**Practical Work n°2 : Hurricane vulnerability analysis for an insurance company**

**Context:**

* We use in this study data from a fake insurance company. Its portfolio is exposed to hurricane risk in Mexico.
* Mexican Pacific and Atlantic ocean coasts are regularly impacted by hurricanes.
* The vulnerability module is a crucial part of a CAT model. It is usually calibrated using claims data obtained from previous hurricanes.

**Goals:**

* Using and leveraging claims and portfolio data
* Getting familiar with the modeliing of hurricane vulnerability
* Developing vulnerability curves related to hurricane risks

**Inputs:**

* A fake portfolio from an insurance Mexico with exposure in Mexico
* Claims datab related to 3 historical hurricanes (Odile, Wilma et Patricia)

You can use the ppt presentation if you need more details.

1. **PORTFOLIO EXPOSURE ANALYSIS FROM AN INSURANCE COMPANY**

The file « 2-TP\_VULNERABILITY/INPUTS/ptf\_exposure.Rdata » is the fake portfolio from the insurance company we want to analyse. Each row of the porfolio gives information on a specific site insured by the company. Each column correspond to one of characteristics (occupation type, structure type, etc.) of insured sites.

* 1. ***– Read portfolio data (ptf\_exposure.Rdata) from the insurance company with the load function.***

***How many sites located in Mexico are insured by the company ?***

* 1. ***– What does each column of the database represent ?***
  2. ***– According to you, which characteristics available in the portfolio could be useful for assessing the vulnerability of buildings to hurricanes ?***
  3. ***Quantify the percentage of each modility for these different variables : LOB, OCCUPANCY\_TYPE, STRUCTURE\_TYPE et GEOCODING\_RESOLUTION.***

*Note : you can use the « table » function which is able to count the number of sites for each modallity of a variable.*

*You can also use the ggplot2 library to show the results using sme barplots. Some examples are available here :* <http://www.cookbook-r.com/Graphs/Bar_and_line_graphs_(ggplot2)/> or here : <https://www.r-bloggers.com/>

* 1. ***– By using the leaflet package, could you represent the different sites of the portfolio on a dynamical map.***

*Note : you can find some examples here on how to use the leaflet package :* <https://rstudio.github.io/leaflet/>

* 1. ***– Knowing now the different site locations and based on you knowledge on tropical cyclones, in which geographical area would the insurance company can have some large claims ?***

***1.6 – By manipulating correcly the portfolio, compute the number of sites and the total insured value per CRESTA. Which are the top 3 CRESTA with the highest number of sites and those with the highest insured value ?***

1. **CLAIMS DATA ANALYSIS**

Here, you will analyse the claims data from 3 historical hurricanes which had strong impacts in Mexico : Patricia in 2015, Odile in 2014 and Wilma in 2005.

* 1. ***– Load the claims database « 2-TP\_VULNERABILITY/INPUTS/claims\_all.Rdata ».***

Each row of the claims database contains the identifier of the damaged site (SITE\_ID), the claim loss amount in $ (LOSS\_AMOUNT), name of the event (EVENT) and an integer index equals to 1 when it actually refers to a claim (IF\_CLAIM).

* 1. ***– What are the total loss and the total number of claims for each of the 3 hurricanes ?***

As explained in the PowerPoint presentation, claims database and portfolio data need to be merged in order to get all site characteristics and loss information in a same database.

* 1. ***– Use the left\_join fonction of the dplyr package for joining the 2 databases.***

In the following parts, we will call « expo\_claim » the new database. After this manipulation, the new database needs to contain 17 columns and 96’519 rows.

* 1. ***– What are the total loss and the number of claims per « line of business » (LOB) for each hurricane ? Are they homogeneous among LOB ?***
  2. ***– Which type of occupancy (OCCUPANCY\_TYPE) contributes the most to the total loss ? What is the proportion (in %) of the total loss for this specific type of occupancy for the 3 hurricanes ?***
  3. ***– According to you, what are the reasons which explain that this type of occupancy contributes so much to the total loss ?***

1. **BUILDING THE VULNERABILITY CURVES**

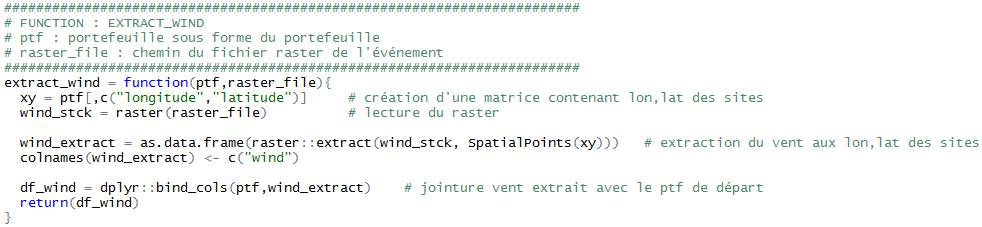
In this part, we will explain you a method which is useful for developing vulnerability curves. A vulnerability curve is defined as the relation between « Mean Damage Ratio » (MDR) and wind speed values.

You could use the ppt presentation to have more information on this part. We remind you that : MDR = (loss amount) / (insured value)

* 1. ***– Compute the « mean damage ratio » (MDR) for all damaged sites by doing the ratio between loss amount (LOSS\_AMOUNT) and total insured value (TIV).***

Then, you need to use raster files of hurricanes footprints. They are 3 footprints for 3 historical hurricanes (Patricia, Odile et Wilma) : « patricia\_swath\_fullgrid.tif », « odile\_swath\_fullgrid.tif » and « wilma\_swath\_fullgrid.tif ». They are available in « 2-TP\_VULNERABILITY/INPUTS/ ».

* 1. ***– For the 3 hurricanes, extract the wind speed at all sites contained in « expo\_claim » database. You can use the « extract\_wind() » function which is available below :***



*Note : Sites located outside from the footprints will get wind speed equal to NaN.*

If you did not succeed to merge wind speed with the « expo\_claim » database, you can load directly the file « OUTPUTS/expo\_claim\_Q3.2.Rdata ». It will be very useful for the following questions.

* 1. ***- Are there any claims (IF\_CLAIM=1) outside from the footprint (wind speed equal to 0)? Do you have an explanation for that ?***
  2. ***– Show the evolution of MDR in function of wind speed. Whqt do you notice on this graphic ?***
  3. ***– Empirical vulnerability***

Each data from the exposure of claims database is associated to a wind speed, which has been extracted from hurricanes footprints. For building vulnerability curves empirically, you can use bins of wind speed of m/s and compute the averaged MDR in each of these bins.

You can create bins of wind speed by using dplyr, magrittr libraries and the cut function as below :



* + 1. ***Create a barplot (using geom\_bar) showing MDR on the y-axis and wind speed on the x-axis. It is a way to represent a vulnerability curve.***
    2. ***Are there any unconsistencies in vulnerability curves when using an empirical approach ?***
    3. ***Do you know others methods for building vulnerability curves ?***
    4. ***Now you have built vulnerability curves, explain how they are used in CAT models ?***