Design storm estimation in Tuscany (Italy) through regional frequency analysis and generalized additive modelling

AUTHORS:

Matteo Pampaloni[1], Virginia Vannacci[2], **Enrica Caporali**[1], Chiara Bocci[2], Valentina Chiarello[1], and Alessandra Petrucci[2]

[1]Department of Civil and Environmental Engineering, University of Florence, Italy (*matteo.pampaloni@unifi.it*)
[2]Department of Statistics, Computer Science, Application University of Florence, Italy (*chiara.bocci@unifi.it*)

Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Dataset description

The complete dataset includes annual maxima of daily rainfall recorded:

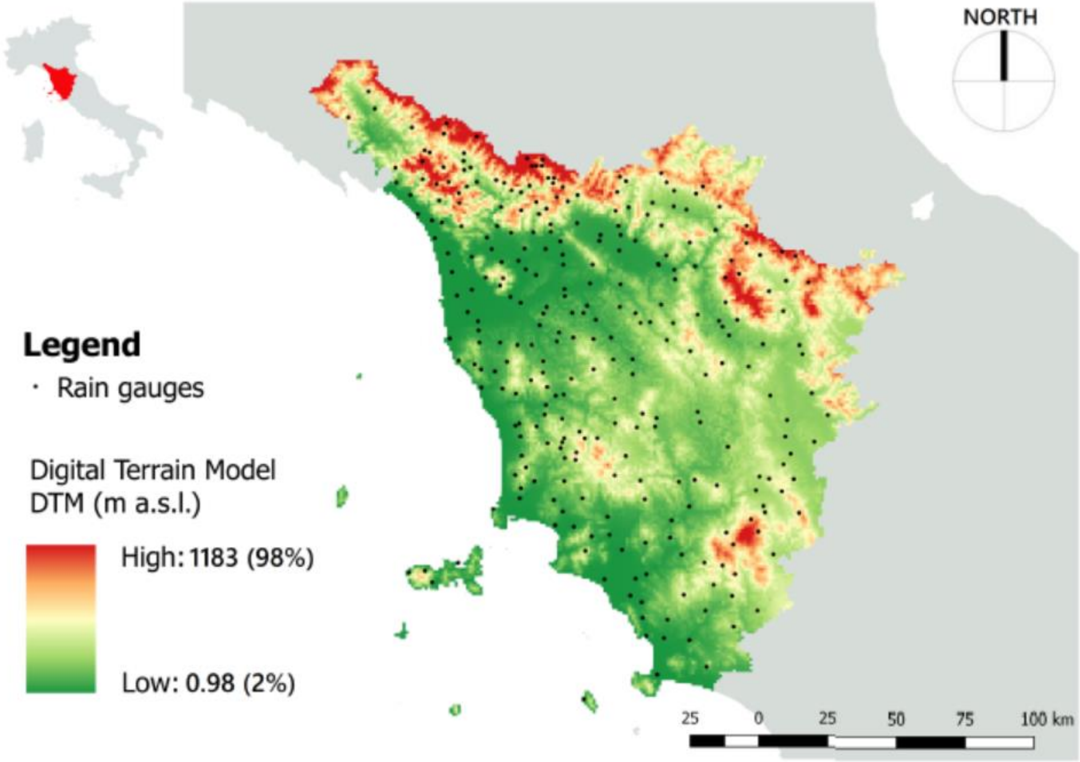
- in 876 rainfall gauges in Tuscany;
- over the period 1916-2017.

They have been considered:

- only stations with 30 or more years of data;
- only the post Second World War period.

The final dataset includes annual maxima of daily rainfall recorded:

- in 289 rainfall gauges in Tuscany;
- over the period 1951-2017;
- 14455 observations.



Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Regional frequency analysis

METHODOLOGY

1. Regions Subdivision ⇒

Preliminary hypothesis of subdivision

Spatial distribution of sample coefficient of skewness G (Lsk) [I level] and of variation Cv (Lcv) [II level]

2. Growth Factor ⇒

TCEV probability distribution function of the standardized variable $X' = X/\mu$

$$F_{X'}(x') = exp\left[-\Lambda_1\, exp(-\eta x') - \Lambda^* \Lambda_1^{\frac{1}{\theta^*}} exp\left(-\frac{\eta x'}{\theta^*}\right)\right]$$

Three levels hierarchical approach of parameters estimation

3. Index Raifall ⇒

For every homogenous region and every rainfall duration multivariate model *(Caporali et al., Environmentrics 2008)*

$$\mu_d = a_0 + a_1 \cdot \ln(M.A.P.) + a_2 \cdot z + a_3 \cdot \left[\sin\left(\frac{Asp}{2} - \frac{\pi}{2}\right) + \pi\right] \cdot |Asp| + a_4 \cdot h_{m,d}$$

as function of climatic and geomorphological characteristics

Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Regional frequency analysis

The **index variable method** *(Dalrymple, Geol. Survey 1543-A, 1960)*

The studied territory is divided into **homogeneous regions** where the **pdf** of the random variable observed at the measurement sites **is the same for all the stations unless of** a characteristic **scale factor** of each site called **INDEX VALUE**

$$x_i(F) = \mu_i \cdot x'(F)$$

“local” Index Rainfall “regional” Growth Factor

The **TCEV - Two Component Extreme Value probabilistic model** *(Rossi et al., WRR, 1984)*, with its hierarchical regionalization procedure on three levels *(Fiorentino et al., Reidel Pub., 1987)*.

$$F_{X'}(x') = exp\left[-\Lambda_1\, exp(-\eta x') - \Lambda^* \Lambda_1^{\frac{1}{\theta^*}} ex\,p\left(-\frac{\eta x'}{\theta^*}\right)\right]$$

where L_1 and L_2, with $L_1 > L_2 \geq 0$, represent the **annual number of events**, on average, due to the basic and the outliers component, while the parameters q_1 and q_2, with $q_2 > q_1 > 0$, represent the **annual mean value of the events**

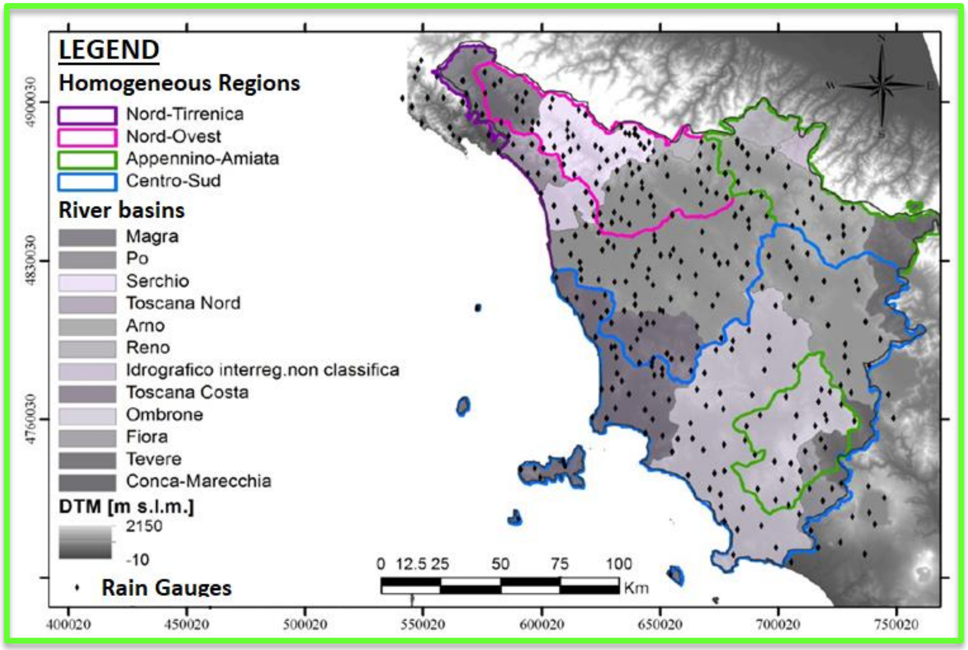
UNIVERSITÀ DEGLI STUDI FIRENZE

Design storm estimation in Tuscany (Italy) through regional frequency analysis and generalized additive modelling

Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Regional frequency analysis: identification of Homogeneous Regions

Subdivision in homogeneous regions and subregions	
Hypothesis 1	UNIQUE homogeneous region and subregion
Hypothesis 2	UNIQUE homogeneous region and 3 subregion NORTH-APPENNINE, NORTH-TYRRHENIAN, CENTER-SOUTH
Hypothesis 3	3 regions NORTH-APPENNINE, NORTH-TYRRHENIAN, CENTER-SOUTH coincident with the 3 sub-regions
Hypothesis 4	4 regions NORTH-TYRRHENIAN, NORTH-WEST, APPENNINE-AMIATA, CENTER-SOUTH coincident with the 4 sub-regions



Comparison between the observed cdf of coefficients G (Lsk) e Cv (Lcv) and the theoretical ones obtained with a Monte Carlo techniques as:

- Difference between ***m(x)obs – m(x)th*** *and* ***σ(x)obs - σ(x)th.***
- ***t Student*** test for the mean, ***Wilcoxon*** test for the mean, the ***χ2*** test.
- ***Discordancy D*** and ***Heterogeneity H*** test (Hosking & Wallis, Cambridge Un. Press, 1997).
- ***Gumbel probability plot test*** of the observed and the theoretical of the TCEV model growth curves.

Matteo Pampaloni*, Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Generalized additive model for Location, Scale and Shape (GAMLSS)

(Rigby and Stasinopoulos, Journal of the Royal Statistical Society, 2005)

- a model for each GEV parameter *(Chavez Demoulin and Davison, Journal of the Royal Statistical Society, 2005)*
  - a univariate smoothing component on the elevation x
  - a bivariate smoothing component on the geographical coordinates s

$$\begin{cases} y_i \sim \mathrm{GEV}\left(\mu(x_i, \mathbf{s}_i), \sigma(x_i, \mathbf{s}_i), \xi(x_i, \mathbf{s}_i)\right) \\ \mu(x_i, \mathbf{s}_i) = g_1(x_i) + f_1(\mathbf{s}_i) \\ \log(\sigma(x_i, \mathbf{s}_i)) = g_2(x_i) + f_2(\mathbf{s}_i) \\ \xi(x_i, \mathbf{s}_i) = g_3(x_i) + f_3(\mathbf{s}_i) \end{cases}$$

- Estimation method: Penalized ML via P-IRLS.
- **Number of knots**: We estimate various models increasing progressively the number K and H of knots until the degrees of freedom of each smoothing component are stable.
- **Type of basis**: Penalized thin plate, with smoothing parameter selected via REML estimation.

Final model ⇨

$$\begin{cases} y_i \sim \mathrm{GEV}\left(\mu(x_i, \mathbf{s}_i), \sigma(x_i, \mathbf{s}_i), \xi(x_i, \mathbf{s}_i)\right), \\ \mu(x_i, \mathbf{s}_i) = g(x_i, K = 20) + f(\mathbf{s}_i, H = 60) \\ \log(\sigma(x_i, \mathbf{s}_i)) = g(x_i, K = 20) + f(\mathbf{s}_i, H = 40) \\ \xi(x_i, \mathbf{s}_i) = g(x_i, K = 20) + f(\mathbf{s}_i, H = 30) \end{cases}$$

Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

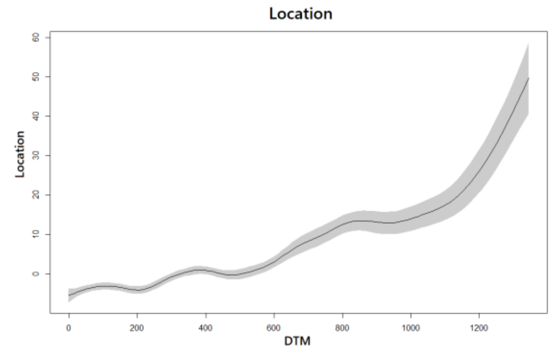
Generalized additive model for Location, Scale and Shape (GAMLSS)

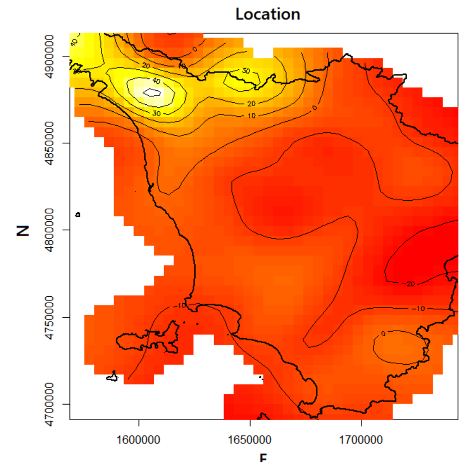
Final model coefficients and terms

Parametric coefficients			
Parameter	Estimate	Standard error	P-value
intercept μ	60.22	0.19	$< 2e-16$ ***
intercept σ	18.17	0.01	$< 2e-16$ ***
intercept ξ	0.13	0.01	$< 2e-16$ ***

Smooth terms		
Term	Degrees of freedom	P-value
Location: DTM	11.28	$< 2e-16$ ***
Location: Coordinates	56.9	$< 2e-16$ ***
Scale: DTM	5.72	2.94e-16 ***
Scale: Coordinates	31.39	$< 2e-16$ ***
Shape: DTM	2.42	0.023 *
Shape: Coordinates	15.27	8.19e-08 ***

Smoothing components in the model for the location parameter

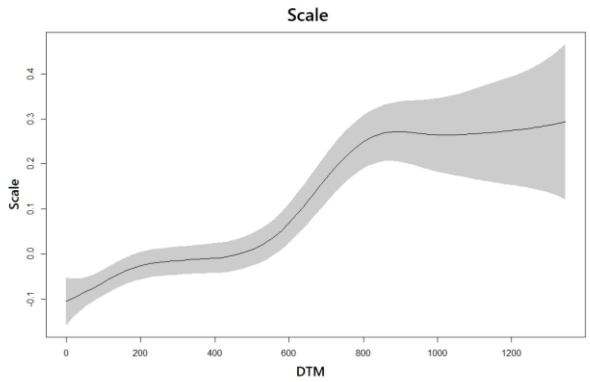


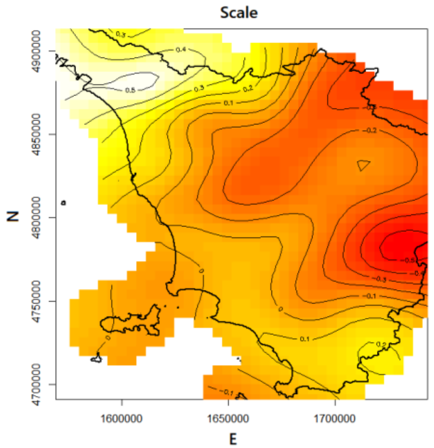


Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

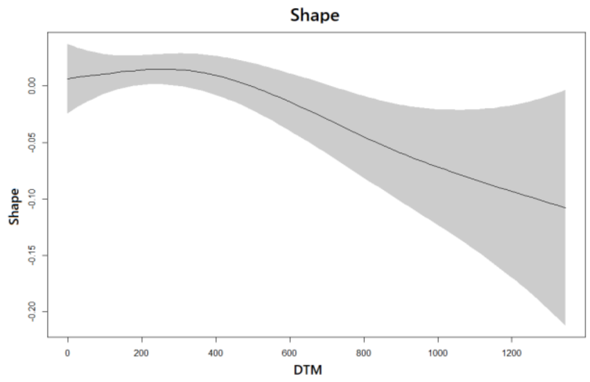
Generalized additive model for Location, Scale and Shape (GAMLSS)

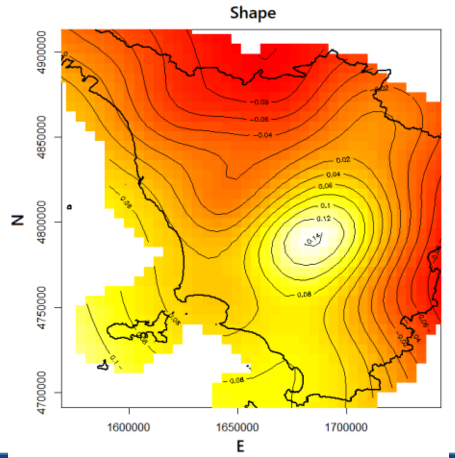
Smoothing components in the model for the scale parameter





Smoothing components in the model for the shape parameter

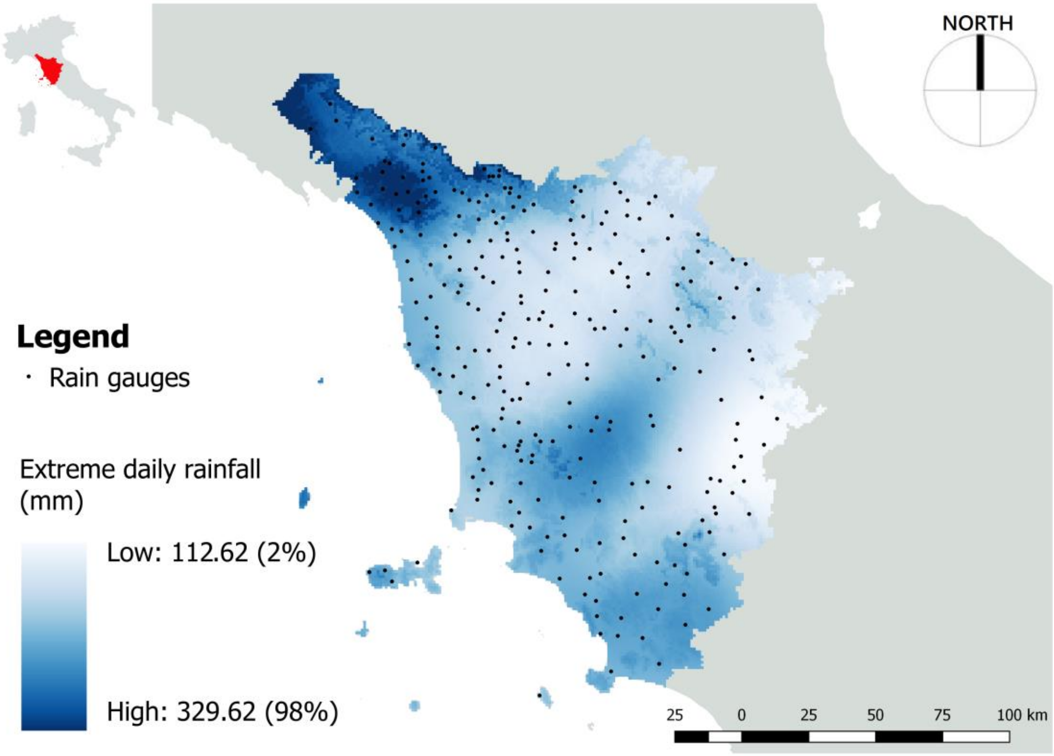




Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Generalized additive model for Location, Scale and Shape (GAMLSS)

Design rainfall with daily storm duration and 200 year return period



Min	100.6
10%	143.59
1st Quart.	160.90
Median	194.30
Mean	199.60
3rd Quart.	228.70
90%	265.45
Max	444.00

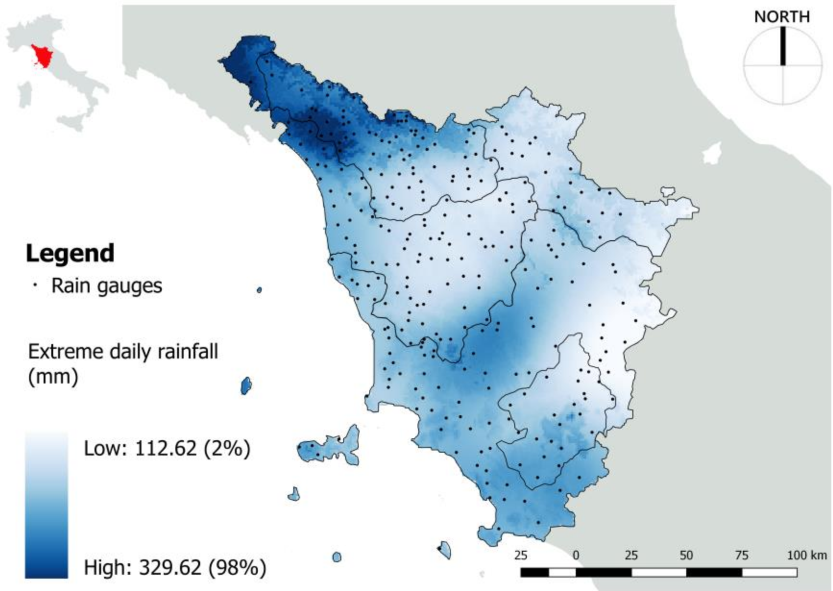
In addition, the magnitudes of design extreme rainfall that are expected to be exceeded in Tuscany once every 2, 20, 50, 100 years, has been predicted

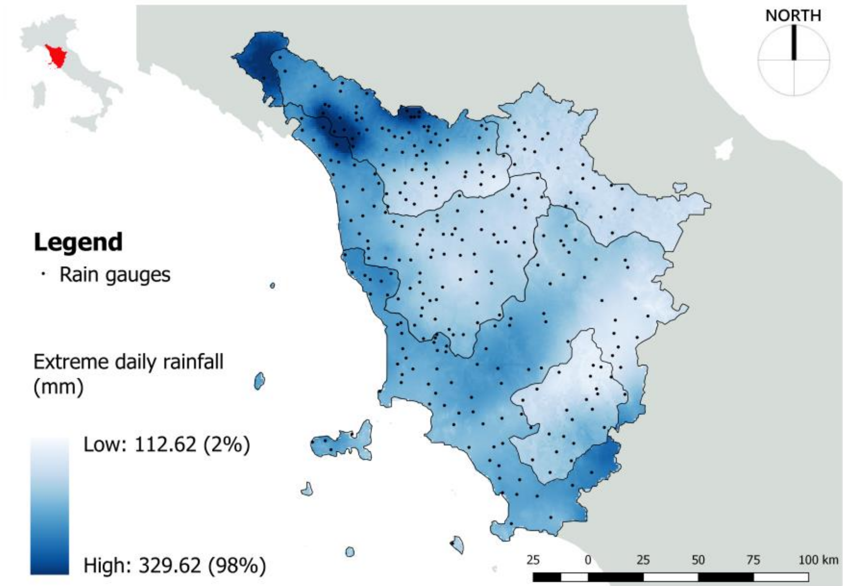
UNIVERSITÀ DEGLI STUDI FIRENZE

Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Generalized additive model compared to Regional Frequency Analysis

Design rainfall with daily storm duration and 200 year return period





GAMLSS

return period	correlation
2	0.88
20	0.91
50	0.89
100	0.88
200	0.87

RFA, using the index method

Matteo Pampaloni, *Virginia Vannacci, Enrica Caporali, Chiara Bocci, Valentina Chiarello and Alessandra Petrucci*

Conclusions

- This analysis referred to the annual maxima of daily rainfall recorded in Tuscany, over the period 1951-2017.
  Aim of the study: to model the distribution of rainfall maxima to predict
  - extreme rainfalls
  - how often they are expected to occur in each location of the region
- A Generalized Additive Model for spatial extreme values has been used
  - Considering the class of Generalized Additive Models for Location, Scale and Shape (GAMLSS).
- **Objective of main interest in the field of hydrology**: prediction of extreme rainfalls associated 2-year, 20-year, 50-year, 100-year, 200-year return periods.
- **Highest values**: Appennines, Alpi Apuane in the North; Colline Metallifere and Maremma in the South
- **Lowest values**: central flat zone including Empoli, Pistoia, Prato and Firenze and Val di Chiana in the East
- **Comparison with the Index method**: the correlation coefficient is nearly 0.9 for all the return periods, detecting a consistence between the two methods.
- GAMLSS method allows for homogeneous subregions to be identified a posteriori, subsequently to the estimation process, which is free from any a priori assumption on the phenomenon spatial behaviour.
- **A future development**: given the additive formulation of our regression models, we could easily consider to model extreme value series with spatio-temporal non-stationarity.