



VRIJE
UNIVERSITEIT
BRUSSEL



ENVIRONMENTAL PROGRAMMING

ASSIGNMENT N

N Tropical Cyclone impact data comparison between Wikimpacts1.0 and EM-DAT database

NI LI

PROF. DR. IR. ELGA SALVADORE

2024-2025





<u>1. INTRODUCTION</u>	2
<u>2. ASSIGNMENT</u>	2
2.1. DATASET	2
2.2. TASKS	2
2.3. OPTIONAL TASK	3
2.4. LIBRARIES	4
<u>3. SUBMISSION AND REPORT</u>	4
3.1. DO YOU NEED CLARIFICATION ON THE ASSIGNMENT?	4
3.2. HOW IS THE ASSIGNMENT EVALUATED?	4

1. INTRODUCTION

In recent years, the study of climate extreme events has garnered considerable attention due to the increasing frequency and intensity of such phenomena and their extensive impacts. Nevertheless, existing databases, such as EM-DAT, are limited by their focus on impact data at the national level, which constrains research on local impacts, climate attribution, and adaptation measures.

To address this gap, we introduce the new climate impact database, Wikimpacts 1.0, which includes impact information at the sub-national level. It is essential to align events from Wikimpacts with those in EM-DAT to establish a benchmark for data consistency regarding impacts.

In this project, you will be provided with the confidential Wikimpacts 1.0 database and tasked with matching the impact information of tropical cyclones to that of EM-DAT, as well as evaluating any discrepancies.

2. ASSIGNMENT

2.1. Dataset

1. Wikimpacts 1.0: contains data on the occurrence and impacts of climate extremes in country and sub-national scales. The database is inferred from Wikipedia and uses generative AI.
2. EM-DAT, please download from [Public EM-DAT platform](#), only "tropical cyclone" needed.

2.2. Tasks

1. Download the Wikimpacts 1.0 database in db format.
2. Load Data:
 - Read the database file and load all tables that start with "Total" into a DataFrame named `L1`.
 - Identify all tables that start with "Specific" and load them into separate DataFrames named `L3_*`, where `*` represents impact categories, only load Deaths, Injuries and Damage.
3. Filter by "Tropical Storm/Cyclone":
 - Using the "Main_Event", filter the Tropical Storm/Cyclone events from L1 into a new dataframe "L1_TC"
 - Using "Event_ID" from "L1_TC", filter the "L3_%" with only impact from Tropical Storm/Cyclone
 - "Start/End_Date_Year," "Start/End_Date_Month," and "Start/End_Date_Day" columns. If these date fields are missing in `L3_*`, fill them with the corresponding information from `L1_TC`.
4. Filter by Date:
 - In each `L3_*` DataFrame, filter the records to include only those events that occurred after the year 1900. Name these filtered DataFrames as `L3_*_1900`.
5. Aggregate by Administrative Area:

- Using the "Administrative_Area_GID" column in each `L3_*_1900` DataFrame obtained from Step 3, for the same "Event_ID", aggregate the impact from the same "Administrative_Area_GID", only consider the rows with one valid GID (specific cases like one country involving several GIDs, only use the one without digits, or the first 3 alphabets), name the new dataframe to `L3_*_1900_aggregated`

6. Comparison with L2 tables

- Read all tables that start with "Instance" and load them into separate DataFrames named `L2_*`, where `*` represents impact categories, only load Deaths, Injuries and Damage.
- Using the same Event_ID from 'L3_*_1900_aggregated', filter the events from 'L2_*', name as 'L2_*_filter'
- For the same Event_ID events, using the "Administrative_Area_GID" from 'L3_*_1900_aggregated' and the "Administrative_Areas_GID" from 'L2_*_filter', map the same GID, compute the impact data difference between 'L3_*_1900_aggregated' and 'L2_*_filter', for each impact category, get the average relative difference score. ('L3_*_1900_aggregated' / 'L2_*_filter') / 'L2_*_filter'.

7. Identify and Analyze same tropical cyclone (TC) Events:

- Using the ISO from EM-DAT, and Administrative_Areas_GID (only consider the row with one GID) in `L2_*_filter`, and "Start/End_Date_Year," "Start/End_Date_Month,", to identify the same TC events, and save a new dataframe as "EM_DAT_Wikimapcts_Matched".
- Calculate the impact (e.g., Deaths, mean of Num_Min and Num_Max) difference of these matched events. Using the relative difference, and category the difference to 5 categories, -50% less, -30% less, Perfect Match, +30% more, +50% more, and visualize the difference in a bar plot. (relative difference: (Wikimpacts-EM_DAT)/EM_DAT)
- Save the plot as "EM_DAT_Wikimpacts_*_comparison.png".

8. Analyze the spatial differences between two databases

- Using the ISO from EM-DAT, and Administrative_Areas_GID (only consider the row with one GID) in `L2_*_filter`, compute the number of impact data entries difference between two databases, and visualize the difference in a world map.
- Save the plot as "EM_DAT_Wikimpacts_Spatial_*_comparison.png".

2.3. Optional Task

Develop a simple graphical user interface (GUI) to interactively execute your code.

Tip: You are free to use any open access Python package and free to design and create the layout of the interface. I suggest you to try Tkinter (<https://wiki.python.org/moin/TkInter>) but this is just a suggestion, many other options are available.

Internet is an incredible source of information (docs, videos, tutorials...).
Just to list some websites and youtube videos...

<https://www.youtube.com/watch?v=eJRLftYo9A0&list=PLQVvaa0QuDcIKx-QpC9wntnURXVJqLyk&index=8>

<http://sebsauvage.net/python/gui/>
https://www.tutorialspoint.com/python/python_gui_programming.htm
<https://docs.python.org/3/library/tkinter.html>
<http://www.tkdocs.com/tutorial/index.html>

2.4. Libraries

Python libraries you may use in the project: Pandas, Geopandas, Matplotlib, NumPy

3. SUBMISSION AND REPORT

Develop a software that solves the steps described in the previous section and submit this together with a short report describing your software (objective, software structure, user manual like) a week before the exam date.

3.1. Do you need clarification on the assignment?

Please contact **Ni Li** and Elga Salvadore through the dedicated CANVAS forum. In the discussion forum you are able to post questions, share your code and request an appointment for an online session.

3.2. How is the assignment evaluated?

EVALUATION MATRIX Assignment	Point
1) Does your software perform the required tasks providing the correct result(s)?	10 points
2) Is your code readable?	± 1 point
3) Are the name of the variables\functions\modules meaningful?	± 1 point
4) Are the data types appropriate for values the hold?	± 0.5 point
5) Are there too many code repetitions?	± 0.5 point
6) Does your code have a coherent structure?	± 0.5 point
7) Is the report clear, to the point and complete?	± 1 point
8) Is your code easy to use?	± 1 point
9) Is there a good balance of comments and code in the scripts?	± 0.5 point
10) Code reuse: are existing codes used appropriately?	± 1 point
11) Code reuse: Did you forget to properly citing the source of codes you are reusing?	- 2 point

12) How flexible is your code? For example, if I use a different data set with a different file name, would it still be working?	± 1 point
13) How easy would be for the next user to adapt/modify your code?	± 1 point
14) Is there any documentation? Installation steps (if additional libraries are used), flow chart	± 0.5 point
15) Did you develop a GUI for your code?	Up to 2 points