# Geoadditive modeling for extreme rainfall data

Chiara Bocci, Alessandra Petrucci
Department of Statistics "Giuseppe Parenti", University of Firenze,
bocci@ds.unifi.it, alessandra.petrucci@unifi.it

Enrica Caporali
Department of Civil and Environmental Engineering, University of Firenze,
enrica.caporali@unifi.it

**Abstract:** Extreme value models and techniques are widely applied in environmental studies to define protection systems against the effects of extreme levels of environmental processes. Regarding the matter related to the climate change science, a certain importance is cover by the implication of changes in the hydrological cycle. Among all hydrologic processes, rainfall is a very important variable as it is a fundamental component of flood risk mitigation and drought assessment, as well as water resources availability and management. We implement a geoadditive mixed model for extremes with a temporal random effect assuming that the observations follow generalized extreme value distribution with spatially dependent location. The analyzed territory is the catchment area of Arno River in Tuscany in Central Italy.

**Keywords:** GEV distribution, geoadditive mixed model, hydrologic processes

## 1 Introduction

Environmental extreme events such as floods, earthquakes, hurricanes, may have a massive impact on everyday life for the consequences and damage that they cause. For this reason there is considerable attention in studying, understanding and predicting the nature of such phenomena and the problems caused by them, not least because of the possible link between extreme climate events and climate change. A number of theoretical modeling and empirical analyses have also suggested that notable changes in the frequency and intensity of extreme events, including intense rainfall and floods, may occur even when there are only small changes in climate (Katz and Brown, 1992).

In this framework, in the past two decades there has been an increasing interest for statistical methods that model rare events (Coles, 2001). The Generalized Extreme Value distribution (GEV) is widely adopted model for extreme events in the univariate context. For modeling extremes of non-stationary sequences it is commonplace to use the GEV as a basic model, and to handle the issue of non-stationarity by regression modeling of the GEV parameters.

Here we implement a geoadditive mixed model for extremes with a temporal random effect. We assume that the observations follow a generalized extreme value distribution whose locations are spatially dependent where the dependence is captured using the geoadditive model. The analyzed territory is the catchment area of Arno River in Tuscany in Central Italy.

# 2 Materials and Methods

The investigation is developed on the catchment area of Arno River almost entirely situated within Tuscany, Central Italy. The time series of annual maxima of daily rainfall recorded in 415 rain gauges are analyzed. In order to have enough rain gauges observations to estimate both the spatial component and the year specific effect, we reduce the time series length to the post Second World War period and we consider only stations with at least 30 hydrologic years of data, even not consecutive. The final dataset is composed by the data recorded from 1951 to 2000 at 118 rain gauges for a total of 4903 observations.

Recently to handle the issue of non-stationarity of the GEV parameters, Padoan and Wand (2008) discuss how generalized additive models (GAM) with penalized splines can be carried out in a mixed model framework for the GEV family.

Geoadditive models, introduced by Kammand and Wand (2003), are a particular specification of GAM that models the spatial distribution of $y$ with a bivariate penalized spline on the spatial coordinates. Suppose to observe $n$ sample maxima $y_{ij}$ at spatial location $\mathbf{s}_{ij}$, $\mathbf{s} \in \mathbb{R}^2$, $j = 1, \ldots, p$ and at time $i = 1, \ldots, t$. In order to model both the spatial and the temporal influence on the annual rainfall maxima, we consider a geoadditive mixed model for extremes with a temporal random effect:

$$\begin{cases} y_{ij} | s_{ij} \sim \text{GEV}(\mu\left(s_{ij}\right), \psi, \xi) \\ \mu\left(\mathbf{s}_{ij}\right) = \beta_0 + \mathbf{s}_{ij}^T \boldsymbol{\beta}_s + \sum_{k=1}^{K} u_k b_{tps}(\mathbf{s}_{ij}, \boldsymbol{\kappa}_k) + \gamma_i, \end{cases} \tag{1}$$

where $\mu$, $\psi$ and $\xi$ are respectively location, scale and shape parameters of the GEV distribution, $b_{tps}$ are the low-rank thin plate spline basis functions with $K$ knots and $\gamma_i$ is the time specific random effect. The model (1) can be written as a mixed model

$$\mathbf{y} | \left(\mathbf{u}, \boldsymbol{\gamma}\right) \sim \text{GEV}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{D}\boldsymbol{\gamma}, \psi, \xi). \tag{2}$$

with

$$\text{E}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \text{Cov}\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \sigma_u^2 \mathbf{I}_K & 0 \\ 0 & \sigma_\gamma^2 \mathbf{I}_t \end{bmatrix}.$$

where

$$\boldsymbol{\beta} = \left[\beta_0, \boldsymbol{\beta}_s^T\right] \qquad \mathbf{u} = [u_1, ..., u_K] \qquad \boldsymbol{\gamma} = [\gamma_1, ..., \gamma_t]$$
$$\mathbf{X} = \left[1, \mathbf{s}_{ij}^T\right]_{1 \le ij \le n} \qquad \mathbf{D} = \left[d_{ij}\right]_{1 \le ij \le n}$$

with $d_{ij}$ an indicator taking value 1 if we observe a rainfall maxima at rain gauge $j$ in year $i$ and 0 otherwise, and $\mathbf{Z}$ the matrix containing the spline basis functions, that is

$$\mathbf{Z} = [b_{tps}(\mathbf{s}_{ij}, \boldsymbol{\kappa}_k)]_{1\leq ij\leq n, 1\leq k\leq K} = [C(\mathbf{s}_{ij} - \boldsymbol{\kappa}_k)]_{1\leq ij\leq n, 1\leq k\leq K} \cdot [C(\boldsymbol{\kappa}_h - \boldsymbol{\kappa}_k)]^{-1/2}_{1\leq h,k\leq K},$$

where $C(\mathbf{v}) = \|\mathbf{v}\|^2 \log \|\mathbf{v}\|$ and $\boldsymbol{\kappa}_1, ..., \boldsymbol{\kappa}_K$ are the spline knots locations.

# 3 Results

The geoadditive mixed model for extremes (2) can be naturally formulated as a hierarchical Bayesian model and estimated under the Bayesian paradigm. Following the specifications of Padoan (2008), our complete hierarchical Bayesian formulation is

$$\begin{array}{ll} \text{1st level} & y_i|\,(\mathbf{u},\boldsymbol{\gamma}) \overset{\text{ind}}{\sim} \text{GEV}(\,[\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{D}\boldsymbol{\gamma}]_i\,, \psi, \xi) \\ \text{2st level} & \mathbf{u}|\sigma_u^2 \sim N(0, \sigma_\gamma^2\mathbf{I}_K) \qquad \boldsymbol{\gamma}|\sigma_\gamma^2 \sim N(0, \sigma_\gamma^2\mathbf{I}_t) \qquad \boldsymbol{\beta} \sim N(0, 10^4\mathbf{I}) \\ & \xi \sim \text{Unif}(-5, 5) \qquad \psi \sim \text{InvGamma}(10^{-4}, 10^{-4}) \\ \text{3st level} & \sigma_u^2 \sim \text{InvGamma}(10^{-4}, 10^{-4}) \qquad \sigma_\gamma^2 \sim \text{InvGamma}(10^{-4}, 10^{-4}). \end{array}$$

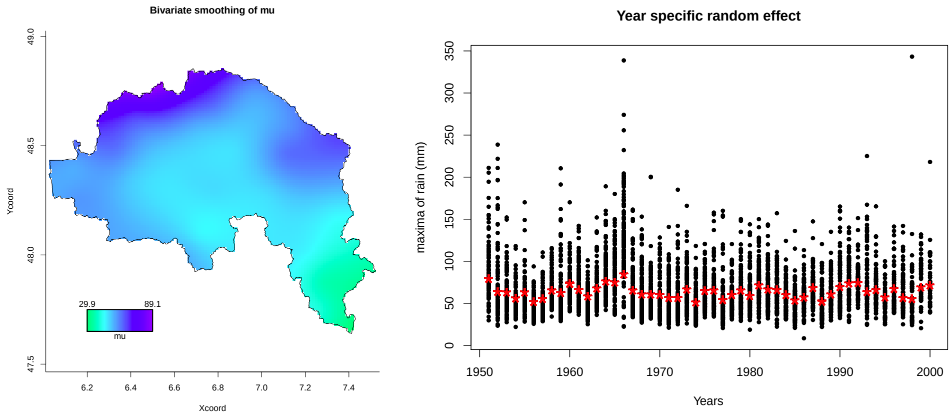
where the parameters setting of the priors distributions for $\xi$, $\psi$, $\boldsymbol{\beta}$, $\sigma_u^2$, $\sigma_\gamma^2$, corresponds to non-informative priors.

Given the complexity of the proposed hierarchical models, we employ `OpenBUGS` Bayesian MCMC inference package to do the model fitting. We access `OpenBUGS` using the package `BRugs` in the `R` computing environment. We implement the MCMC analysis with a burn-in period of 40000 iterations and then we retain 10000 iterations, that are thinned by a factor of 5, resulting in a sample of size 2000 collected for inference. Finally, the last setting concern the thin plate spline knots that are selected setting $K = 30$ and using the *clara* space filling algorithm of Kaufman and Rousseeuw (1990), available in the `R` package `cluster`.

The resulting spatial smoothing component and time specific component of $\mu(s_{ij})$ are presented in Figures 1(a) and 1(b). Observing the map, it is evident the presence of a spatial trend in the rainfall extreme dynamic, even after controlling for the year effect. The spline seems to capture well the spatial dependence as it produce the same same patter of the Average Total Annual Precipitation. The time influence is pointed out by the estimated year specific random effects, that present a strong variability through years.

# 4 Conclusions

We have implemented a geoadditive modeling approach for explaining a collection of spatially referenced time series of extreme values. We assume that the obser-




(a) (b)

Figure 1: Estimated spatial component (a) and year specific random effects (b) of $\mu(s_{ij})$. Black dots indicate the observed values.

vations follow generalized extreme value distributions whose locations are spatially dependent.

The results show that this model allows us to capture both the spatial and the temporal dynamics of the rainfall extreme dynamic.

Under this approach we expect to reach a better understand of the occurrence of extreme events which are of practical interest in climate change studies particularly when related to intense rainfalls and floods, and hydraulic risk management.

# References


Coles, S.G. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer, London.

Kammann, E.E., Wand, M.P. (2003) Geoadditive models, *Applied Statistics*, 52, 1-18.

Katz R.W., Brown B.G. (1992) Extreme events in a changing climate: variability is more important than averages, *Climate Change*, 21, 289-302.

Kaufman, L., Rousseeuw, P.J. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.

Padoan, S.A. (2008) *Computational methods for complex problems in extreme value theory*, Ph.D. thesis, Ph.D. in Statistical Science, Department of Statistical Science, University of Padova.

Padoan, S.A., Wand, M.P. (2008) Mixed model-based additive models for sample extremes, *Statistics and Probability Letters*, 78, 2850-2858.