TP - CAT RISK MODEL

December 2022

General Set-up of the Practical Work

- **The main concepts for creating two important modules for a CAT model are studied in this practical work:**
  - Modelling physical risk (hazard modelling) : Part 1
  - Modelling vulnerability (vulnerability modelling) : Part 2
- **We studied here these concepts by analysing hurricane risks in Mexico for a fake insurance company**
- **We use several datasets:**
  - Data from the portfolio of the fake insurance company based in Mexico
  - Fake claims data from 3 historical hurricanes (Odile, Wilma and Patricia)
  - Wind speed footprints from these 3 hurricanes
  - IBTrACS database

General set-up

- **Friday 2 December 2022 :**
  - General presentation
  - Explanation of the practical work

- **Work at home**
  - Prepare a ppt presentation with the main graphics answering the different questions
  - Send me your R script (or Python if you prefer) and the ppt presentation

- **Wednesday 11 January 2023 :**
  - Correction

R Packages

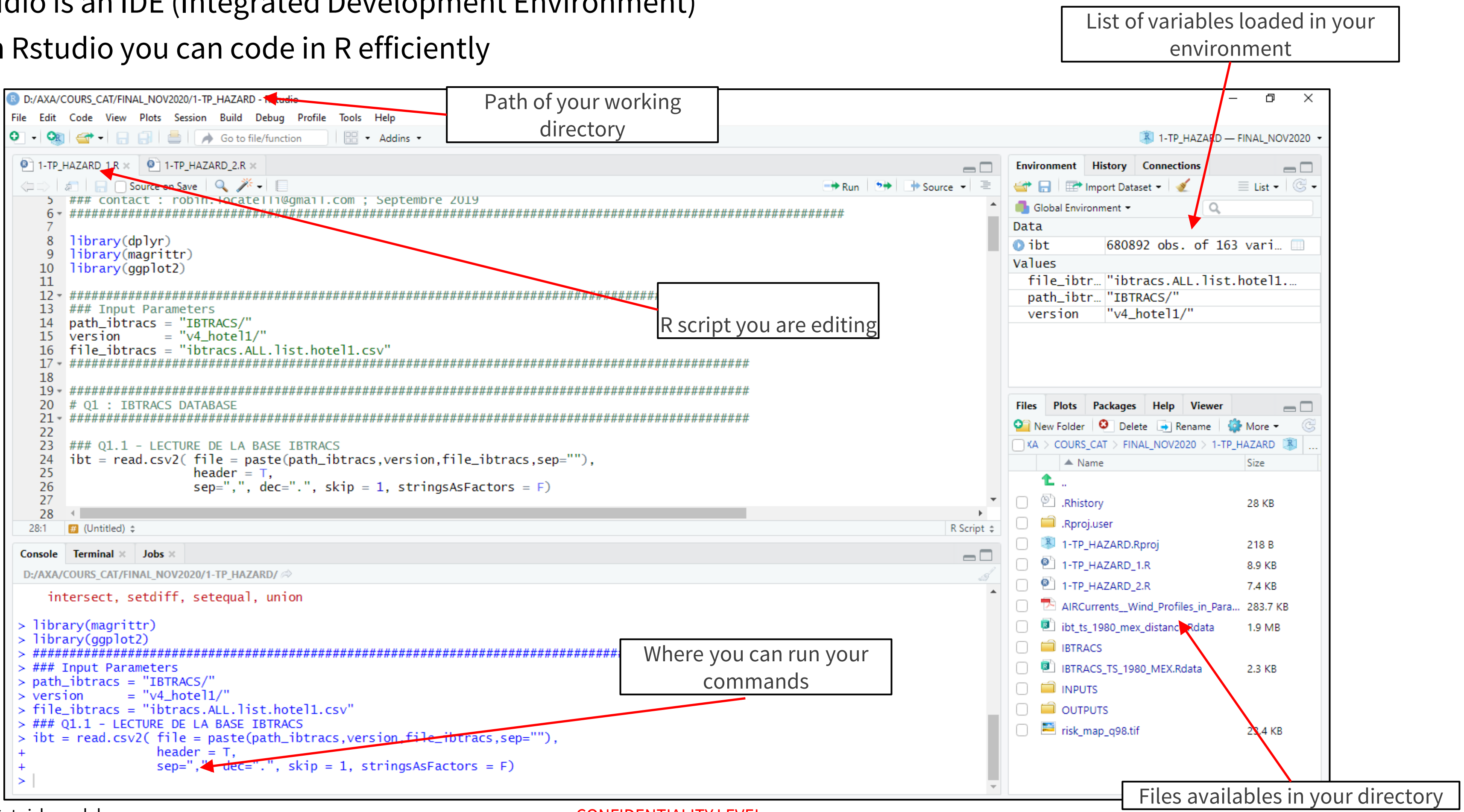
- Data.frames manipulation
  - → Useful packages: *dplyr* ; *magrittr*
  - → https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

- Rasters and shapefiles
  - → Useful packages : *raster* ; *rgdal* ; *sp* ; *geosphere*

- Graphical
  - → Packages : *ggplot2* ; *leaflet*

RSTUDIO Interface

- Rstudio is an IDE (Integrated Development Environment)
- With Rstudio you can code in R efficiently

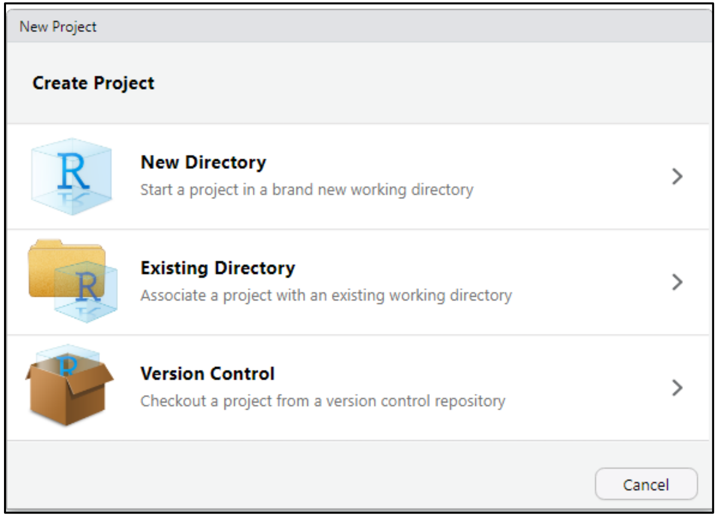


How to create a R project

- In « File » tab, click on « New Project »
- You get the same window that the one on the right
- You can choose « Existing Directory » option
- Choose the path of your working directory
- Click on « Create Project »

- When it is created, your project appears in your directory as a .Rproj file format
- You can then create a R script (Ctrl + Shift + N) and you can start to code

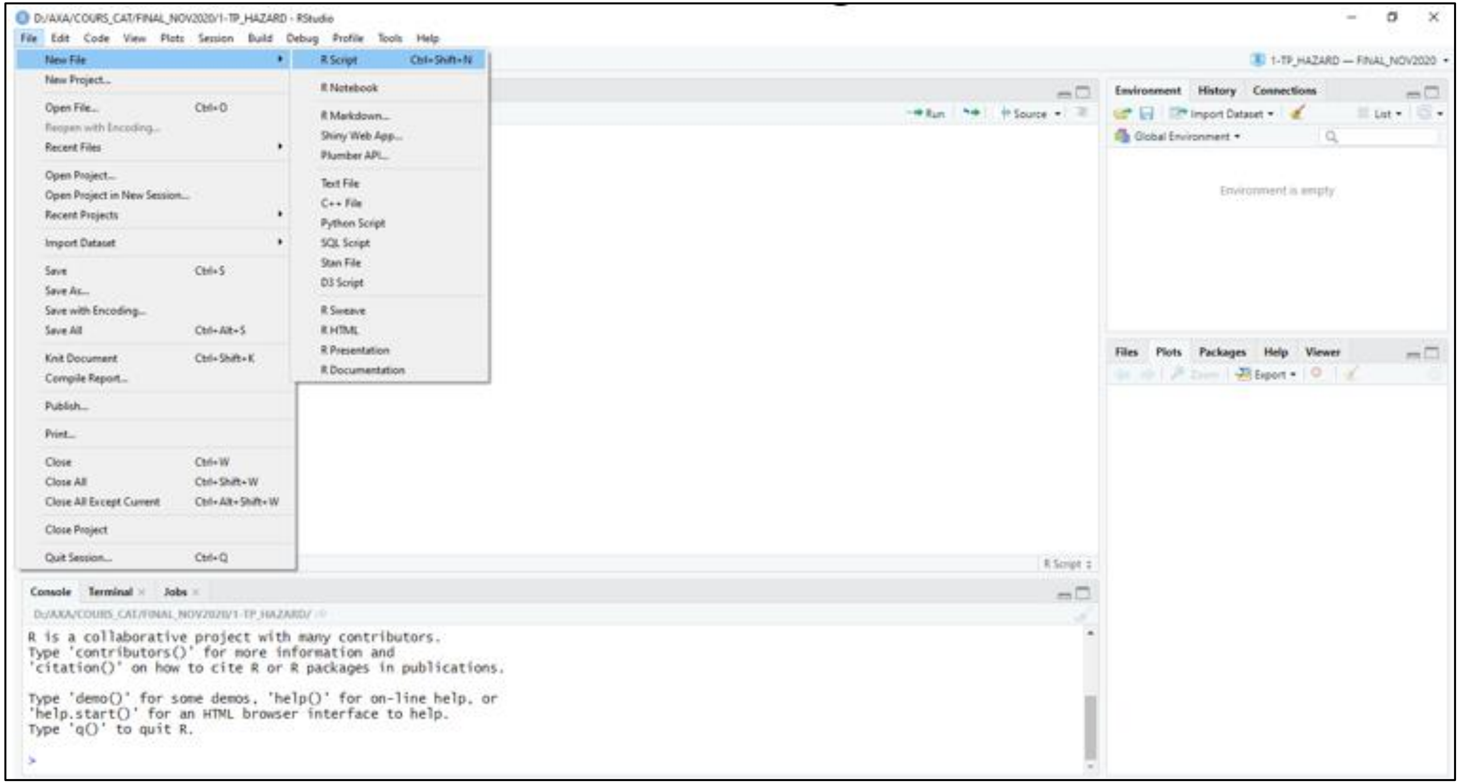
- When you come back in your working directory, you can open directly your project by clicking on the .Rproj file for resuming your work



Create a R script

- You can create a script by going into « File »
- Then, choose « New File » and « R script »
- You can also create directly a R script with « CTRL + Shift + N »

- You can save the script using « CTRL + S »
- For writing a commentary, you have to start the line by « # »
- This line wont be executed



Few useful commands

Load a .Rdata file into your environment – Using « load » function

→ df = get(load(‘’filename.Rdata’’))

Know all modalities of a specific variable - Using « table » function

For example, you want to know all modalities of COLOR variable in df dataframe:

→ table(df$COULEUR) ; you will get Red : 23 ; Blue : 45 ; Yellow 13

Manipulate dataframe – Using dplyr

→ df_out = dplyr::inner_join(df1,df2,by=‘’SITE_ID’’) : join df1 and df2 by keeping all common rows

→ df_out = dplyr::left_join(df1,df2,by=‘’SITE_ID’’) : join df1 and df2 by keeping all rows from df1

→ df_out = dplyr::mutate(df1, MDR = LOSS_AMOUNT/TIV) : create MDR variable which is the ratio between LOSS_AMOUNT and TIV.

TP – PART 1 : HAZARD MODELLING

- **IBTrACS : International Best Track Archive for Climate Stewardship**
- IBTrACS database give many information on tropical systems for all basins
- IBTrACS database is a reference for the topical system community

- You can find the database “*ibtracs.ALL.list.hotel1.csv*” in the folder 1-TP_HAZARD/INPUTS/
  - This database has 163 columns and 680’892 rows

SID	SEASON	NUMBER	BASIN	SUBBASIN	NAME	ISO_TIME	NATURE	LAT	LON	WMO_WIND	WMO_PRES	WMO_AGENCY	TRACK_TYPE	DIST2LAND	LANDFALL
1951037S07074	1951	11	SI	MM	NOT_NAMED	1951-02-11 15:00:00	TS	-27.96550	60.9343	NA	NA		main	898	89
1951037S07074	1951	11	SI	MM	NOT_NAMED	1951-02-11 18:00:00	TS	-28.40000	60.9000	NA	NA		main	940	94
1951037S07074	1951	11	SI	MM	NOT_NAMED	1951-02-11 21:00:00	TS	-28.80710	60.8899	NA	NA		main	981	98
1951037S07074	1951	11	SI	MM	NOT_NAMED	1951-02-12 00:00:00	TS	-29.17670	60.8942	NA	NA		main	1020	102
1951037S07074	1951	11	SI	MM	NOT_NAMED	1951-02-12 03:00:00	TS	-29.50130	60.8980	NA	NA		main	1049	104
1951037S07074	1951	11	SI	MM	NOT_NAMED	1951-02-12 06:00:00	TS	-29.80000	60.9000	NA	NA		main	1077	N
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 01:00:00	NR	-11.00000	99.0000	NA	1002	bom	main	828	82
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 03:00:00	NR	-11.17810	99.0765	NA	NA		main	832	82
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 06:00:00	TS	-11.42990	99.1897	NA	NA		main	836	83
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 09:00:00	TS	-11.68450	99.2661	NA	NA		main	849	84
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 12:00:00	TS	-11.93480	99.3168	NA	NA		main	863	86
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 13:00:00	TS	-12.01460	99.3257	NA	NA		main	870	87
1951038S11099	1951	12	SI	WA	NOT_NAMED	1951-02-07 15:00:00	TS	-12.18400	99.3363	NA	NA		main	885	88

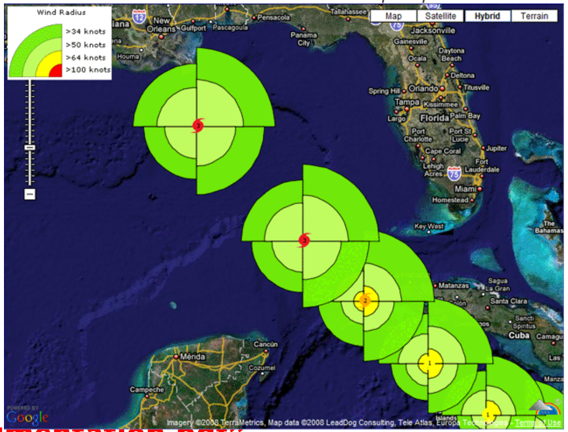
TP – PART 1 : HAZARD MODELLING

Présentation de la bae de données IBTrACS

- “Best Track” means that available information is the best estimation of the different parameters (eye location, max sustained wind, pressure, ..) with a temporal resolution of 6 hours
- Several agencies (NHC, JMA, IMD, MFLR, NZMS, etc.) are in charge to measure, correct and archive data of tropical systems

- **Most important parameters**
  - Location (longitude, latitude)
  - Intensity – Max sustained wind and minimal pressure
  - Classification (Tropicale, extra-tropical, sub-tropical)
  - Wind radii maximum extent in different direction for different wind speed (34kt, 50kt et 64kt). The unit is nautic miles.
  - Landfall: Yes or No the system went onshore

Illustration de l’extension des vents par quadrants



- **Useful information on the dataset is available here “IBTrACS_v04_column_documentation.pdf”**
- **General informations may be found here : https://www.ncdc.noaa.gov/ibtracs/**

TP – PARTIE 1 : HAZARD MODELLING

- Be careful all agencies do not use the same averaging period for measuring wind speed

- Despite data reanalysis efforts, problems are persisting

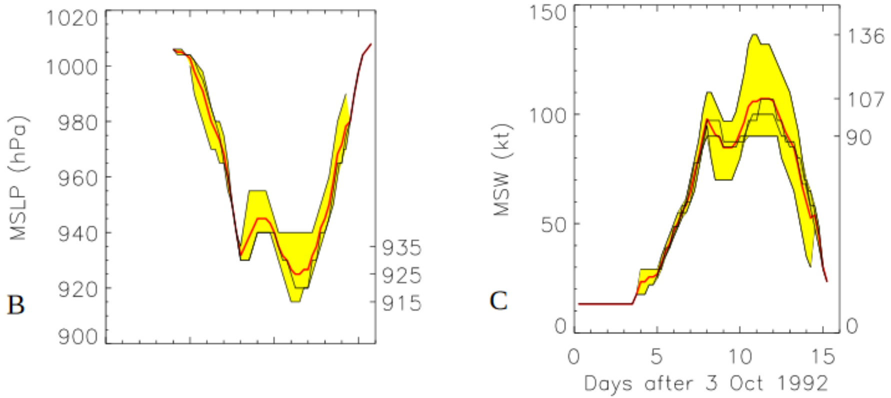
- For example, biases exist on the evolution of observation methosd to detect and monitor tropical systems
  - 1851 - today: data from cargoes
  - 1950 - today: using aircrafts and buoys
  - 1960 - today: satellite data

- It is usually recommended to use the most recent period of the database (after 1960 for example)

Table 3 - Wind speed averaging period by agency.

1-min wind	2-min wind	3-min wind	10-min wind
US Agencies (NOAA and JTWC)	CMA (China)	IMD (India)	JMA (Japan) BoM (Australia) La Reunion Nadi (Fiji) Wellington (New Zealand)

Uncertainties in MSW and MSLP for Yvette (1992) – Records for different agencies



TP – PARTIE 1 : HAZARD MODELLING

Présentation de la base de données IBTrACS

- A few orders of magnitude of uncertainties on the intensity and tracks of tropical systems
  - We notice a strong improvement on intensity forecast on the North Atlantic basin
  - We notice also a decrease in the track uncertainty when the intensity of the system increases

Table 1 - Qualitative uncertainty level for intensity in wind speed (knots). Blank boxes imply the level of uncertainty is too difficult to quantify (and possibly larger than 30 knots).

Period	SI	NI	SP	WP	EP	NA
pre1950						±30
1950-1965				±30		±30
1965-1973	±30	±30	±30	±20		±20
1973-1978	±20	±20	±20	±20	±20	±20
1978-1984	±15	±20	±20	±20	±20	±15
1984-1987	±15	±20	±15	±10	±20	±10
1987-1995	±15	±15	±15	±15	±15	±10
1995-2000	±10	±15	±15	±10	±15	±10
2000- now	±10	±10	±10	±10	±10	± 7

Table 2 - Uncertainty of TC position based on TC intensity

Approximate intensity of system	**Approximate uncertainty of position**
Weak TC (Winds < 60 kt)	~ 30-40 km (and larger before 1980)
Moderate TC (60 kt < Winds < 100 kt)	~ 20-25 km (and larger before 1980)
Strong TC (Winds > 100 kt)	~ 10-15 km (and larger before 1980)

TP – PART : VULNERABILITY MODELLING

I – Exposure data - Portfolio

- **Exposure data (called “portfolio”) contains the list of all sites insured by the company with all characteristics (localisation, structure type, …).**
- The portfolio of an insurance company is a dataset strictly confidential.
- In this exercise, we use a fake portfolio of an insurance company with lots of insured sites in Mexico.

- The accurate analysis of a portfolio may answer to the following questions :
  - **What are the different characteristics of the insured sites ?**
  - **Where are located these sites ?**

TP – PART 2 : VULNERABILITY MODELLING

I – Exposure data - Portfolio

- **Exposure data (called “portfolio”) contains the list of all sites insured by the company with all characteristics (localisation, structure type, …).**
- The portfolio of an insurance company is a dataset strictly confidential.
- In this exercise, we use a fake portfolio of an insurance company with lots of insured sites in Mexico.
- Below, you will find a sample of the portfolio :

SITE_ID	POLICY_ID	OCCUPANCY_TYPE	NUM_STORIES	COUNTRY	CRESTA	POSTCODE	longitude	latitude	GEOCODING_RESOLUTION	LOB	TIV
POL_00002_S00001	POL_00002	PERMANENT_DWELLING_SINGLE	14	MEX	JAL	48390	-105.23984	20.59518	STREET	RESIDENTIAL	26109.530
POL_00003_S00001	POL_00003	PERMANENT_DWELLING_SINGLE	1	MEX	COA	NA	-100.94000	25.53280	STREET	RESIDENTIAL	47799.508
POL_00004_S00001	POL_00004	PERMANENT_DWELLING_SINGLE	1	MEX	DIF	1280	-99.18700	19.49390	POSTCODE	RESIDENTIAL	23792.900
POL_00005_S00001	POL_00005	PERMANENT_DWELLING_SINGLE	2	MEX	CHP	29200	-92.63035	16.74093	BUILDING	RESIDENTIAL	9889.548
POL_00006_S00001	POL_00006	PERMANENT_DWELLING_SINGLE	1	MEX	JAL	NA	-103.36000	20.70800	POSTCODE	RESIDENTIAL	20401.362
POL_00008_S00001	POL_00008	PERMANENT_DWELLING_SINGLE	1	MEX	SON	NA	-110.91000	27.88040	STREET	RESIDENTIAL	483750.000
POL_00009_S00001	POL_00009	PERMANENT_DWELLING_SINGLE	1	MEX	PUE	NA	-98.22800	19.08650	STREET	RESIDENTIAL	431732.569
POL_00010_S00001	POL_00010	PERMANENT_DWELLING_SINGLE	1	MEX	DIF	NA	-99.06700	19.40990	POSTCODE	RESIDENTIAL	86276.757
POL_00011_S00001	POL_00011	PERMANENT_DWELLING_SINGLE	1	MEX	DIF	NA	-99.14300	19.51560	STREET	RESIDENTIAL	44539.184
POL_00012_S00001	POL_00012	PERMANENT_DWELLING_SINGLE	1	MEX	CHH	NA	-106.14000	28.73850	STREET	RESIDENTIAL	12152.898
POL_00013_S00001	POL_00013	PERMANENT_DWELLING_SINGLE	1	MEX	TAM	NA	-98.22600	26.03510	STREET	RESIDENTIAL	20401.362
POL_00014_S00001	POL_00014	PERMANENT_DWELLING_SINGLE	1	MEX	JAL	NA	-103.36000	20.72880	STREET	RESIDENTIAL	9413.306
POL_00012_S00002	POL_00012	PERMANENT_DWELLING_SINGLE	1	MEX	CHH	NA	-106.12000	28.73760	BUILDING	RESIDENTIAL	9889.548

TP – PART 2 : VULNERABILITY MODELLING

I – Exposure data - Portfolio

- We describe below the different variables available in the portfolio.
  1. **SITE_ID : site indentifier**
  2. **POLICY_ID : policy identifier.** Some sites have same policy (ex : commercial centers)
  3. **OCCUPANCY_TYPE : type of occupancy**
     - Hotel, Permanent Dwelling Single ; Apart ; Retail_Trade ; …
  4. **« LOB – Line Of Business » : business activity**
     - commercial, residential, industrial
  5. **NUM_STORIES : Number of stories**
  6. **COUNTRY : country where the site is located**
  7. **CRESTA : administrative region**
     - ROO : Quintana Roo ; BCS : Baja California Sur ; …
  8. **POSTCODE**
  9. **LONGITUDE**
  10. **LATITUDE**
  11. **GEOCODING_RESOLUTION : resolution of geocoding**
      - Geocoding can be done at Cresta, postal code, street or building levels.

TP – PART 2 : VULNERABILITY MODELLING

I – Exposure data - Portfolio

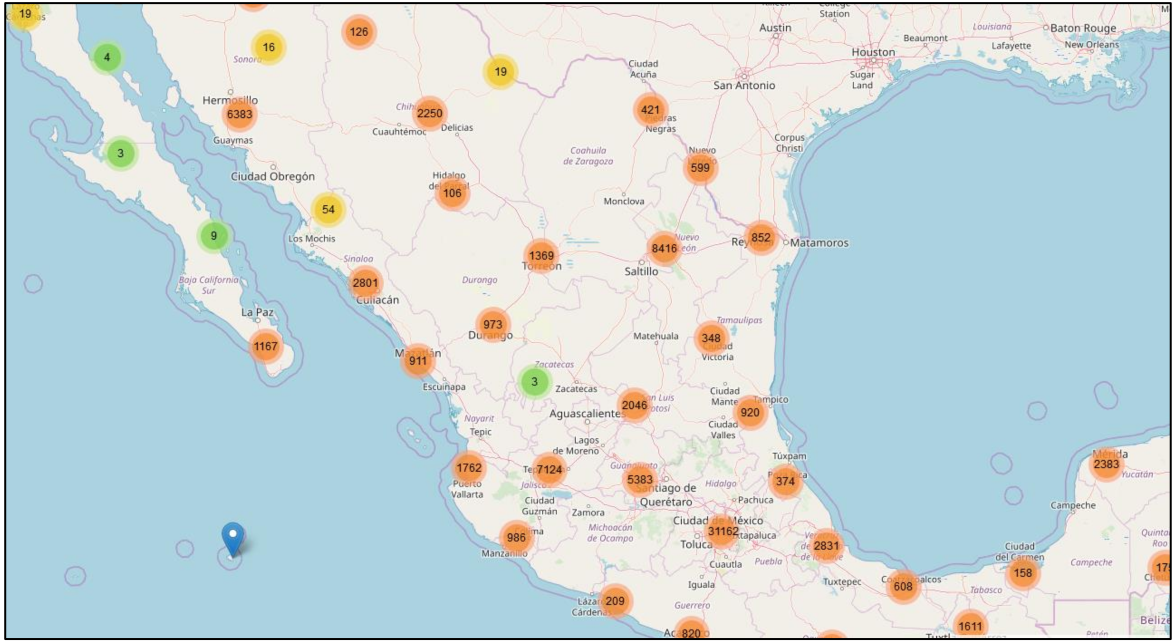
- We describe below the different variables available in the portfolio.
  12. **TIV (Total Insured Value) : value in $ of the insured site**
  13. **DEDUCT (Deductible) : Deductible included in the insurance contract**
  14. **STRUCTURE_TYPE**

- Financial conditions are the characteristics of insurance contract. These conditions can be very complex. In this exercise, we won't go in details on financial conditions.
- However, for your information, the maximum loss for a specific site is equal to the difference between the insured value (TIV) and the deductible (DEDUCT).

TP – PART 2 : VULNERABILITY MODELLING

I – Exposure data - Portfolio

- Geocoded data may give a spatial representation of insured sites.
- In one of the question, you will be asked to create a map like the one below. It shows the localization of all sites contained in the portfolio.



CRESTA

- CRESTA are administrative regions. They are shown below.

TP – PART 2 : VULNERABILITY MODELLING

II – Claims data

- **The claims dataset contains all damages suffered by insured sites of the portfolio during historical events.**

- Using claims data by insurance company give information on:
  - The reasons of the claims
  - Define insurance premiums based on characteristics of the site
  - Set up prevention measures

Sample of the claims data

SITE_ID : identifier of the site
EVENT : name of the event
LOSS_AMOUNT : loss amount in $
IF_CLAIM equals to 1, it means it is a claim data
(useful for jointure)

SITE_ID	EVENT	LOSS_AMOUNT	IF_CLAIM
POL_22775_S00001	WILMA	8.006172e+04	1
POL_22876_S00001	WILMA	1.476115e+04	1
POL_23021_S00001	WILMA	3.229228e+04	1
POL_23550_S00001	WILMA	2.045144e+05	1
POL_23580_S00001	WILMA	7.367268e+05	1
POL_00314_S00031	WILMA	1.322823e+04	1
POL_04452_S00004	WILMA	5.000000e+00	1
POL_24025_S00001	WILMA	7.831497e+02	1
POL_24186_S00001	WILMA	1.204852e+04	1

TP – PART 2 : VULNERABILITY MODELLING

II – Claims data

- **In general, claims database does not contain all characteristics of sites**
- Consequently, you need to merge the claims database with the exposure database in order to have all available information in the same database.
- In R, you can use “dplyr” library with “left_join”, “right_join”, “inner_join”,…

Claims database

SITE_ID	EVENT	LOSS_AMOUNT	IF_CLAIM
POL_22775_S00001	WILMA	8.006172e+04	1
POL_22876_S00001	WILMA	1.476115e+04	1
POL_23021_S00001	WILMA	3.229228e+04	1

X

Portfolio – Exposure database

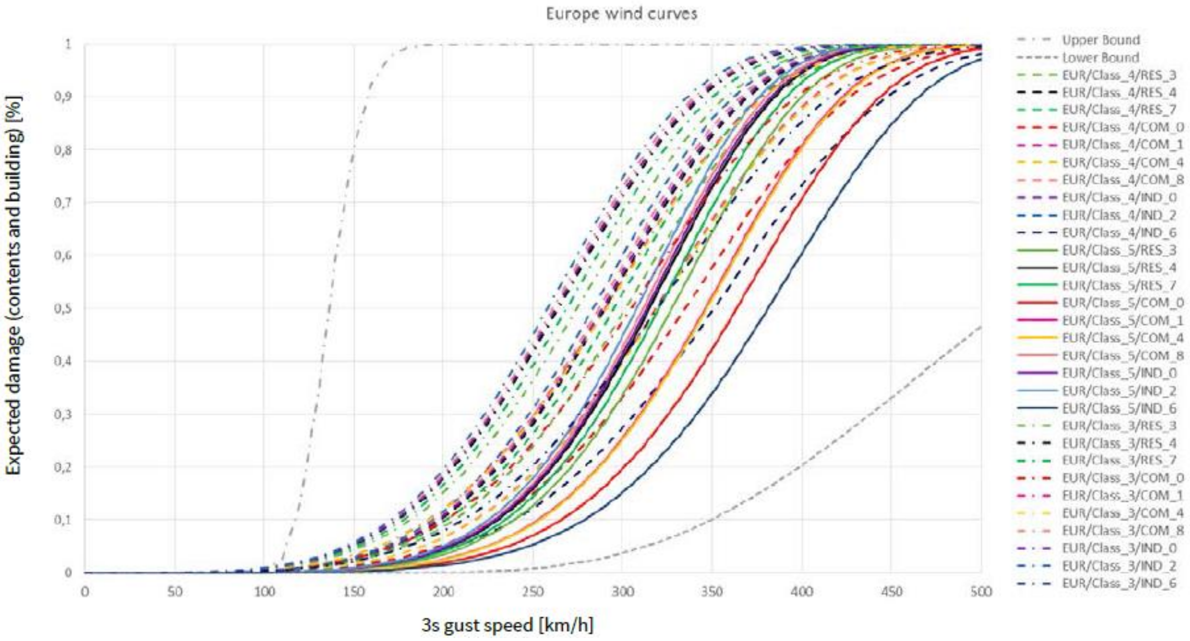
SITE_ID	POLICY_ID	OCCUPANCY_TYPE	NUM_STORIES	COUNTRY	CRESTA	POSTCODE
POL_00002_S00001	POL_00002	PERMANENT_DWELLING_SINGLE	14	MEX	JAL	48390
POL_00003_S00001	POL_00003	PERMANENT_DWELLING_SINGLE	1	MEX	COA	NA
POL_00004_S00001	POL_00004	PERMANENT_DWELLING_SINGLE	1	MEX	DIF	1280

Full database with characteristics from both exposure and claims databases

TP – PART 2 : VULNERABILITY MODELLING

Vulnerability curves

- In this exercise, you will try to create vulnerability curves
- A vulnerability curve links wind speed values with mean damage ratio (MDR).
- MDR (mean damage ratio) is defined as the ratio between claims loss and total insured value : MDR = (claims loss) / (insured value)

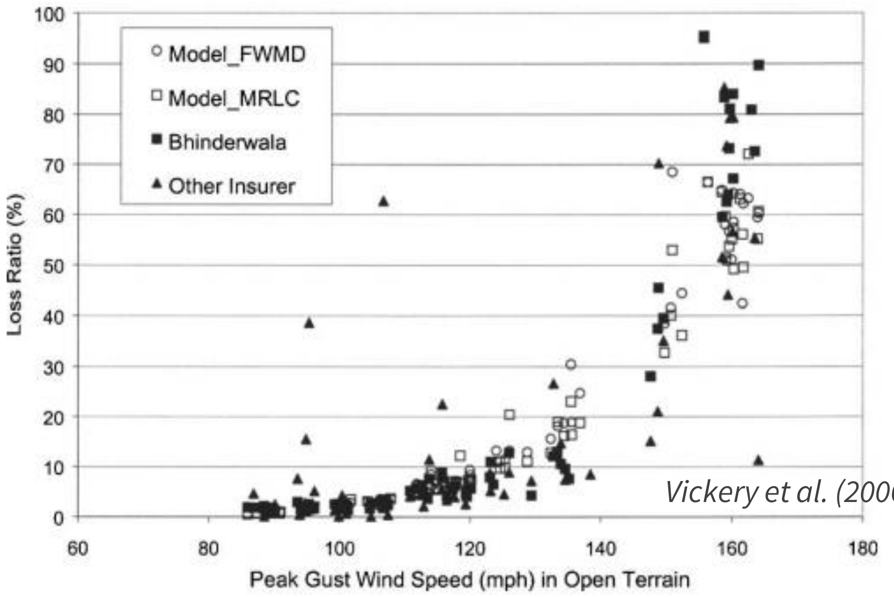


TP – PART 2 : VULNERABILITY MODELLING

Vulnerability curves

- Vulnerability curves are usually based on the using of historical claims data. But it is not the case every time…

- There are several methods:
  - Empirical method
  - Analytical method
  - Engineering method
  - Hybrid method



Vickery et al. (2006)

TP – PART 2 : VULNERABILITY MODELLING

Vulnerability curves

- For building these vulnerability curves, you will need to extract wind speed at insured site locations by using longitude/latitude coordinates

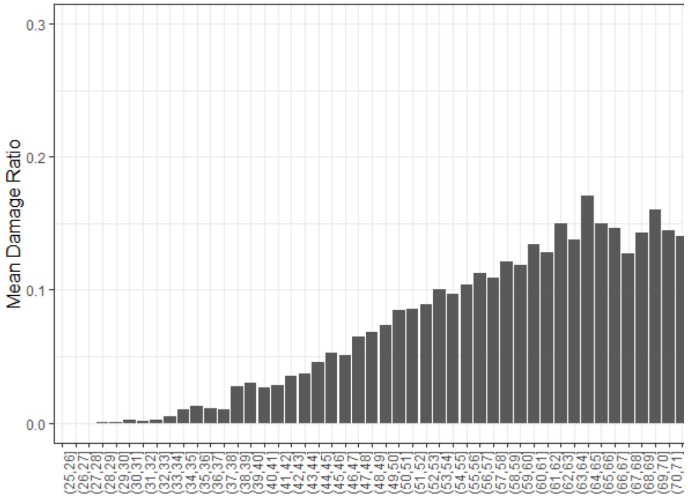
- Files of hurricanes footprints are available in the folder INPUTS/
- These files are « .tif » format files which can be read with the function raster of raster library
  - Example : odile = raster(fichier_odile_tif)
- You can also use the extract_wind(ptf,raster) function which is specified in the exercise

TP – PART 2 : VULNERABILITY MODELLING

Vulnerability curves – Empirical Method

- The empirical method builds vulnerability curves using only claims data.
- In general, we compute MDR mean by wind speed bin. We can for example use a bin of 1m/s and averaged MDR for all claims contained in each bin.

- Then, we can plot the evolution of MDR in function of wind speed and you will find a graphic similar to the one below:



TP – PART 2 : VULNERABILITY MODELLING

Vulnerability curves – Analytical method

- This method uses an analytical expression which has a similar shape than a vulnerability curve.
- These analytical expressions have several parameters which are generally optimized using claims data.
- Below, you can find 2 formulas which are usually used:
  - Emanuel (2011) expression
  - ATC-13 exporession
- Emanuel (2011) expression depends on V_{thresh} and V_{half}
- ATC-13 expression depend on 3 parameters : V_{lim}, ρ and γ.

Emanuel (2011)

$$\boldsymbol{MDR(V) = \frac{V_n^3}{1 + V_n^3} \quad with : \; V_n = \frac{MAX(V - V_{thresh}, 0)}{V_{half} - V_{thresh}}}$$

ATC-13

$$\boldsymbol{MDR(V) = 1 - 0.5^{\left(\frac{V - Vlim}{\gamma}\right)^{\rho}}}$$

TP – PART 2 : VULNERABILITY MODELLING

Vulnerability curves

Summary of the methodology used to build vulnerability curves. This is the method used in the practical work n2.

1. Merge claims and exposure database
2. Compute MDR for all insured sites. Others sites not damaged in the past have a MDR equals to 0.
3. Extract wind speed for all sites
4. Apply an empiral method to compute MDR by wind bins