of the part and analysis section used) while if all the (seventy) partitions are run simultaneously in a parallel way the computation time is about one day.

2) PROCESSING OF THE RAW OUTPUT

The tables exported from the GEE code need to be processed and merged before proceeding to the next analysis. The data processing involves the following steps:

- 1- Generate the *Summary Results* tables for each analysis section. In order to manage the table files, the *Results#.m* codes must be copied into each Google drive folder (*Results#*). In order to execute the codes correctly, you may need to change the variable "folderpath" at line 2 of each code, by indicating the correct position in your pc of the folders "*Results#*". If the code works properly, then two files (.xlsx) for each folder are generated (except for the folder *Results5* where only a file will be created).
- 2- Generate the *overall summary Result* table. To merge all the summary result tables, create a new folder where the merged tables will be saved. Run the Matlab codes *ResultsTot.m* and *ResultsTotE.m*. In order to execute the codes correctly, you may need to change the variables "folderpath#", indicating the correct position in your pc of the file .xlsx, created at the previous step. If the code works properly, then two new files (.xlsx) are generated (*ResultsTot.xlsx* and *ResultsTotE.xlsx*).

3) SIGNATURE CLASSIFICATION AND GENERATION OF THE FINAL OUTCOMES

At this step, the files *ResultsTot.xlsx*, *ResultsTotE.xlsx*, and *CentroidInRegions.xlsx* serve as input files. Firstly, the code "*Classification.m*" must be executed by paying attention to select the file *ResultsTot.xlsx* as input (change the file path in line 2 of the code if needed). If the code runs properly, a new output file will be generated (*Class.xlsx*). Finally, run the Matlab code *Analysis.m*, providing as input the files *ResultsTot.xlsx*, *ResultsTotE.xlsx*, and *Class.xlsx*. The result will be the generation of Supplementary Datasets S6 and S7.

4) DEMO VERSION: Application to a single ROI

The demo version of the code analyzes a single ROI and generates all outcomes related to such ROI. In the code was set the ROI n1 of the Ucaialy river as an example (the user can choose any other ROI contained in Supplementary Table S4, by changing the identification code appropriately on line 117 of the demo script). The GEE code has to be run following the instructions in the section above "USE OF THE GEE SCRIPT". In the demo version, 7 files (.csv) will be generated as output

To process the GEE outcomes, use the Matlab code *demoResults.m*. In order to execute the code correctly, you may need to change the variable "folderpath" at line 2 of the code, indicating the correct position in your pc of the folders "*ResultsDemo#*". If the code runs properly, a new output file will be generated (*ResultsDEMO.xlsx*).

Run the Matlab code *AnalysisDemo.m*, which requires the table *ResultsDEMO.xlsx* as input. As a result, an excel row containing all the results of the analyzed ROI is generated.