

# Report for Project, ASRM 569

Nov. 25, 2020

**Project title:** Extreme Precipitation Analysis and Modeling of Floods Damage

**Team Member:** Panyi Dong

## 1 Introduction

Flood is one of most frequent catastrophes mankind currently facing. Additionally, due to the climate change during the past decades, all categories of catastrophes demonstrate drastically increase in frequency and occurrences of extreme cases, which results in tremendous amount of economic, properties and lives loss and significantly impact the trend of development of flood-suffering regions. Thus, it's essential to learn how to quantify the damage of floods not only on the perspective of insurance companies but also beneficial for house-owners to raise the idea of flood precautions, since, both market-based insurance feedback [1] and adaptation methods [2] are the keys to reduce the flood impact and keep the society continuously moving forward.

In this project, I would like to estimate the parameters of extreme rainfall using at-site statistics by the method of extreme value analysis and plot the map of hazard of extreme rainfall. After at-sites analysis, I con-

struct a explicit probability model using spatial regression techniques [3]. And with the knowledge of hazard of rainfall, I can use the multidisciplinary models based on actuarial science, physics and engineering to calculate the physical, economic damage of floods to private properties [4]. The model traces back to observed precipitation statistics, constructs the surface cutoff and river flow using physics differential equation(Shallow water equation) and calculate the damage of overflow under extreme cases, which, then can be used to determine premiums of such flow insurance. These statistics can then help insurance companies determine the flood insurance market and help private house-owners prepare for flows.

## 2 Part I: Extreme Precipitation Analysis

### 2.1 Method

#### 2.1.1 Visualization of Dataset

The dataset of precipitation is taken from Santander Meteorology

Group [5] [6]. The dataset contains the daily precipitation from 1/1/1950 to 12/31/2015 (in total of 66 years or 24106 days) with 0.1 degree (around 10km geographically) precision scale of the entire Spain and neighboring sea area. Based on the map, it's a square area with bottom left point of 36.0°N - 9.3°E and the top right point of 43.8°N 4.4°E (in total of 10902 points).

For this part, I try to visualize the dataset, since it's easier to understand how precipitation distributed on the map. Three maps are constructed, 1. average annual precipitation map; 2. average annual daily maximum precipitation map; 3. the average ratio of daily maxima and annual precipitation map. And there are two ways of such visualization, one is simply use the idea of heatmap and color all values at each points, which can be visually acceptable if the requirements of map is not that precise or the map hasn't been stretched too much; another one is to use the method of Gaussian kernel, which means every individual point value follow Gaussian distribution spatially and add up all overlapped value at chosen coordinates (illustration as seen in fig1).

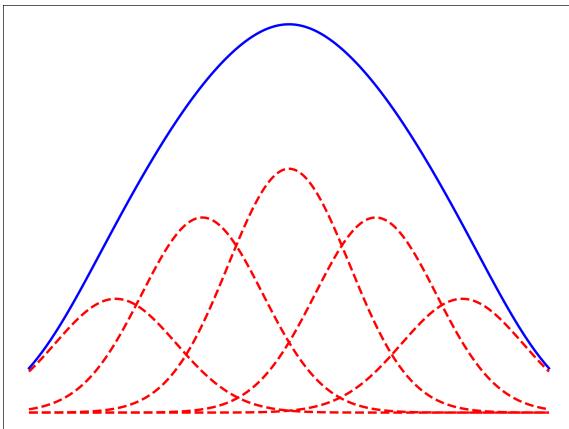


Figure 1: The Illustration of Gaussian Kernel

## 2.1.2 Extreme Precipitation Analysis [3]

For this part, I select 300 most-populated cities in Spain and by the precision scale of the dataset, 53 cities overlapped and 17 cities out of the boundaries of the database (these cities distributed a bit north of North Africa), which means only 230 cities are taken into the extreme precipitation analysis. While analyzing the extreme precipitation data, three more cities are consist of all 0s. I believe it's because these cities are so close to the land boundaries that with the precision scale of the database, they are considered as in-sea areas, which leads to the all 0s. And later in the report, I call these 230 cities as stations. Fig8 shows how these 230 stations distributed.



Figure 2: Stations Distribution

Then, I try to fit the at-sites extreme precipitation data to Generalized Pareto(GP) distribution, and examine the goodness-of-fit.

For GP distribution, cumulative distribution function is

$$F(x) = \begin{cases} 1 - (1 - \kappa \frac{x}{\alpha})^{1/\kappa}, & \kappa \neq 0 \\ 1 - e^{-x/\alpha}, & \kappa = 0 \end{cases} \quad (2.1)$$

where  $\alpha$  is the scale parameter and  $\kappa$  is the shape parameter. For  $\kappa \leq 0$ , the range  $x > 0$ ; for  $\kappa > 0$ , the range  $x \in [0, \frac{\alpha}{\kappa}]$ .

And the probability distribution function has form

$$f(x) = \begin{cases} \frac{1}{\alpha} (1 - \kappa \frac{x}{\alpha})^{1/\kappa-1}, & \kappa \neq 0 \\ e^{-x/\alpha}/\alpha, & \kappa = 0 \end{cases} \quad (2.2)$$

### a.Examine the Data

To begin with, I have to examine how the data is distributed and if it is possible to fit the data to a GP distribution. One frequently used method is to examine the mean excess plot of the data.

By the definition of mean excess function,  $e(t) = E(X - t | X > t)$  and incorporate it with the GP distribution. I get

$$\begin{aligned} e(t) &= \frac{\int_t^{+\infty} (x - t) \frac{1}{\alpha} (1 - \kappa \frac{x}{\alpha})^{1/\kappa-1} dx}{\int_t^{+\infty} \frac{1}{\alpha} (1 - \kappa \frac{x}{\alpha})^{1/\kappa-1} dx} \\ &= \frac{\alpha}{1 + \kappa} - \frac{\kappa}{1 + \kappa} t \end{aligned}$$

which means if the data is possible for fitting to GP distribution, mean excess plots should linearly increase as the threshold of data increases.

### b.GP Distribution Fitting

In this project the GP distribution fitting employs the method of probability-weighted moments(PWM). Compared to the method of maximum likelihood estimation(MLE) and method of moments estimation(MME), PWM method can be applied to dataset with large data samples, which in our case of studying

extreme precipitation distribution, is crucially helpful.

By the method of PWM, if the random sample  $X$  has cumulative distribution function  $F(X)$ , we can define [7]

$$M_{p,r,s} = E[X^p(F(X))^r(1 - F(X))^s] \quad (2.3)$$

and

$$\beta_s = M_{1,0,s} \quad (2.4)$$

Then, combining (2.3) and (2.4),

$$\beta_s = E[X(1 - F(X))^s] \quad (2.5)$$

For the case of GP distribution, calculate (2.5) as

$$\beta_s = \frac{\alpha}{(s+1)(s+1+\kappa)} \quad (2.6)$$

Then, based on (2.6), for  $s = 0$ , and  $s = 1$ ,

$$\beta_0 = \frac{\alpha}{1 + \kappa} \quad (2.7)$$

$$\beta_1 = \frac{\alpha}{2(2 + \kappa)} \quad (2.8)$$

Solving for (2.7) and (2.8),

$$\kappa = \frac{\beta_0}{\beta_0 - 2\beta_1} - 2 \quad (2.9)$$

$$\alpha = \frac{2\beta_0\beta_1}{\beta_0 - 2\beta_1} \quad (2.10)$$

And for  $n$  observations, empirical probability weighted moments have form of

$$\beta_s = \frac{1}{n} \sum_{j=1}^n \frac{(n-j)(n-j-1)\dots(n-j-r+1)}{(n-1)(n-2)\dots(n-r)} x_{j,n} \quad (2.11)$$

where  $x_{j,n}$  is  $j$ -th largest value of the sorted  $n$  observations.

Then I get the PWM estimators of scale parameter,  $\alpha$  and shape parameter  $\kappa$ .

### c.Examine Goodness of Fit

After calculating the PWM estimators, I try to examine the goodness

of fit of these estimators by Probability Plot (PP), Probability Plot–Weighted Mean Bias Error (PPWMBE), Probability Plot–Weighted Root-Mean-Square Error (PPWRMSE) and Probability Plot–Weighted Correlation Coefficient(PPWCC).

For  $n$  observations of  $\{y_i\}$  ( $i=1,2,\dots,n$ ) extreme precipitation events, where  $y_i$  is the exceedance of precipitation over threshold  $x_0$ , which can be written as  $y_i = x_i - x_0$ . Empirical cumulative distribution function of extreme precipitation events can be approached by

$$EmpCDF(j) = \frac{j + A}{n + B} \quad (2.12)$$

for this case,  $A = -0.15$ ,  $B = 0$  [8] is recommended.

Weighing function

$$\omega_j = \frac{1}{1 - EmpCDF(j)} \quad (2.13)$$

The estimated observations can be calculated using the quantile function of the empirical distribution. Solving (2.1) and replacing  $x_{j,n}$  by  $Q(p)$  or  $\hat{y}_j$  in this case, and replacing  $F(x_{j,n})$  by  $EmpCDF(j)$ , we can get

$$\hat{y}_j = \frac{\hat{\alpha}}{\hat{\kappa}}[1 - (1 - EmpCDF(j))^{\hat{\kappa}}] \quad (2.14)$$

The weighted mean of the observations and mean of empirical estimated observations are

$$\bar{y} = \frac{1}{n \sum_{j=1}^n \omega_j} \sum_{j=1}^n y_j \omega_j \quad (2.15)$$

$$\bar{\hat{y}} = \frac{1}{n \sum_{j=1}^n \omega_j} \sum_{j=1}^n \hat{y}_j \omega_j \quad (2.16)$$

The Probability Plot can compare the original observations and the empirical estimated observations so I can

visually and qualitatively understand how good our fitting is.

The Probability Plot–Weighted Mean Bias Error (PPWMBE) uses the definition

$$PPWMBE = \frac{1}{n \sum_{j=1}^n \omega_j} \sum_{j=1}^n (\hat{y}_j - y_j) \omega_j \quad (2.17)$$

The Probability Plot–Weighted Root-Mean-Square Error (PPWRMSE) can be written as

$$PPWRMSE = \sqrt{\frac{1}{n \sum_{j=1}^n \omega_j} \sum_{j=1}^n (\hat{y}_j - y_j)^2 \omega_j} \quad (2.18)$$

And Probability Plot–Weighted Correlation Coefficient(PPWCC) is defined as

$$PPWR^2 = \frac{[\sum_{j=1}^n (y_j \hat{y}_j - n \bar{y} \bar{\hat{y}}) \omega_j]^2}{[\sum_{j=1}^n (y_j^2 - n \bar{y}^2) \omega_j][\sum_{j=1}^n (\hat{y}_j^2 - n \bar{\hat{y}}^2) \omega_j]} \quad (2.19)$$

These error statistics can help us quantitatively assess the goodness-of-fit of our estimators.

### 2.1.3 Spatial Probability Distribution

To predict extreme rainfall events at random coordinates, I have to determine the spatial distribution of the parameters of the precipitation distribution (here,  $\alpha$ ,  $\kappa$ ,  $x_0$  are the parameters of interest). In this project, the method of multiple regression is used to predict the value of climatic variable.

$$z(x) = b_0 + b_1 P_1 + \dots + b_n P_n \quad (2.20)$$

where  $z(x)$  is the predicted value at point  $(x)$ ,  $b_0$ ,  $b_1$ , ...,  $b_n$  are the multiple regression coefficients,  $P_1$ ,  $P_2$ , ...,  $P_n$  are the values of the different independent variables at point  $(x)$ .

In this project, I apply the independent variables as follows:

1. The first order of trend surface. I use  $x^{(1)} = \text{lat} + \text{lon}$ .
2. The second order of trend surface. I use  $x^{(2)} = (\text{lat} + \text{lon})^2$ .
3. Mean elevation within the circle of 0.2 degree (around 20km),  $x^{(3)}$ .
4. Mean slope within the circle of 0.2 degree (mean elevation difference),  $x^{(4)}$ .
5. Mean relief energy within the circle of 0.2 degree (the max elevation - elevation at point),  $x^{(5)}$ .
6. Barrier effect to the four cardinal directions within the range of 0.2 degree (max elevation within the radius of 0.2 degree and at the di-

rection of north/south/east/east - elevation at point),  $x^{(6)}$ .

Then, if I replace  $z(x)$  with  $y$ , the multiple regression equation can be written as  $y = b_0 + b_1x_i^{(1)} + \dots + b_6x_i^{(6)}$ . Respectively, the mean square of regression error,  $L = \sum_{i=1}^n (y_i - b_0 - b_1x_i^{(1)} - \dots - b_6x_i^{(6)})^2$ . Based on the idea of least mean square of regression error, I get

$$\begin{cases} \frac{\partial L}{\partial b_0} = -\sum_{i=1}^n 2(y_i - b_0 - b_1x_i^{(1)} - \dots - b_6x_i^{(6)}) = 0 \\ \frac{\partial L}{\partial b_1} = -\sum_{i=1}^n 2x_i^{(1)}(y_i - b_0 - b_1x_i^{(1)} - \dots - b_6x_i^{(6)}) = 0 \\ \dots \\ \frac{\partial L}{\partial b_6} = -\sum_{i=1}^n 2x_i^{(6)}(y_i - b_0 - b_1x_i^{(1)} - \dots - b_6x_i^{(6)}) = 0 \end{cases} \quad (2.21)$$

Additionally, write the equation in matrix form

$$\begin{pmatrix} n & \sum_{i=1}^n x_i^{(1)} & \sum_{i=1}^n x_i^{(2)} & \dots & \sum_{i=1}^n x_i^{(6)} \\ \sum_{i=1}^n x_i^{(1)} & \sum_{i=1}^n (x_i^{(1)})^2 & \sum_{i=1}^n x_i^{(1)}x_i^{(2)} & \dots & \sum_{i=1}^n x_i^{(1)}x_i^{(6)} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_i^{(6)} & \sum_{i=1}^n x_i^{(6)}x_i^{(1)} & \sum_{i=1}^n x_i^{(6)}x_i^{(2)} & \dots & \sum_{i=1}^n (x_i^{(6)})^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \dots \\ b_6 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i^{(1)}y_i \\ \dots \\ \sum_{i=1}^n x_i^{(6)}y_i \end{pmatrix} \quad (2.22)$$

With the spatial prediction of values of climatic variable,  $\hat{\alpha}$ ,  $\hat{\kappa}$  and  $\hat{x}_0$ . I can construct the predicted distribution of extreme precipitation using

$$P(U) = \frac{\hat{\alpha}}{\hat{\kappa}}(1 - (1 - U)^{\hat{\kappa}}) + \hat{x}_0 \quad (2.23)$$

where  $P(U)$  is the predicted individual extreme precipitation events and  $U$  is a uniformly distributed random variable with interval of  $[0, 1]$ . And to examine the goodness of fit of such prediction, frequency plot and Probability Plots are constructed.

## 2.2 Results and Discussions

For the method of heatmap, three maps are shown as fig3, fig4, fig5.

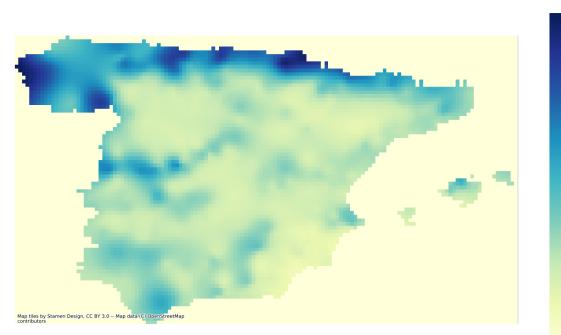


Figure 3: The Heatmap of Average Annual Precipitation

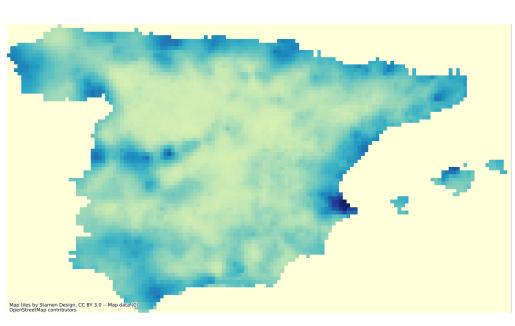


Figure 4: The Heatmap of Average Annual Daily Maximum Precipitation

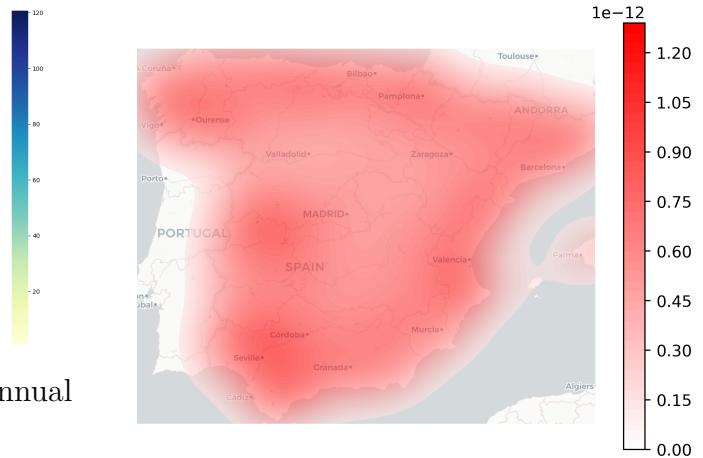


Figure 7: The Gaussian Kernel Map of Average Annual Daily Maximum Precipitation

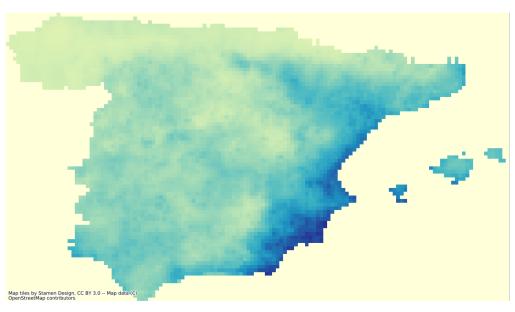


Figure 5: The Heatmap of Ratio of Daily Maxima to Annual Precipitation

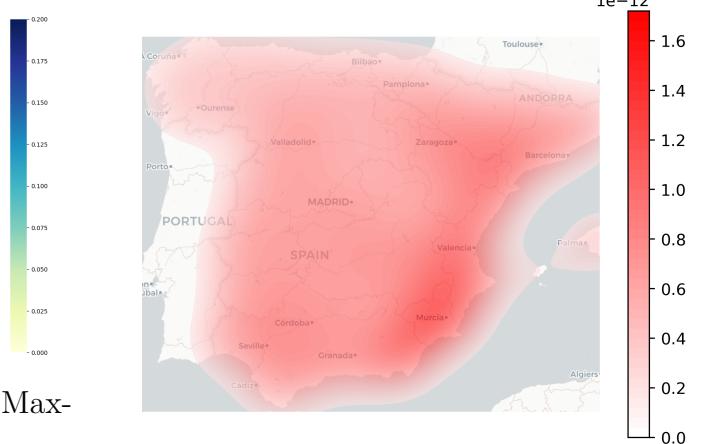


Figure 8: The Gaussian Kernel Map of Ratio of Daily Maxima to Annual Precipitation

For the method of Gaussian kernel, three maps are shown as fig5, fig6, fig8.



Figure 6: The Gaussian Kernel Map of Average Annual Precipitation

Using the calculation of Gaussian kernel, we are able to compare the precipitation data of any randomly chosen points instead of approximated coordinates (by the limitation of the database). But since the values at the points has changed after the Gaussian kernel calculation, it can't reflect the precipitation data itself but only shows the comparison of intensity. From the heatmap of the ratio of daily maxima to annual precipitation, I notice that at some places, the annual daily maximum precipitation is nearly 20 percent of the annual precipitation, which indicates that at Mediterranean area, at least in Spain, extreme rainfall events occur frequently, and concentrate on

period of few days. This phenomenon confirms our reason of choosing this area for extreme precipitation analysis.

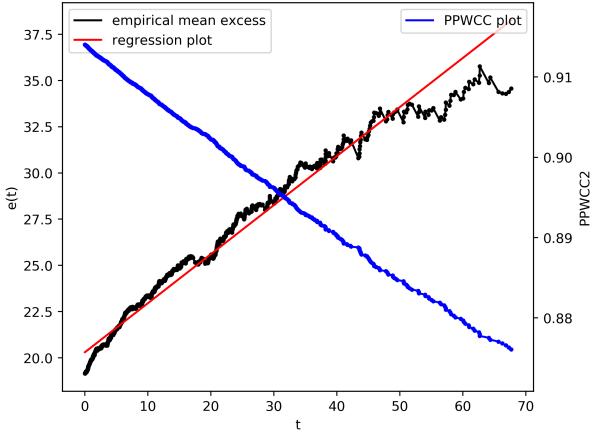


Figure 9: The mean excess plot and PPWCC plot of Calp

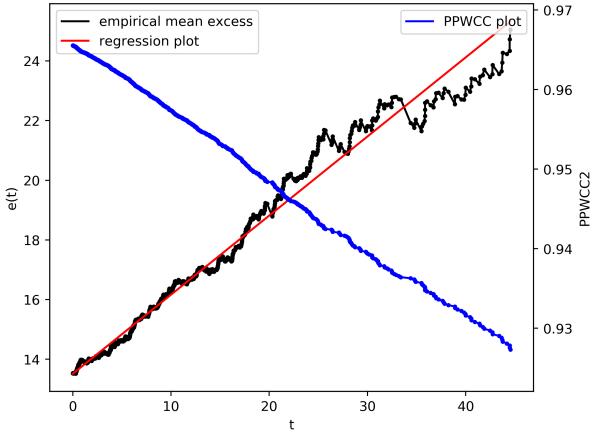


Figure 10: The mean excess plot and PPWCC plot of El prat de llobregat

To determine whether the dataset is possible for GP distribution fitting,

mean excess plots and PPWCC plots are constructed. As Fig9 and Fig10, the mean excess plot and PPWCC plot of two randomly chosen stations, Calp and El prat de llobregat are shown. Since the tail behavior of empirical mean excess function converges to 0, which is not ideal for demonstration and regression, the largest five percent precipitation data is omitted. Based on the trend of these two plots, mean excess increases almost perfectly linearly as the threshold  $t$  increases and PPWCC decreases slightly as threshold  $t$  increases but still maintains high value near perfect correlation 1. These plots help convince that the precipitation data can be approached by GP distribution and even though as the  $t$  increases, the tail fluctuates a bit, it still is in appropriate agreement with GP distribution behaviors.

Then, the GP distribution fitting using PWM method and the goodness-of-fit examination are performed. Distribution parameters and goodness-of-fit examination of 10 randomly chosen stations are shown in Table 1.

Table 1: Values of climatic variable and goodness of fit examination for 10 randomly chosen stations

Station Names	$\alpha$	$\kappa$	$x_0$	PPWMBE	PPWRMSE	PPWCC
Madrid	6.70915	-0.0018349	8.10879	-0.0047909	0.22894	0.99709
Barcelona	9.87890	-0.13866	8.95497	0.032329	1.52909	0.97069
Valencia	8.43013	-0.33039	6.28861	0.13572	6.53462	0.90317
Sevilla	11.31092	-0.15666	12.13588	0.041630	1.91990	0.98103
Zaragoza	5.97986	-0.19601	5.70100	0.033162	1.54059	0.96513
Malaga	11.44522	-0.10627	9.86191	0.014735	0.76286	0.9882
Murcia	6.81192	-0.24383	3.91237	0.038080	1.80478	0.98548
Palma	12.12907	-0.19254	12.62173	0.052719	2.60763	0.95790
Bilbao	10.35019	-0.06852	18.95900	-0.0082901	0.44096	0.99576
Valladolid	5.42917	-0.07094	7.26499	0.0052749	0.27016	0.99165

GP distribution possesses one important property that the shape parameter remains constant regardless of the threshold value. Thus, we should apply a lower threshold for distribution fitting but not so small that tail behaviors vanish. In the simulation, 1200 largest precipitation data is selected among 24106 total data (around 5%). The threshold of parameter of interest is thus being determined. For  $\kappa < 0$ , the range of GP distribution is  $[0, +\infty)$ , and the around 10mm extreme precipitation thresholds shown in Table1 correspond to the classification of moderate rain.

For PPWMBE, since it is category of mean bias error, the closer to 0, the better the estimation; for PPWRMSE, since it is category of mean square error, the closer to 0, the better the estimation; for PPWCC, it is category of correlation coefficient, thus, the closer to 1, the better the estimation. In goodness-of-fit examination, at-sites parameters of all 227 stations are being examined, and all mean bias examination of PPWMBE remains in the scale of 0.01; majority of PPWRMSE keeps in the range of 0.5 but some reaches the mean square error of 8; and in case of PPWCC, most correlation coefficients are at the level above 0.95. Thus, overall, the performance of GP distribution fitting reaches our expectations.

The Probability Plots are constructed for each station. Two of the randomly chosen stations' Probability Plots of observed extreme precipitation and fitted extreme precipitation are shown as Fig11 and Fig12. The fitted extreme precipitation matches ob-

served data well and the regression plots of these Probability Plots are almost overlapped with ideal probability plots. Nevertheless, there are some stations where the tail fittings are not good enough due to the instability of tail behaviors.

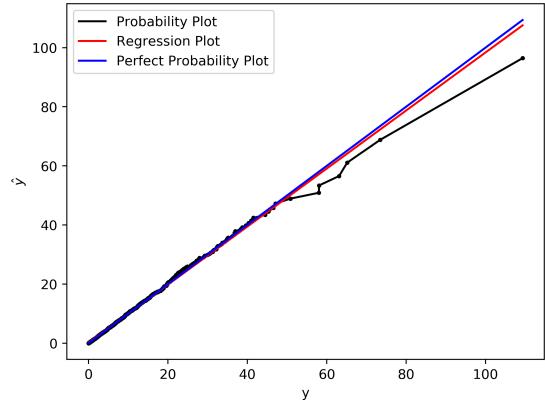


Figure 11: Probability Plot of observed/fitted extreme precipitation at Arteixo

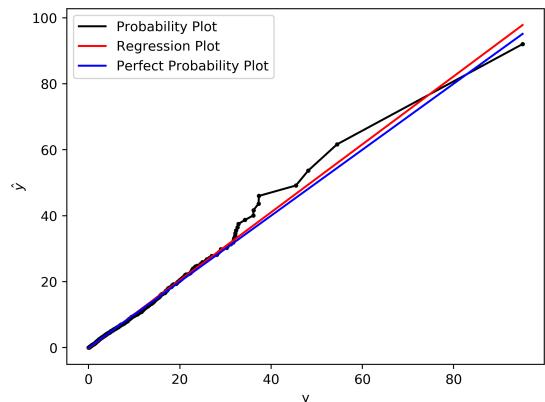
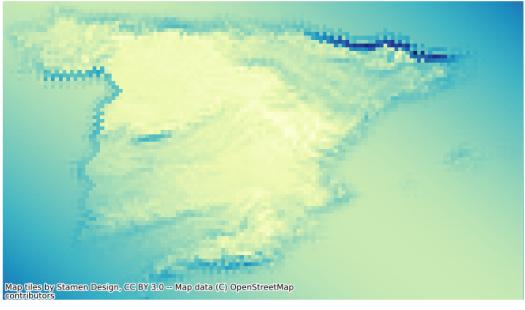


Figure 12: Probability Plot of observed/fitted extreme precipitation at Ciudadreal

With the parameter estimations at stations, I can then apply multiple regression method to construct the spatial distribution of these parameters. The three spatial distribution heatmaps are shown as Fig13, Fig14 and Fig15. Here, the method of heatmap is used. Specifically, the distribution of  $\kappa$  in the simulation are in the range of  $[-0.3, 0.2]$ , which are in the range of  $[-0.5, 0.5]$ , the restriction range provided by Hoksing [7], for

both theoretical and practical applications. One thing to be stated is that since the boundary is not restricted in the simulation, the neighboring sea areas are included in the multiple regression process.



With the spatial distribution of these parameters, we can predict extreme precipitation of any randomly chosen coordinates. We can apply the predicted parameters to construct randomly GP-distributed precipitation events. Additionally, with 1200 extreme predicted precipitation events constructed, we can use frequency plots to qualitatively compare the predicted/observed data and Probability Plots for quantitative comparison. As shown in Fig16 and Fig17, the predicted extreme precipitation frequency plot is close to the observed data, and the Probability Plot of these two extreme precipitation data is shown as Fig18. Nevertheless, there are coordinates where the predicted extreme precipitation data only follows the trends but not in acceptable error range. I think it may be resulted from the choice of our independent variables of multiple regression model and the lack of high resolution geographic, topographic database. The introduction of potential radiation and other climate-related factors will increase the accuracy of this prediction.

Figure 13: Spatial distribution of scale parameter  $\alpha$

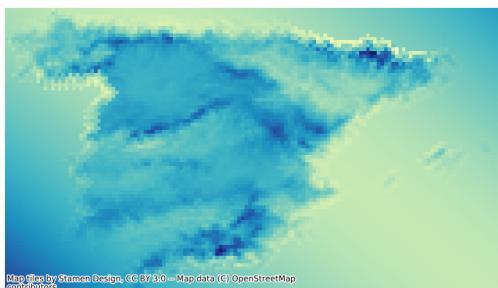


Figure 14: Spatial distribution of shape parameter  $\kappa$

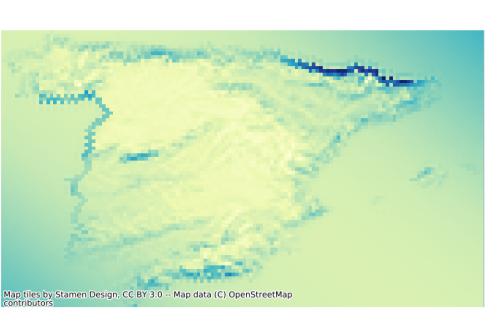


Figure 15: Spatial distribution of threshold  $x_0$

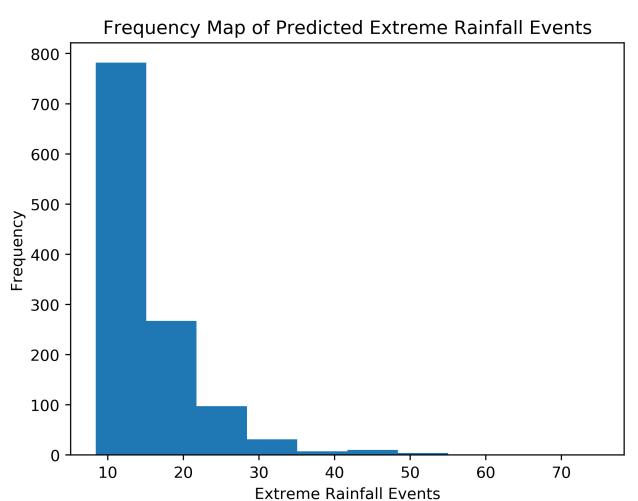


Figure 16: Frequency plot of predicted extreme precipitation

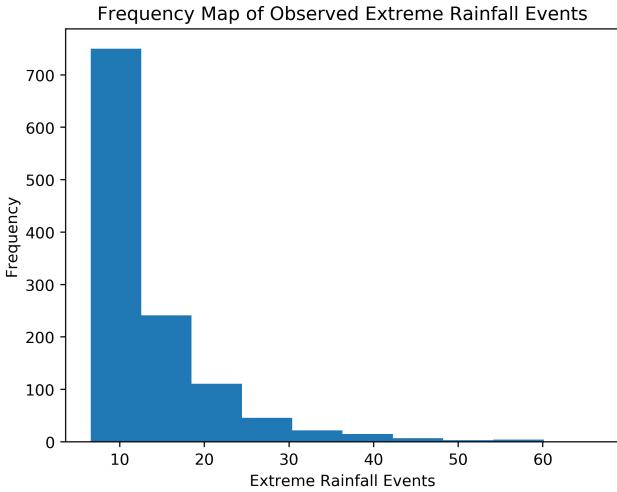


Figure 17: Frequency plot of observed extreme precipitation

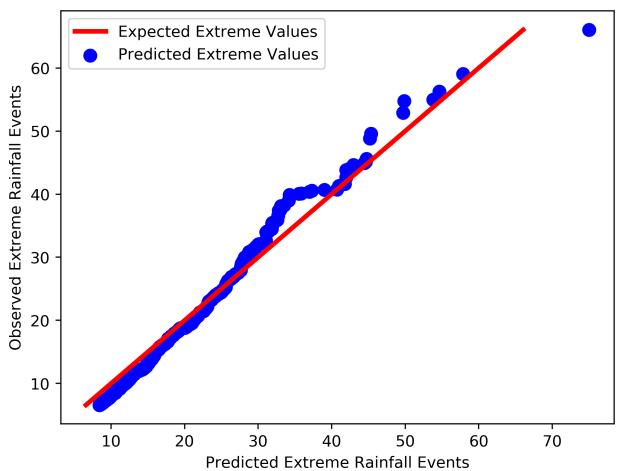


Figure 18: Probability plots of observed/predicted extreme precipitation

### 2.3 Conclusions

In this project, our goal is to predict the extreme precipitation events distribution at any random coordinates in Spain with the 50-year-long daily precipitation data at some given stations. I first construct the heatmap and Gaussian kernel map of mean annual precipitation, mean annual daily maximum precipitation, the mean ratio of daily maximum to annual precipitation with the daily Spain precipitation dataset, which confirms our expectation that extreme rainfall events occur frequently in Spain. Additionally, I examine the validity of GP distri-

bution assumption using mean excess plot. As theoretical GP distribution suggested, if the data follows GP distribution, as threshold  $t$  increases, mean excess function  $e(t)$  linearly increases. By empirical mean excess calculation and omitting the largest five percent data, mean excess function increases approximately linearly and Probability Plot-Weighted Correlation Coefficient (PPWCC) remains high level around 1, which confirms that our approach of GP distribution fitting is viable.

After examining the data, I apply the method of Probability Weight Moments (PWM), comparing the weighted empirical mean and theoretical weighted mean to calculate the scale parameter  $\alpha$  and shape parameter  $\kappa$  at 230 most populated cities in Spain (as our observation stations). The goodness-of-fit of these PWM estimations can be examined by Probability Plot (PP), Probability Plot-Weighted Mean Bias Error (PPWMBE), Probability Plot-Weighted Root-Mean-Square Error (PPWRMSE) and Probability Plot-Weighted Correlation Coefficient(PPWCC), which by our calculation, all demonstrate that our distribution fittings are within acceptable error. With the distribution fitting at stations, I then apply the multiple regression method using geographic data, coordinates and elevations to construct the spatial distribution of values of climatic parameters. Finally, we are able to predict extreme precipitation events at any randomly chosen coordinates. By the examination of frequency map and Probability Plot, some of the predicted data fits well

with original observed precipitation data and some of them are only meaningful for qualitative analysis.

One of the possible reason the distribution fitting demonstrates almost perfect correlation relationship is that the precipitation database I employed in this project is spatially interpolated one. Thus, spatial interpolation relationships are already contained in the database. For further modifications, the elevation dataset applied in this project is so primitive that we are not able to calculate independent variables of multiple regression precisely. Additionally, one significant factor, potential radiation that drastically influence the precipitation distribution is not considered in this project. Thus, the application of digital elevation model (DEM) at high resolution using the Mi-raMon geographic information system (GIS) software will increases the precision scale of our prediction on extreme precipitation events.

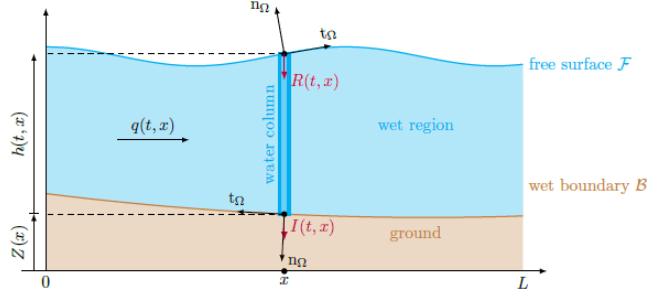


Figure 19: Illustration for River Flow [9]

The flow model is as shown in Fig19. The system contains the wet region, or the water flow in our simulation and ground, or topographic of river bed, which separated by free surface between air and water flow and wet boundary between water flow and river bed. The St. Venant equations can be derived from Navier-Stokes equations combined with boundary conditions of the kinematic friction law and permeable stress condition. The one-dimensional St. Venant equations considering recharge (fluid adding to the system by free surface) and infiltration (fluid leaving the system through It boundary) can then be written as:

$$\partial_t h + \partial_x q = S = R - I \quad (3.1)$$

$$\partial_t q + \partial_x [q^2/h + g h^2/2] = -gh \partial_x Z + S \frac{q}{h} - (k_+(R) + k_-(I) + k_0(\mu)) \frac{q}{h} \quad (3.2)$$

where  $h$  stands for flow height,  $q$  stands for flow momentum with  $q = h\mu$ ,  $\mu$  stands for flow velocity,  $R$  stands for recharge rate,  $I$  stands for infiltration rate with  $S = R - I$ ,  $Z$  stands for topography of ground,  $k_+(R)$  for recharge friction coefficient,  $k_-(I)$  for infiltration friction coefficient, and  $k_0(\mu)$  for kinetic friction coefficient.

For numerical method, we introduce kinetic scheme for St. Venant

## 3 Part II: Quantifying Flood damage [9] [10]

### 3.1 Method

For this part, due to time limits, this project focus on quantifying river flow at certain time/coordinates. For river flow, one frequently used hydro-mechanical model is St. Venant system, or sometimes called Shallow Water Equations.

system. We introduce kinetic density function  $M(t, x, \xi)$  as density of particles with speed  $\xi$  at time-space coordinates  $(t, x)$ . Additionally, the density function follows macroscopic-microscopic relations,

$$\int_R \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix} M(t, x, \xi) d\xi = \begin{pmatrix} h(t, x) \\ h(t, x)\mu(t, x) \\ h(t, x)\mu(t, x) + \frac{g}{2}h(t, x)^2 \end{pmatrix} \quad (3.3)$$

Defined nonlinear flux integral operator  $\hat{W}$  as

$$\hat{W}(t, x) = Z(x) + \int_0^x \frac{(k_+(R) + k_-(I) + k_0(\mu))\mu}{gh}(t, s) ds \quad (3.4)$$

To construct the numerical iteration steps of St. Venant equations for

cell  $c_i \in [x_{i-0.5}, x_{i+0.5}]$ , where  $x_i = i\Delta x$ ,  $\Delta x = \text{len}(c_i)$  at time  $t_n = n\Delta t$  (which means  $i$  stands for space coordinates and  $n$  stands for time), we set the vector of our interest  $U_i^n = \int_R \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix} M_i^n(\xi) d\xi = \begin{pmatrix} h_i^n \\ h_i \mu_i^n \\ h_i \mu_i^n \end{pmatrix}$ .

The discretized numerical scheme

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\Delta x} (F_{i+0.5}^n - F_{i-0.5}^n) + \Delta t \begin{pmatrix} S_i^n \\ S_i^n \mu_i^n \end{pmatrix} \quad (3.5)$$

where  $\frac{\text{CFL}}{\text{CFL} \cdot \Delta x}$  time step  $\Delta t = \frac{\text{CFL} \cdot \Delta x}{\max_i(|\mu_i^n| + \sqrt{2gh_i^n})}$ , and numerical flux  $F_{i\pm 0.5}^n = \int_R \xi \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix} M_{i\pm 0.5}^\mp(\xi) d\xi$  can be written as

$$F_{i+0.5}^n = \int_0^{+\infty} \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} M_i^n(\xi) d\xi + \int_{-\sqrt{2g\Delta W_{i+0.5}^n}}^0 \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} M_i^n(-\xi) d\xi + \int_{-\infty}^{-\sqrt{2g\Delta W_{i+0.5}^n}} \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} M_i^n(-\sqrt{|\xi|^2 - 2g\Delta W_{i+0.5}^n}) d\xi \quad (3.6)$$

$$F_{i-0.5}^n = \int_{-\infty}^0 \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} M_i^n(\xi) d\xi + \int_0^{\sqrt{2g\Delta W_{i-0.5}^n}} \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} M_i^n(-\xi) d\xi + \int_{\sqrt{2g\Delta W_{i-0.5}^n}}^{\infty} \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} M_i^n(\sqrt{|\xi|^2 - 2g\Delta W_{i-0.5}^n}) d\xi \quad (3.7)$$

where semi-discretized kinetic energy

$$M_i^n = \sqrt{h_i^n} \chi \left( \frac{\xi - \mu_i^n}{\sqrt{h_i^n}} \right) \quad (3.8)$$

and in this simulation, we employ Barrenbaltt kinetic Weighting function  $\chi(\omega) = \frac{1}{\pi g} \sqrt{(2g - \omega^2)_+}$ .

### 3.2 Progress

For the St. Venant iteration, I've completed and debugged the coding part (around 210 lines of Python

work). However, while running the simulation, the CFL time steps I introduced are at levels of  $10^{-5}$  seconds and the integrals per step take about 30 seconds to 1 minute. Thus, in practice, it takes about 1 day for 0.1 seconds of iteration of the water flow simulation. And if speeding up the iteration (like ten times faster for CFL time steps or 0.0001 seconds per step), the numerical iterations causes the  $F$  function faster than the recharge rate, and thus leads to negative water height, which is against physical laws. So, it's difficult to examine the water flow.

## References

- [1] Holley, E. R. *et al.* Climate change and market-based insurance feedbacks. *SOA* (2020).
- [2] Hinkel, J. *et al.* Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences* **111**, 3292–3297 (2014).
- [3] Beguería, S. & Vicente-Serrano, S. M. Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *Journal of applied meteorology and climatology* **45**, 108–124 (2006).
- [4] David, D. E. *et al.* Residential flood risk in the united states: Quantifying flood losses, mortgage risk and sea level rise. *SOA* (2020).
- [5] Herrera, S. *et al.* Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over spain (spain02). *International Journal of Climatology* **32**, 74–85 (2012).
- [6] Herrera, S., Fernández, J. & Gutiérrez, J. M. Update of the spain02 gridded observational dataset for euro-cordex evaluation: assessing the effect of the interpolation methodology. *International Journal of Climatology* **36**, 900–908 (2016).
- [7] Hosking, J. R. & Wallis, J. R. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics* **29**, 339–349 (1987).
- [8] Landwehr, J. M., Matalas, N. & Wallis, J. Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles. *Water resources research* **15**, 1055–1064 (1979).
- [9] Ersoy, M., Lakkis, O. & Townsend, P. A saint-venant shallow water model for overland flows with precipitation and recharge. *arXiv preprint arXiv:1705.05470* (2017).
- [10] Perthame, B. & Simeoni, C. A kinetic scheme for the saint-venant system with a source term. *Calcolo* **38**, 201–231 (2001).