ISUP

Stochastic Modeling
Challenge : "Prediction of extremal precipitation"

The second assignment consists in participating in a challenge to predict spatio-temporal extremes. The aim of the challenge is to estimate high quantiles and to extrapolate them both in time and space.

About the data : The data are daily maxima of precipitation $P_{j,t}$, $j = 1, \ldots, 40$ at 40 stations active during the 44 years period from t=12/31/72 to t=12/31/16. The training sample corresponds to the 24 years period from t=12/31/72 to t=12/31/95. The aim is to predict from the training sample a quantile of level corresponding to extreme monthly precipitation over the next 20 years (the test period from t=01/01/96 to t=12/31/16) station by station. On the daily range, this event corresponds to a quantile of level

$$0.998 = 1 - 0.002 \approx 1 - \frac{1}{20 * 30 \text{ days}}.$$

Under strict stationarity, the quantile $q_{j,k}$ satisfies

$$\mathbb{P}(P_{j,t} > q_{j,k}) = 0.002, \qquad \text{day t of the test period in month k},$$

for any $j = 1, \ldots, 40$ and $k = 1, \ldots, 12$.

Evaluation : The performances of the quantile predictions $\hat{q}_{j,k}$ for any $j = 1, \ldots, 40$ and $k = 1, \ldots, 12$, are evaluated thanks to the quantile loss function

$$\ell(x, y) = \alpha(x - y)1_{x>y} + (1 - \alpha)(y - x)1_{y \geq x}, \qquad x, y \in \mathbb{R},$$

at level $\alpha = 0.998$. The quantile predictions $\hat{q}_{j,k}$ are compared with the daily maxima $P_{j,t}$ at each station $j = 1, \ldots, 40$ and each month $k = 1, \ldots, 12$:

$$S_{j,k}(\hat{q}_{j,k}) = \sum_{\text{day t of the test period in month k}} \ell(P_{j,t}, \hat{q}_{j,k}).$$

Notice that, under strict stationarity, the risk $q \to \mathbb{E}[S_{j,k}(q)]$ is minimized in $q_{j,k}$, see [1]. The final score of the predictive algorithm $\hat{q} = (\hat{q}_{j,k})$ will be the sum of the quantile losses over the stations and the months

$$S_i(\hat{q}) = \sum_{j \in C_i} \sum_{k=1}^{12} S_{j,k}(\hat{q}_{j,k}), \qquad i = 1, 2.$$

There are two different challenges based on two different sets C_i, $i = 1, 2$ of test stations :

1. The final score is the sum over C_1, the set of the 29 stations of the training sample that were still open after the training period (see below),
2. The final score is the sum over C_2, the set of the 34 stations open during the test period.

Registration : The teams have to sign in by sending an email to `olivier.wintenberger@upmc.fr`.

The email must contain the name and the names of the members of the team with their emails in CC. The link to the online storage of the data will be provided after registration. The training sample of the daily maxima of the precipitation is provided in a data frame in the file precip_sample.csv. The longitude and latitude of the stations are provided in a data frame in the file stations_coord.csv. A row number from 1 to 40 has been assigned randomly to every stations. The coordinates of the stations have been shifted to be centered at Paris. Relative distances are unchanged.

Predictions format : Each team shall provide a data frame in csv format with the first column listing the numbers associated to the stations that are, for each challenge (see as examples the benchmarks in Figures 1 and 2 below)

1. $C_1 =$ "2" "4" "5" "6" "11" "12" "13" "15" "16" "18" "19" "20" "21" "22" "23" "24" "25" "26" "28" "29" "30" "32" "33" "34" "35" "36" "38" "39" "40",
2. $C_2 = C_1 \cup$ "7" "8" "9" "10" "37".

The other 12 columns shall correspond to the monthly predictions of the 0.998-th quantile of daily maximum ($\hat{q}_{j,k}$), $k = 1, \ldots, 12$ ($k = 1$ corresponding to January).

Assignment :

1. Register your group officially as a team participating in the challenge to get the data.
2. Provide a code that creates the benchmarks (Figures 1 and 2 below) [1].
3. Provide a code that creates another prediction table for each of the 2 problems. You could use the tutorial *http ://sites.lsa.umich.edu/eva2015/wp-content/uploads/sites/44/2015/07/EVA2015.pdf*
4. Provide graphical justifications of your alternative methodology.

Références

[1] Koenker, R. (2005) *Quantile regression.* Cambridge University Press, Cambridge.

1. Any reverse-engineering algorithm is prohibited. The use of external data (i.e. data other than the provided data) is also prohibited.

"stations"	"X1"	"X2"	"X3"	"X4"	"X5"	"X6"	"X7"	"X8"	"X9"	"X10"	"X11"	"X12"
"2"	0.98	2.2	2.44	6.34	1.38	1.34	11.81	1.22	1.06	4.45	4.06	3.94
"4"	0.94	1.02	6.38	0.98	1.38	6.73	13.23	7.13	6.81	1.57	1.5	0.98
"5"	3.36	0.98	3.19	1.38	10.16	4.13	1.54	5.04	1.73	1.85	1.93	1.3
"6"	0.87	0.31	0.43	0.24	0.31	0.63	1.4	0.31	0.98	1.38	0.43	1.14
"11"	2.09	0.98	3.94	4.09	6.73	1.38	2.44	1.61	8.62	6.73	2.05	1.38
"12"	1.06	10.67	1.06	11.83	1.3	2.09	1.65	1.85	1.02	1.22	2.17	3.98
"13"	0.94	0.28	0.2	0.43	0.28	0.83	0.4	0.55	0.87	1.57	0.91	0.91
"15"	0.87	0.39	0.51	0.31	0.87	0.59	1.02	0.83	0.79	1.02	0.43	1.1
"16"	0.87	2.13	5.35	3.82	11.42	4.06	1.73	11.89	3.58	10.08	1.77	0.87
"18"	0.91	0.43	0.35	0.63	0.98	0.59	0.31	0.35	2.05	0.63	1.02	0.91
"19"	2.2	2.24	1.54	0.98	1.57	2.24	2.09	1.54	1.26	2.83	2.13	2.8
"20"	0.98	0.28	0.28	0.24	0.96	0.75	1.5	0.71	0.87	0.98	0.43	0.43
"21"	0.83	0.51	0.55	0.31	1.46	0.47	0.43	0.28	0.91	0.79	0.87	1.06
"22"	0.98	0.59	0.44	0.47	1.18	0.47	0.71	0.35	1.85	1.02	0.83	0.87
"23"	1.97	1.5	1.14	0.98	1.15	3.98	10.12	0.98	10.2	6.38	1.3	2.17
"24"	0.71	0.59	0.39	0.39	1.58	0.43	0.55	0.12	0.94	1.06	0.43	1.02
"25"	0.98	0.31	0.28	0.28	0.87	0.63	0.39	0.55	2.05	0.83	0.63	0.47
"26"	3.98	2.05	1.34	2.01	1.38	8.7	1.34	3.98	11.42	1.5	1.61	5.51
"28"	4.5	1.14	0.98	0.98	10.08	1.89	1.65	1.61	1.26	2.17	1.14	1.34
"29"	0.61	0.51	0.35	0.16	0.31	0.08	0.28	0.1	0.91	0.83	0.91	0.71
"30"	0.67	0.47	0.2	0.24	0.2	0.55	0.71	0.31	0.94	1.14	0.71	1.22
"32"	0	0	0	0	0.71	0.39	0.94	0.47	0.94	0.24	0.94	0.61
"33"	1.73	1.02	3.15	2.52	1.97	1.46	3.78	1.18	3.86	1.28	3.07	5.55
"34"	0.59	0.47	0.31	0.28	0.79	0.75	0.16	0.16	1.26	1.22	0.13	1.46
"35"	0.94	0.98	1.42	2.05	1.97	1.26	1.65	6.39	1.89	3.98	1.54	1.1
"36"	1.34	2.37	1.18	0.79	0.94	1.3	3.54	1.22	10.35	2.03	1.26	1.42
"38"	0.67	0.51	0.41	0.35	0.39	0.28	0.35	0.04	1.46	0.55	0.31	0.71
"39"	1.97	2.09	0.98	1.1	1.89	7.99	1.77	1.34	1.97	1.97	0.98	1.18
"40"	0.67	0.91	0.24	0.24	0.59	0.32	0.08	0.2	1.06	1.1	0.22	0.71

FIGURE 1 – Benchmark 1

"stations"	"X1"	"X2"	"X3"	"X4"	"X5"	"X6"	"X7"	"X8"	"X9"	"X10"	"X11"	"X12"
"2"	0.98	2.2	2.44	6.34	1.38	1.34	11.81	1.22	1.06	4.45	4.06	3.94
"4"	0.94	1.02	6.38	0.98	1.38	6.73	13.23	7.13	6.81	1.57	1.5	0.98
"5"	3.36	0.98	3.19	1.38	10.16	4.13	1.54	5.04	1.73	1.85	1.93	1.3
"6"	0.87	0.31	0.43	0.24	0.31	0.63	1.4	0.31	0.98	1.38	0.43	1.14
"11"	2.09	0.98	3.94	4.09	6.73	1.38	2.44	1.61	8.62	6.73	2.05	1.38
"12"	1.06	10.67	1.06	11.83	1.3	2.09	1.65	1.85	1.02	1.22	2.17	3.98
"13"	0.94	0.28	0.2	0.43	0.28	0.83	0.4	0.55	0.87	1.57	0.91	0.91
"15"	0.87	0.39	0.51	0.31	0.87	0.59	1.02	0.83	0.79	1.02	0.43	1.1
"16"	0.87	2.13	5.35	3.82	11.42	4.06	1.73	11.89	3.58	10.08	1.77	0.87
"18"	0.91	0.43	0.35	0.63	0.98	0.59	0.31	0.35	2.05	0.63	1.02	0.91
"19"	2.2	2.24	1.54	0.98	1.57	2.24	2.09	1.54	1.26	2.83	2.13	2.8
"20"	0.98	0.28	0.28	0.24	0.96	0.75	1.5	0.71	0.87	0.98	0.43	0.43
"21"	0.83	0.51	0.55	0.31	1.46	0.47	0.43	0.28	0.91	0.79	0.87	1.06
"22"	0.98	0.59	0.44	0.47	1.18	0.47	0.71	0.35	1.85	1.02	0.83	0.87
"23"	1.97	1.5	1.14	0.98	1.15	3.98	10.12	0.98	10.2	6.38	1.3	2.17
"24"	0.71	0.59	0.39	0.39	1.58	0.43	0.55	0.12	0.94	1.06	0.43	1.02
"25"	0.98	0.31	0.28	0.28	0.87	0.63	0.39	0.55	2.05	0.83	0.63	0.47
"26"	3.98	2.05	1.34	2.01	1.38	8.7	1.34	3.98	11.42	1.5	1.61	5.51
"28"	4.5	1.14	0.98	0.98	10.08	1.89	1.65	1.61	1.26	2.17	1.14	1.34
"29"	0.61	0.51	0.35	0.16	0.31	0.08	0.28	0.1	0.91	0.83	0.91	0.71
"30"	0.67	0.47	0.2	0.24	0.2	0.55	0.71	0.31	0.94	1.14	0.71	1.22
"32"	0	0	0	0	0.71	0.39	0.94	0.47	0.94	0.24	0.94	0.61
"33"	1.73	1.02	3.15	2.52	1.97	1.46	3.78	1.18	3.86	1.28	3.07	5.55
"34"	0.59	0.47	0.31	0.28	0.79	0.75	0.16	0.16	1.26	1.22	0.13	1.46
"35"	0.94	0.98	1.42	2.05	1.97	1.26	1.65	6.39	1.89	3.98	1.54	1.1
"36"	1.34	2.37	1.18	0.79	0.94	1.3	3.54	1.22	10.35	2.03	1.26	1.42
"38"	0.67	0.51	0.41	0.35	0.39	0.28	0.35	0.04	1.46	0.55	0.31	0.71
"39"	1.97	2.09	0.98	1.1	1.89	7.99	1.77	1.34	1.97	1.97	0.98	1.18
"40"	0.67	0.91	0.24	0.24	0.59	0.32	0.08	0.2	1.06	1.1	0.22	0.71
"7"	1.34	1.22	1.33	1.37	1.98	1.75	2.12	1.91	2.53	1.98	1.25	1.49
"8"	1.34	1.22	1.33	1.37	1.98	1.75	2.12	1.91	2.53	1.98	1.25	1.49
"9"	1.34	1.22	1.33	1.37	1.98	1.75	2.12	1.91	2.53	1.98	1.25	1.49
"10"	1.34	1.22	1.33	1.37	1.98	1.75	2.12	1.91	2.53	1.98	1.25	1.49
"37"	1.34	1.22	1.33	1.37	1.98	1.75	2.12	1.91	2.53	1.98	1.25	1.49

Figure 2 – Benchmark 2