M1 - Marine Physics 2019–2020

Data Analysis

#5 Statistical analysis of extreme values

1 The data

The data you will analyze are stored here

`http://mespages.univ-brest.fr/~gula/TS1/extreme_data.mat`

The file contains daily maximum temperature, precipitation, and snowfall for the 6 locations plotted in Fig.1 during 2 different time periods (1979-2001 and 2050-2060). Data have been obtained by downscaling global climate model results (CESM) using an atmospheric regional model (WRF) at 10 km resolution. Snow depth is given in water equivalent mm, which represents the depth of water that would theoretically result if you melted the entire snowpack instantaneously. To compute the actual snow depth one needs to divide by the snow density, which will take equal to 0.1.

2 Changes in precipitation

1. Plot the time-series of total precipitation (rain) in Toronto during the historical period (1979-2001), then plot its pdf.
2. Superimpose the pdf of the projected precipitation (2050-2060)

name	size	meaning	period
`tmax`	6 x n_1	daily maximum temperature (K)	1979-2001
`rain`	6 x n_1	daily total precipitation (mm)	1979-2001
`snow`	6 x n_1	daily snow precipitation (w.e. mm)	1979-2001
`tmax2050`	6 x n_2	daily maximum temperature (K)	2050-2060
`rain2050`	6 x n_2	daily total precipitation (mm)	2050-2060
`snow2050`	6 x n_2	daily snow precipitation (w.e. mm)	2050-2060

Table 1 – List of variables in the data file

3. How are the mean and standard deviation projected to change ?
4. Redo the same for the snow precipitation
5. Compute the following precipitation indices (Days with rain, Days with snow, Very wet days, Heavy P days, and Heavy snow days, see definitions in table 2) for the historical period in Toronto.
6. Compute the projected changes for the precipitation indices (in %)

3 Analysis of extreme precipitation events

1. Plot the pdf of the historical precipitation measurements in Ottawa.
2. Fit a Generalized Pareto Distribution to the data using a threshold of 5 mm. You can use [matlab : `gpfit;`, python : `scipy.stats.genpareto.fit`] to compute maximum likelihood estimates of the parameters.
3. Compute and plot the empirical return periods for the data (historical precipitation measurements in Ottawa).
4. Superimpose the empirical return periods for the fitted Generalized Pareto Distribution.
5. What is the return period of a daily precipitation > 200 mm ?
6. Compute and plot the 70% confidence interval for the fitting curve using a bootstrapping method.

TABLE 2 – Definitions of temperature and precipitation indices.

	Definitions	Units
Precipitation Indices		
Days with rain	Number of days with total precipitation (> 0.1 mm) per year	days
Days with snow	Number of days with snow (> 0.1 w.e. mm) per year	days
Very wet days (> 95^{th} percentile)	Number of days with total precipitation > 95^{th} percentile per year	days
Heavy P days (> 20 mm)	Number of days with total precipitation > 20 mm per year	days
Heavy snow days (> 5 cm)	Number of days with snow > 5 cm per year	days
Temperature Indices		
Cold days	Number of days with $T_{max} < 10^{th}$ percentile per year	days
Summer days	Number of days with $T_{max} > 25°C$ per year	days
Warm days	Number of days with $T_{max} > 90^{th}$ percentile per year	days

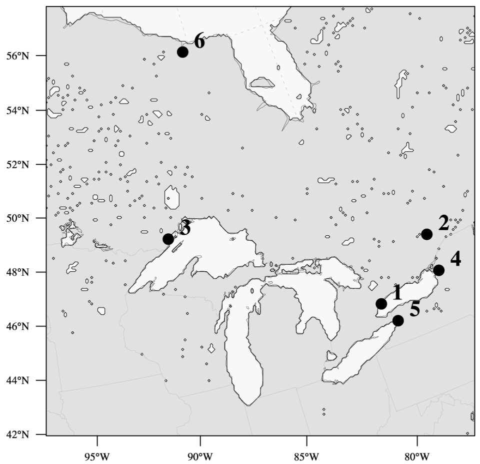

FIGURE 1 – Positions of the 6 data locations : (1) Toronto, (2) Ottawa, (3) Thunder Bay, (4) Watertown, (5) Buffalo and (6) Peawanuck.

7. Redo the same for the future period and compare the two curves (and confidence intervals). What does it say about future changes ?
8. Are the changes similar for all locations ?

4 Changes in maximum temperature

1. Plot the pdf for the maximum temperature in Buffalo during the historical period (1979-2001)
2. Superimpose the pdf of the projected maximum temperature (2050-2060)
3. How are the mean and standard deviation projected to change ?
4. Compute the temperature indices for the historical period (Cold days, Summer days, Warm days) in Buffalo.
5. Compute the projected changes for the temperature indices.

5 Analysis of extreme temperature events

1. Fit a Generalized Pareto Distribution to the data (maximum temperature in Buffalo) using a threshold of 300 K. You can use [matlab : `gpfit`;, python : `scipy.stats.genpareto.fit`] to compute maximum likelihood estimates of the parameters.
2. Compute and plot the empirical return periods for the data.
3. Superimpose the empirical return periods for the fitted Generalized Pareto Distribution.
4. Compute and plot the 95% confidence interval for the fitting curve using a bootstrapping method.
5. Redo the same for the future period and compare the two curves (and confidence intervals). What does it say about future changes ?
6. Are the changes similar for all locations ?