



Article

Developing a Hierarchical Model for the Spatial Analysis of PM₁₀ Pollution Extremes in the Mexico City Metropolitan Area

Alejandro Ivan Aguirre-Salado ^{1,*}, Humberto Vaquera-Huerta ², Carlos Arturo Aguirre-Salado ³, Silvia Reyes-Mora ¹, Ana Delia Olvera-Cervantes ¹, Guillermo Arturo Lancho-Romero ¹ and Carlos Soubervielle-Montalvo ³

¹ Department of Physics and Mathematics, Universidad Tecnológica de la Mixteca, 69000 Huajuapán de León, Oax., Mexico; sreyes@mixteco.utm.mx (S.R.-M.); ana.olvera@mixteco.utm.mx (A.D.O.-C.); lancho@mixteco.utm.mx (G.A.L.-R.)

² Department of Statistics, Colegio de Postgraduados, Campus Montecillo, Texcoco, 56230 Montecillo, Mex., Mexico; hvaquera@colpos.mx

³ Faculty of Engineering, Universidad Autónoma de San Luis Potosí, 78280 San Luis Potosí, S.L.P., Mexico; carlos.aguirre@uaslp.mx (C.A.A.-S.); carlos.soubervielle@uaslp.mx (C.S.-M.)

* Correspondence: aleaguirre@mixteco.utm.mx; Tel.: +55-953-532-0214

Academic Editor: Kim Natasha Dirks

Received: 22 May 2017; Accepted: 3 July 2017; Published: 6 July 2017

Abstract: We implemented a spatial model for analysing PM₁₀ maxima across the Mexico City metropolitan area during the period 1995–2016. We assumed that these maxima follow a non-identical generalized extreme value (GEV) distribution and modeled the trend by introducing multivariate smoothing spline functions into the probability GEV distribution. A flexible, three-stage hierarchical Bayesian approach was developed to analyse the distribution of the PM₁₀ maxima in space and time. We evaluated the statistical model's performance by using a simulation study. The results showed strong evidence of a positive correlation between the PM₁₀ maxima and the longitude and latitude. The relationship between time and the PM₁₀ maxima was negative, indicating a decreasing trend over time. Finally, a high risk of PM₁₀ maxima presenting levels above 1000 µg/m³ (return period: 25 yr) was observed in the northwestern region of the study area.

Keywords: air pollution; particulate matter; extreme value theory; Markov Chain Monte Carlo (MCMC); nonstationary

1. Introduction

Air pollution in urban areas has become a major problem [1]. Increases in industry and urban traffic due to economic and population growth have led to an increase in gas and particulate emission that contribute to air pollution [2]. Among air pollutants, fine inhalable particles, known as particulate matter (PM) are associated with respiratory illnesses such as bronchitis, emphysema, asthma and other chronic obstructive pulmonary diseases [3]. PM can be classified by size; particles of 10 µm or less in diameter are called PM₁₀. They can consist of diverse solid and liquid particles, which may contain chemical constituents such as nitrates, sulfates and organic carbon [4]. PM originates from factories, internal combustion engines, heating systems, volcanoes, and deserts, and can include dust particles formed through the mechanical breakdown of larger particles during agricultural and mining processes [5]. In the Mexico City metropolitan area, a study conducted by the U.S. Department of Energy (DOE) and Mexico's Petróleos Mexicanos (PEMEX) showed that approximately 50% of the PM₁₀ was composed of dust from roadways, construction, and bare land [6].

Previous studies on the damage caused to human health by breathable particulate matter have revealed an association between high concentrations of PM₁₀ and mortality due to respiratory diseases [1,7]. To reduce exposure and minimize the adverse effects of these particulates on public health, several studies have been conducted with a focus on understanding the causes and factors related to the origin and flow of these particles [4,8]. Most of these studies have relied on continuous measurements of PM₁₀ to predict future concentrations based on various models, such as multiple regression models, neural networks [9] and generalized linear models [5]. However, these short-term forecasting methodologies were developed for use in locations with limited infrastructure, and where obtaining continuous measurements of PM₁₀ concentrations is difficult. In other, more densely populated regions, such as the Mexico City metropolitan area, there are systems that measure these concentrations and send information in real time to a central location for the immediate activation or deactivation of alerts or emergency procedures. In these cases, it is more convenient to study air pollutants through the use of robust methodologies such as the theory of extreme values.

Extreme value theory is used in many fields, such as environmental studies, engineering and finance, to monitor the quantitative risk of future extreme events [10,11]. In the environmental sciences, it is used to model extremes of temperature, rainfall, wind and pollution, among other phenomena. The asymptotic distribution of maxima is known as the generalized extreme value distribution, which is described by three parameters corresponding to location, scale and shape [12]. These parameters are estimated using the maximum likelihood method [13]. However, this method is not robust with a small sample size, so many other estimation methods have also been proposed, such as the method of moments, the use of L-moments and the use of weighted probability moments [14–16].

Recently, new methodologies have been proposed for the study of extreme values, mainly for application to hydroclimatological and environmental data, all of them based on the generalized extreme value distribution. For example: Gaetan and Grigoletto [17] proposed the use of Markov random fields approximated based on smoothing kernel parameters for modeling the parameters of the GEV distribution. Reich et al. [18] studied heat waves using a Bayesian hierarchical model with the generalized Pareto distribution (GPD). Cooley and Sain [19] studied maximum rainfall events by assigning a normal prior to the parameters of the GPD. Sang and Gelfand [11] studied the extreme values of spatial stochastic processes and modeled the observed trend as a function of covariates.

In a real scenario, it is common that conditions change and that the assumptions of stationarity that are required in a traditional analysis of extreme values not met; this is because there are often trends of extreme values [20–22]. Recent studies have introduced covariate functions for describing extreme value distributions to model either the location parameter, the scale parameter or both. Regarding the location parameter, Weissman [23] used a sine function, Tawn [24] and Scarf [25] proposed a linear function, Rosen and Cohen [26] and Pauli and Coles [27] used splines, and Bocci et al. [28] used a geoadditive model. In the case of the scale parameter, because this parameter is assumed to be positive, it is more common to model its logarithm; therefore, Yee and Stephenson [29] used additive models, El Adlouni et al. [10] and Rodriguez et al. [30] used linear functions, and Cannon [31] proposed the use of neural networks.

In this article, we present a spatio-temporal analysis of extreme PM₁₀ values in the Mexico City metropolitan area. The objective of this study was to analyse PM₁₀ data collected over time and in different spatial locations to gain insight into the distribution of PM₁₀ maxima and to quantify the risk of future extreme PM₁₀ pollution events, even in non-monitored regions.

2. Materials and Methods

2.1. Study Area

The Mexico City metropolitan area is one of the world's largest urban agglomerations, with approximately 25.4 million inhabitants spread over 7866 km² at an average elevation of 2240 m above mean sea level. It is surrounded by mountains to the east and west, creating a basin with

low points to the the north which leads to air pollution problems because of limited ventilation [6]. Two synoptic wind regimes prevail throughout the year: an anticyclonic westerly wind from November to April and a moist trade wind associated with the rainy season from May to October. Despite weak prevailing synoptic winds, the thermally induced local wind converges toward the city, which tends to restrict the ventilation of polluted air [32]. The study area and the locations of the primary sampling sites are shown in Figure 1.

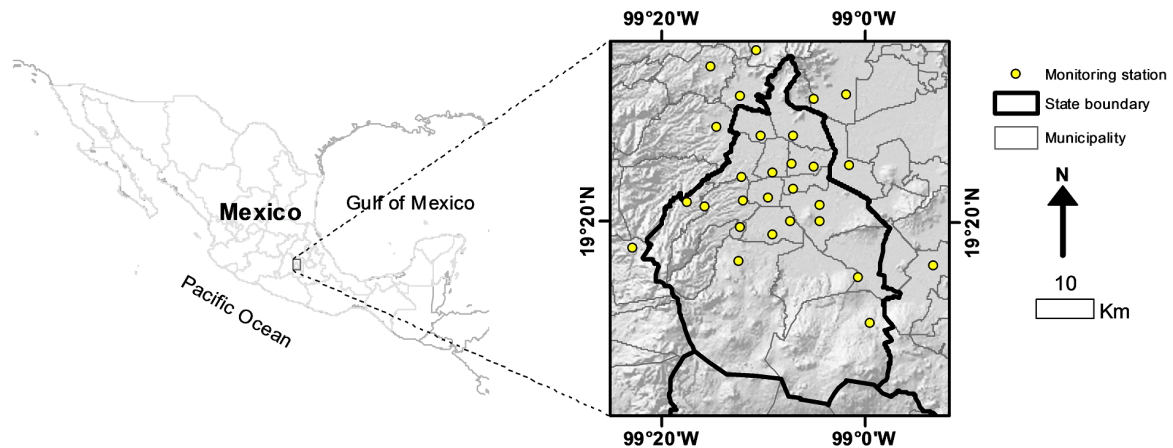


Figure 1. Study area.

2.2. Methodology

2.2.1. A Nonstationary GEV Model

Let Y_1, \dots, Y_n be an independent and identically distributed set of random variables with distribution function $F_X(x)$ and let $M_n = \max(Y_1, \dots, Y_n)$. Let F_{M_n} be the distribution function of M_n , because $F_{M_n} = F_X(x)^n$ we have that M_n is a degenerate distribution when $n \rightarrow \infty$. The extreme value theory considers that the only nondegenerate limiting distribution $G_n = (M_n - a_n)/b_n$ (if such a sequences of constants $\{b_n > 0\}$ and $\{a_n\}$ exist) as $n \rightarrow \infty$ is the generalized extreme value (GEV) distribution [12]:

$$F_{GEV}(y) = \begin{cases} \exp \left\{ - \left(1 + \kappa \frac{(y-\mu)}{\sigma} \right)^{-\frac{1}{\kappa}} \right\}, & \kappa \neq 0, \\ \exp \left\{ - \exp \left(-\frac{(y-\mu)}{\sigma} \right) \right\}, & \kappa = 0 \end{cases}$$

for $y : 1 + \kappa \frac{(y-\mu)}{\sigma} > 0$ when $\kappa \neq 0$, where $-\infty \leq y \leq \mu + \sigma/\kappa$ when $\kappa < 0$ (Weibull), $-\infty \leq y \leq +\infty$ when $\kappa = 0$ (Gumbel) and $\mu + \sigma/\kappa \leq y \leq +\infty$ when $\kappa > 0$ (Fréchet). Here, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\kappa \in \mathbb{R}$ are the location, scale and shape parameters, respectively; see [33].

In the nonstationary case, the parameters are expressed as a function of a vector of covariates x_t : $GEV(\mu(x_t), \sigma(x_t), \kappa(x_t))$ [34]. To ensure a positive value for the scale parameter, $\log \sigma(x_t)$ is used instead of $\sigma(x_t)$ in the estimation process.

2.2.2. Proposed Approach

For the non-stationary case, consider assigning a linear predictor to the location and scale parameters. The κ parameter is usually assumed to be constant [29]. Therefore, we propose using the following conditions for the above parameters:

$$\begin{aligned} \mu_t &= \sum_{i=1}^{P_1} X_{ti}\beta_{1i} + \sum_{i=1}^{P_2} Z_{ti}u_{1i}, \\ \log \sigma_t &= \sum_{i=1}^{P_1} X_{ti}\beta_{2i} + \sum_{i=1}^{P_2} Z_{ti}u_{2i}, \\ \kappa_t &= \kappa \end{aligned} \tag{1}$$

where X is a scaled and centered $n \times p_1$ matrix of covariates that includes the intercept; $\beta_i, i = 1, 2$, is a vector of length p_1 ; $u_i, i = 1, 2$, is a vector of length p_2 ; Z is an $n \times p_2$ matrix such that $\{Z\}_{ij} = \exp \left[- \left(\left\| \underline{x}_i - \underline{k}_j \right\| \right)^2 \right], i = 1, \dots, n, j = 1, \dots, p_2$; \underline{x}_i is the vector of covariates for the i -th observation, scaled and centered; and \underline{k}_j the j -th centroid obtained using the method of average linkage hierarchical clustering [28,35].

2.2.3. Maximum Likelihood Estimation

For a sample of n observations: $\underline{y} = (y_1, \dots, y_n)$, the maximum likelihood estimator for the non-stationary GEV can be determined by maximizing the likelihood function

$$L(\mu_t, \sigma_t, \kappa_t | \underline{y}) = \prod_{t=1}^n \frac{1}{\sigma_t} \exp \left\{ - \left[1 + \kappa_t \left(\frac{y_t - \mu_t}{\sigma_t} \right) \right]^{-\frac{1}{\kappa_t}} \right\} \times \left[1 + \kappa_t \left(\frac{y_t - \mu_t}{\sigma_t} \right) \right]^{-\left(1 + \frac{1}{\kappa_t}\right)}$$

where n is the number of observations. The function of κ_t is usually assumed constant [29], so the log-likelihood is:

$$\ell(\mu_t, \sigma_t, \kappa | \underline{y}) = -n \log \sigma_t - \sum_{t=1}^n \left[1 + \kappa \left(\frac{y_t - \mu_t}{\sigma_t} \right) \right]^{-\frac{1}{\kappa}} - \sum_{t=1}^n \left(1 + \frac{1}{\kappa} \right) \log \left[1 + \kappa \left(\frac{y_t - \mu_t}{\sigma_t} \right) \right]$$

Let $C = [X \ Z]$ and $\underline{b}'_{(i)} = \left[\underline{\beta}'_{(i)} \ \underline{u}'_{(i)} \right]$, where C is a $n \times p$ matrix, with $p = p_1 + p_2$; $\underline{b}'_{(i)}$ is a vector of $p \times 1$ parameters, the linear predictors can be written as:

$$\mu_t = Cb_1; \log \sigma_t = Cb_2; \kappa_t = \kappa$$

For this study, we assigned a penalization (P) to the vector of parameters, such that:

$$P = \begin{bmatrix} \frac{1}{\sigma_\beta^2} \otimes I_{p_1} & 0 \\ 0 & \frac{1}{\sigma_u^2} \otimes I_{p_2} \end{bmatrix}$$

where I_{p_1} and I_{p_2} are identity matrices of order p_1 and p_1 respectively, σ_β^2 and σ_u^2 are values that control the degree of regularization of the model.

Therefore the penalized log-likelihood of the model is:

$$\ell_n^p(\mu_t, \sigma_t, \kappa | \underline{y}) = \sum_{t=1}^n \ell_t(\mu_t, \sigma_t, \kappa | \underline{y}) - \sum_{i=1}^2 \underline{b}'_{(i)} P \underline{b}_{(i)} - \frac{1}{\sigma_\kappa^2} \kappa^2$$

where:

$$\ell_t(\mu_t, \sigma_t, \kappa | \underline{y}) = -\log \sigma_t - \left[1 + \kappa \left(\frac{y_t - \mu_t}{\sigma_t} \right) \right]^{\frac{1}{\kappa}} \log \left[1 + \kappa \left(\frac{y_t - \mu_t}{\sigma_t} \right) \right]$$

Defining $\ell_n(\mu_t, \sigma_t, \kappa | \underline{y}) = \sum_{t=1}^n \ell_t(\mu_t, \sigma_t, \kappa | \underline{y})$ and $\ell_{PL}(b_{(1)}, b_{(2)}, \kappa) = \left[- \sum_{i=1}^2 \underline{b}'_{(i)} P \underline{b}_{(i)} - \frac{1}{\sigma_\kappa^2} \kappa \right]$ we can rewrite the log-likelihood as:

$$\ell_n^p(\mu_t, \sigma_t, \kappa | \underline{y}) = \ell_n(\mu_t, \sigma_t, \kappa | \underline{y}) + \ell_{PL}(b_{(1)}, b_{(2)}, \kappa) \tag{2}$$

In order to compare two models, let M_1 with θ_1 a parametric vector against another model M_0 with θ_0 a subset vector such as $M_1 \subset M_0$, a simple way to compare two models is to use the deviance statistic defined by [34]

$$D = 2 (l_n^* (M_1) - l_n^* (M_0))$$

where $l_n^* (M)$ is the maximized log likelihood function of model M . Values of D greater than χ_k^2 are considered significant, where k is the difference between the dimensions of M_1 and M_0 , thus model M_1 explains better data variation than model M_0 .

Penalized maximum likelihood estimators are used in this work for two reasons, the first is to use these estimators to perform a procedure of variables selection through the deviance, and second, we will use these estimators as initial values for the Bayesian hierarchical model to reduce the number of samples necessary to reach the stationary distribution of the MCMC algorithm.

2.2.4. Bayesian Implementation

Under the assumption that $\pi (y_t | \theta_t)$ is the GEV distribution with parameters of $\theta_t = (\mu_t, \sigma_t, \kappa)$, a Bayesian formulation for the model of extreme values is as follows:

$$\pi (\theta_t, \omega | y_t) \propto \pi (y_t | \theta_t) \pi (\theta_t | \omega) \pi (\omega) \tag{3}$$

where $\omega = (\beta_1, \beta_2, u_1, u_2)$ is such that the set of equations given in Equation (1) is satisfied. This hierarchical model consists of three levels: a data level, denoted by $\pi (y_t | \theta_t)$; a process level, denoted by $\pi (\theta_t | \omega)$; and a prior level, denoted by $\pi (\omega)$. Alternatively, the model given in Equation (3) with the conditions given in Equation (1) can be reformulated as a function of the parameter space of $\omega^* = (\beta_1, \beta_2, u_1, u_2, \kappa)$ and $\omega^{**} = \{\sigma_1, \sigma_2\}$; therefore, we can write the posterior distribution as follows:

$$\pi (\omega^*, \omega^{**} | y_t) \propto \pi (y_t | \omega^*) \pi (\omega^* | \omega^{**}) \pi (\omega^{**}) \tag{4}$$

where $\pi (y_t | \omega^*)$ is the GEV density under the conditions on the parameters given in Equation (1). The prior distribution for ω^* is such that

$$\begin{aligned} \beta_1 &\sim N (0, 10^4 I) \\ \beta_2 &\sim N (0, 10^4 I) \\ u_1 | \sigma_1 &\sim N (0, \sigma_1^2 I_{K_x}) \\ u_2 | \sigma_2 &\sim N (0, \sigma_2^2 I_{K_x}) \\ \kappa &\sim Uniform(-5, 5) \end{aligned}$$

The prior distribution for the hyperparameters ω^{**} is given by

$$\begin{aligned} \sigma_1^2 &\sim Half - Cauchy(25) \\ \sigma_2^2 &\sim Half - Cauchy(25) \end{aligned}$$

To sample the a posteriori distribution, we use a MCMC method with an acceptance probability given by

$$\alpha (\theta^* | \theta) = \min \left(1, \frac{\pi (x | \theta^*) \pi (\theta^*) Q (\theta^*, \theta)}{\pi (x | \theta) \pi (\theta) Q (\theta, \theta^*)} \right)$$

where $\pi (\theta)$ is the prior distribution for the parameters, $\pi (x | \theta)$ is the likelihood, and Q is the proposal function.

2.2.5. Simulation Study

In this section, we examine the performance of the hierarchical GEV model defined in Equation (4) using simulated data. We simulated 500 extremes from a GEV model with two covariates, x_1 and x_2 ,

corresponding to the longitude and latitude, respectively. The x_1 values were generated with equally spaced data in the range of 0 to 4π , and the values of the covariate x_2 were randomly selected from the interval $[0, 4]$, with

$$\mu_t = \sin \left(\frac{\left(\sqrt{(x_1 - 2\pi)^2 + (x_2 - 2\pi)^2} \right)^{\frac{3}{2}}}{5} \right)$$

$$\sigma_t = \sigma = 0.3$$

$$\kappa_t = \kappa = -0.1$$
(5)

For our model, we ran 70,000 iterations to obtain samples of the *a posteriori* density, after a burn-in period of 60,000 iterations, and applied thinning in every fifth iteration. The hierarchical model given in Equation (4) was fitted by setting $p_2 = 10$ in equation set Equation (1). The estimate of the shape parameter was -0.14 . The true functions of μ and σ as functions of the covariates x_1 and x_2 are shown in Figure 1a,c, respectively. The function corresponding to μ that was chosen for the simulation is a function that cannot be separated based only on the main effects of the covariates, so it cannot be adjusted as in most traditional models for extreme values. The function for sigma is a flat function in the space covariate.

The true surface and local spatial patterns of the location parameter (Figure 2a) that were used to simulate the extreme values were recovered reasonably well by the smoothing function (Figure 2b). Similarly, a comparison of Figure 2c,d reveals that the smoothing function for the scale parameter of the extreme values based on the model given in Equation (1) recovered the true flat function for σ given in Equation (5).

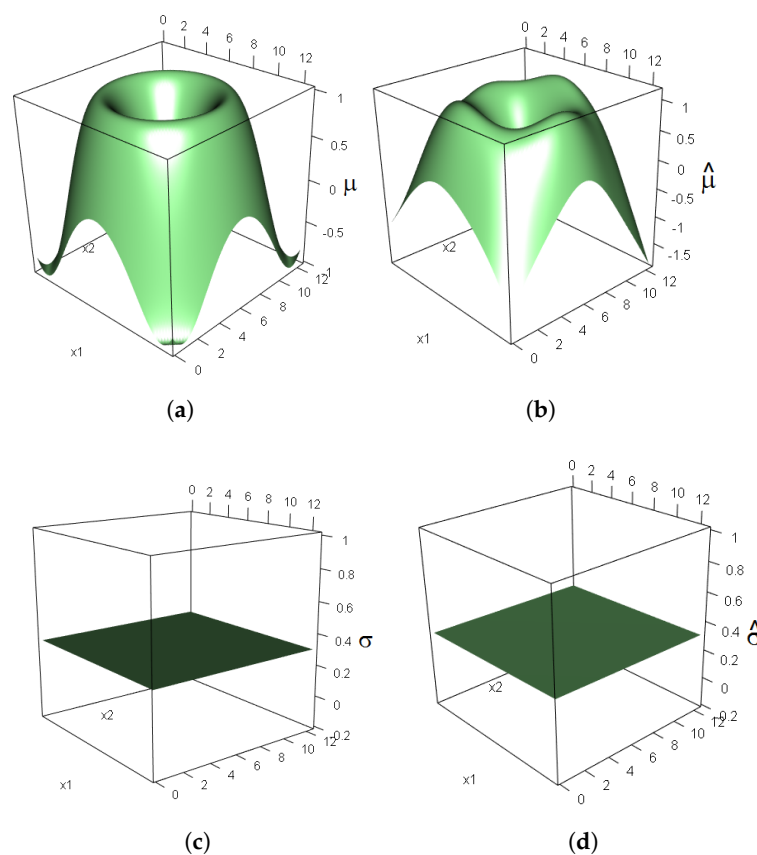


Figure 2. Real functions (a,c) and functions obtained by fitting the parameters (b,d) of a non-stationary generalized extreme value (GEV) model to simulated data with a sample size of $n = 500$.

To evaluate the performance of our model, we reserved a set of 1000 locations to serve as the testing set and then calculated the correlation with respect to their estimated values. Specifically, for the location parameter, we obtained a correlation of 0.99 between the predicted values and the testing data. Typically, when the trend found in the maxima can be modeled using nonlinear functions of a set of covariates, the interpretations of the individual coefficients of the smoothing function lose importance, and they are often meaningless. Therefore, the main information about the trend is provided by the smoothing function constructed from the complete set of estimated parameters.

2.2.6. Maxima in PM₁₀ Pollution Levels

Data Collection

The data correspond to 1238 observations of quarterly block maxima in PM₁₀ between 1 January 1995, and 31 December 2016, obtained at 31 fixed monitoring stations of the Red Automática de Monitoreo Atmosférico (RAMA) network established by the Comisión Ambiental Metropolitana de Mexico City to monitor compliance with ambient air quality standards; this network is part of the Sistema de Monitoreo Atmosférico (SIMAT), a program responsible for ongoing measurements of the main air pollutants in Mexico City.

Data Analysis

We constructed a GEV model of the PM₁₀ maxima in the Mexico City metropolitan area, using multivariate smoothing functions of spatio-temporal covariates, latitude (s_1), longitude (s_2) and time (t), grouped into $X_{(1)}$ and meteorological covariates, wind speed (ws) and relative humidity (rh), grouped into $X_{(2)}$, to fit the trends in the non stationary GEV model. We defined two models to estimate the GEV parameters, the GEV0 model which included the spatio-temporal covariates and the GEV1 model that included the spatio-temporal covariates as well as the meteorological covariates.

According to the proposed approach in Equation (1), we define the joint contribution $\delta_{\theta_t}(X)$ of the set of covariates X corresponding to the θ_t parameter of the GEV distribution, as follows

$$\begin{aligned}\delta_{\mu_t}(X) &= \sum_{i=1}^{p_1} X_{ti}\beta_{1i} + \sum_{i=1}^{p_2} Z_{ti}u_{1i}, \\ \delta_{\log \sigma_t}(X) &= \sum_{i=1}^{p_1} X_{ti}\beta_{2i} + \sum_{i=1}^{p_2} Z_{ti}u_{2i}, \\ \delta_{\kappa_t}(X) &= \kappa\end{aligned}\quad (6)$$

The GEV0 model is a baseline model that incorporates the spatio-temporal covariates $X_{(1)}$ using a multivariate spline structure that implicitly includes the interaction between the covariables of the group, as follows

$$\begin{aligned}\mu_t &= \delta_{\mu_t}(X_{(1)}), \\ \log \sigma_t &= \delta_{\log \sigma_t}(X_{(1)}), \\ \kappa &= \delta_{\kappa_t}(X_{(1)})\end{aligned}\quad (7)$$

The GEV1 model is an extension of the GEV0 model that incorporates in an additive way the meteorological set of covariates $X_{(2)}$ with the structure given by Equation (6).

$$\begin{aligned}\mu_t &= \delta_{\mu_t}(X_{(1)}) + \delta_{\mu_t}(X_{(2)}), \\ \log \sigma_t &= \delta_{\log \sigma_t}(X_{(1)}) + \delta_{\log \sigma_t}(X_{(2)}), \\ \kappa &= \delta_{\kappa_t}(X_{(1)}) + \delta_{\kappa_t}(X_{(2)})\end{aligned}\quad (8)$$

3. Results and Discussion

A descriptive summary of the data is shown at Table 1. Four examples of quarterly block maxima of PM₁₀ levels are presented in Figure 3. The U.S. and Mexican standard for PM₁₀ pollution levels is 150 µg/m³. An analysis of Table 1 shows that in the area of Villa de las Flores, the permissible level was exceeded by more than three quarters of all measured extreme values. At three of the monitoring stations, CUA = Cuautitlán, NET = Netzahualcoyotl and XAL = Xalostoc, all observations exceeded the permitted standard level, this is because these locations have a high population density and more concentrated industry. The peak PM₁₀ concentrations exceeded 1000 µg/m³ at CES = Cerro de la Estrella, MER = Merced, SAG = San Agustín, VIF = Villa de las Flores and XAL = Xalostoc; at SAG = San Agustín station, the level recorded was 10 times higher than the recommended safe limit.

Table 1. Descriptive summary information on the extreme values of particulate matter less than 10 micrograms in diameter (PM₁₀) at the stations considered in the study.

Name	Symbol	Long(W)	Lat(N)	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Acolman	ACO	99°07' 03.89"	19°28'04.4"	76	166.2	197.5	230.5	252.5	535
Ajusco Medio	AJM	99°12' 16.54"	19°31'44.67"	92	99	109	121.1	137	175
Atizapan	ATI	99°04' 56.64"	19°31'33.58"	108	137	185	203.9	256	387
Benito Juárez	BJU	99°07' 10.53"	19°25'28.59"	102	118.5	127	265.8	274.2	707
Camarones	CAM	99°12' 14.88"	19°19'30.52"	102	150	181.5	189.8	213.2	462
Cerro de la Estrella	CES	99°04' 28.84"	19°20'05.03"	130	279	373	444	617.5	1023
Chalco	CHO	99°01' 34.02"	19°25' 16.14"	150	207.5	283	272.3	341	401
Cuajimalpa	CUA	99°10' 37.82"	19°36'09.15"	75	108	119	131.7	168	191
Cuautitlán	CUT	99°05' 47.72"	19°39'29.60"	198	267.2	297	289.5	305	427
FES Acatlán	FAC	99°00' 38.03"	19°14' 47.25"	98	167.5	248	272.3	364	758
Hangares	HAN	99°08' 59.97"	19°18' 52.12"	117	228	302	333.8	369	959
Hospital General de México	HGM	99°14' 36.68"	19°28' 56.90"	96	143.8	153	190	193.2	376
Investigaciones Nucleares	INN	99°05' 01.04"	19°25' 13.86"	69	71.25	120	142	190.8	259
Iztacalco	IZT	99°12' 00.39"	19°21' 57.12"	78	128.5	186	230.8	317	569
La Villa	LVI	99°01' 49.16"	19°31' 58.68"	118	203.8	286	309	355.5	871
Merced	MER	99°07' 23.53"	19°20' 08.48"	109	187.5	290.5	357.8	437	1233
Miguel Hidalgo	MGH	99°07' 03.50"	19°23' 03.88"	100	109	121	134.7	137	230
Milpa Alta	MPA	98°54' 43.21"	19°38' 07.8"	119	123.5	128	150.3	166	204
Netzahualcoyotl	NET	99°10' 11.25"	19°28' 06.25	298	580.5	737	722.4	887	991
Pedregal	PED	99°04' 25.96"	19°21' 38.85	94	146	189	233	284	884
Plateros	PLA	99°09' 07.94"	19°24' 41.82	112	178	233	241.2	294	440
San Agustín	SAG	99°15' 46.31"	19°21' 26.48"	104	216.5	346	430.9	571	1570
Santa Fe	SFE	99°11' 54.96"	19°43' 19.86"	91	131	149	159.2	182	267
Santa Ursula	SUR	98°53' 09.91"	19°16' 01.01"	100	164	237	265.8	335	603
Tlahuac	TAH	99°15' 14.87"	19°34' 37.06"	91	183	281	336.9	463.2	977
Taxqueña	TAX	99°17' 30.13"	19°21' 55.12"	128	188.5	262	280.6	334.5	513
Tlalnepantla	TLA	99°12' 09.57"	19°24' 14.58"	121	190	236	317.3	374	912
Tultitlán	TLI	99°12' 27.87"	19°16' 19.77"	41	203.2	293	303	368.5	828
UAM Iztapalapa	UIZ	99°09' 34.54"	19°22' 13.67"	105	158	172	201.8	238	539
Villa de las Flores	VIF	99°22' 49.87"	19°17' 31.08"	82	243.8	380.5	405.8	470.8	1269
Xalostoc	XAL	98°59' 24.68"	19°10' 36.83"	187	330	443	497	609	1076

Figure 4 shows boxplots of the PM₁₀ maxima at each of the 31 monitoring stations considered in the study. The sites with lower PM₁₀ maxima, less than 151 µg/m³ on average, are located in MGH = Miguel Hidalgo, MPA = Milpa Alta, AJM = Ajusco Medio and CUA = Cuajimalpa. Relatively well-preserved areas can still be found at these sites, which also have the lowest industrial activity and urban growth in the study area. By contrast, we also found sites with more than 723 µg/m³ PM₁₀ on average, such as NET = Netzahualcoyotl, XAL = Xalostoc, CES = Cerro de la Estrella and MER = Merced, which are characterized by high industrial activity and heavily traveled paved and unpaved roads with heavy vehicular traffic [6].

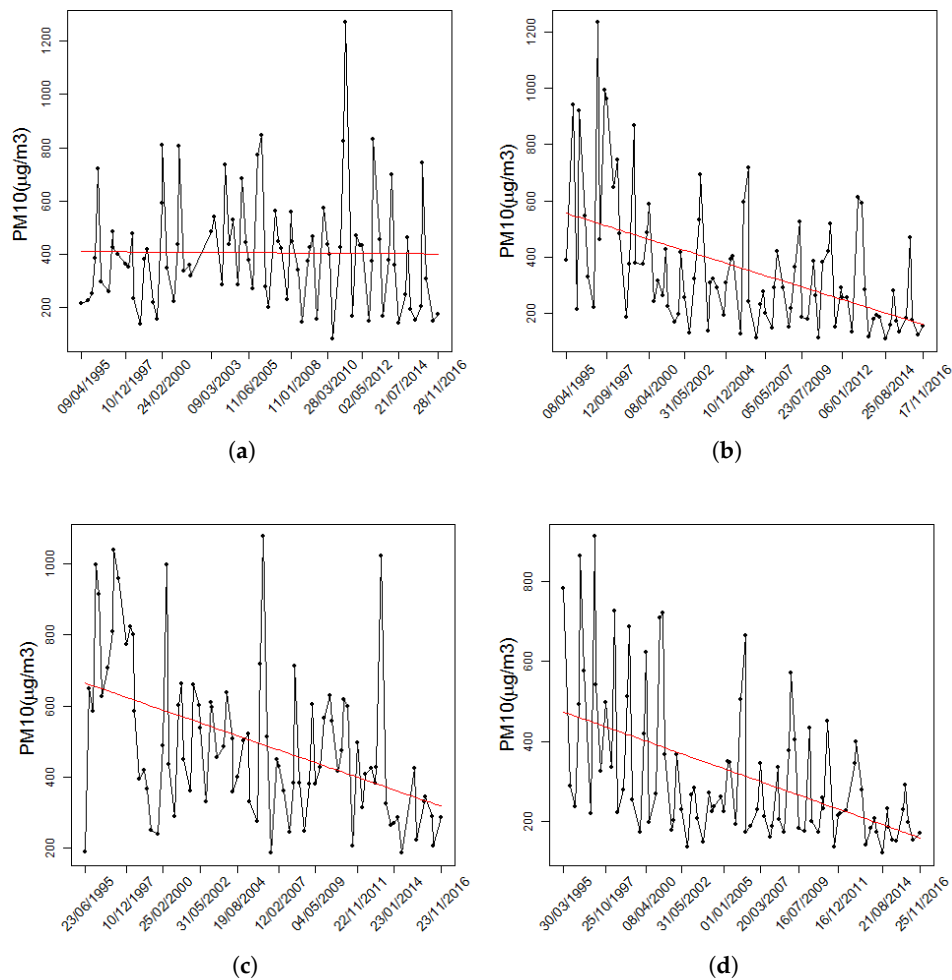


Figure 3. Example of quarterly block maxima of PM₁₀ levels at (a) VIF = Villa Flores; (b) MER = Merced; (c) XAL = Xalostoc and (d) TLA = Tlanepantla. Linear trend over time are indicated in red.

In order to determine the significance of the meteorological variables in the non-stationary GEV model, we adjusted the GEV0 and GEV1 models by using the method of maximum likelihood (ML) and penalized maximum likelihood (PML) and we compared these two models through the deviance statistics. The results presented in Table 2 show that the model GEV1 is not statistically better than the baseline model GEV0, therefore the set of meteorological covariates wind speed (ws) and relative humidity (rh) did not present evidence at the 95% level to be included in the non-stationary model GEV for the modeling of PM₁₀ maxima. Table 2 shows the effect of penalization of the parameters in the model, in the case of maximum likelihood method none regularization was performed, therefore $\ell_{PL}(b_{(1)}, b_{(2)}, \kappa)$ is considerably greater than the case of the penalized maximum likelihood. The maximum likelihood estimators have desirable statistical properties, however penalized estimators are preferred because they control the overfitting and reduce the instability of the estimates. In this work penalized estimators are used as initial values for the Bayesian hierarchical model, in order to reduce the number of chains necessary to reach the stationary distribution of the sampling MCMC algorithm.

Table 2. Statistical comparison of the GEV0 model against the GEV1 model.

% Method	Model	$\ell_n(\underline{y} \mid \mu_t, \sigma_t, \kappa)$	$\ell_{PL}(b_{(1)}, b_{(2)}, \kappa)$	Deviance	<i>p</i> -value
ML	GEV0	−4607.8	−21,763.3	32	0.12
	GEV1	−4591.8	−21,737.2		
Penalized ML	GEV0	−4845.8	−28.1	26.6	0.33
	GEV1	−4832.5	−28.7		

Once the set of covariates involved in the model is determined, a hierarchical Bayesian model was fitted to analyse the PM₁₀ maxima in the Mexico City metropolitan area. At the first level, we modeled the PM₁₀ maxima using the generalized extreme value distribution; at the second level, we used multivariate smoothing spline functions to model nonstationary spatio-temporal extremes; and finally, at the third level, we assumed a priori distribution functions for the parameters of the model. Unfortunately, the posterior density in Equation (4) is not analytically tractable, so we implemented our model via the random walk Metropolis-Hastings algorithm to draw samples of the unknown parameters.

We performed the analysis using the statistical software *R* (version 3.3.1, R Foundation for Statistical Computing, Vienna, Austria). We fitted our Bayesian hierarchical model with the conditions expressed in Equation (6), setting $p_2 = 10$. We ran 70,000 MCMC samples after a burn-in period of 60,000 iterations, with thinning every fifth iteration. A look at the log-likelihood of the chain revealed convergence toward the stationary distribution of the MCMC algorithm.

Evidently, the extreme values of PM₁₀ vary over time and from area to area because of local conditions, such as the topographic setting or the local wind. This fact justifies the modeling of the location parameter with respect to time and space. We verified this behavior by means of the estimates presented in Table 3, which show that the magnitudes of the studied maxima tend to decrease over time. Similarly, the boxplots in Figure 4 show that the distribution of the maxima is not the same in all monitored locations; therefore, a suitable model for the PM₁₀ maxima must include temporal and spatial variations and their possible interactions. One of the strengths of our study is that the proposed model implicitly includes interactions between covariates, whereas most models consider only the main effects through additive or linear functions.

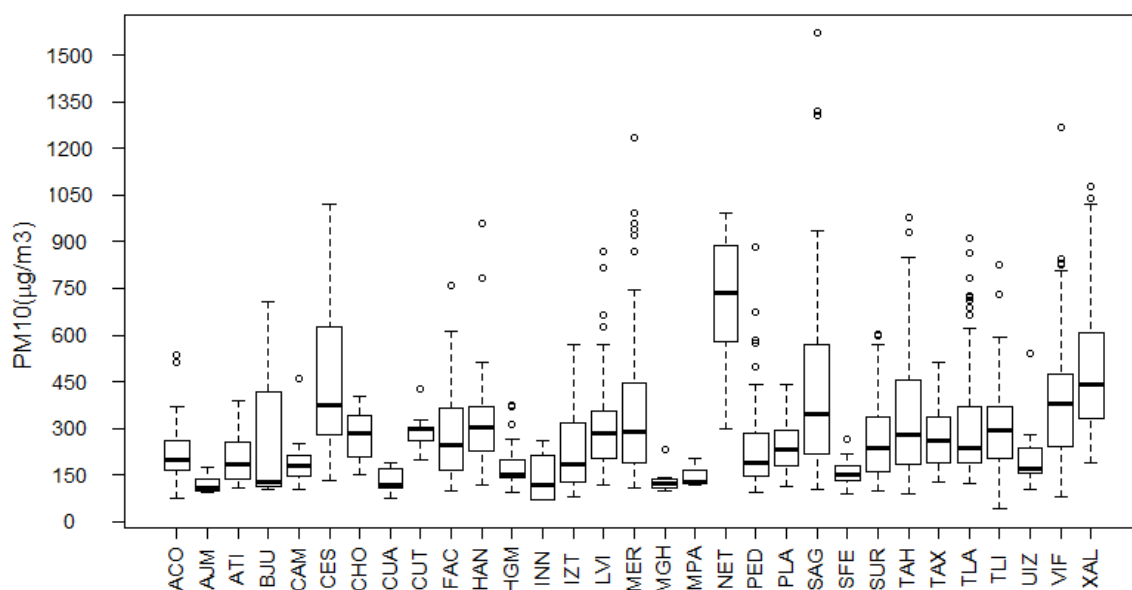


Figure 4. Boxplots of the PM₁₀ maxima at 31 monitoring stations in the Mexico City metropolitan area.

The estimates for the parameters of model Equation (1) corresponding to the coefficients of the main effects of the space-time covariates in the location and scale smoothing functions as well as the shape parameter are shown in Table 3. In this case, the vector of parameters β_1 corresponds to μ , and the vector β_2 corresponds to $\log \sigma_t$. We noticed that the effect $\beta_{(1)t}$ of time on the location parameter is negative and that the effects $\beta_{(1)s1}$ and $\beta_{(1)s2}$ of the longitude and latitude are both positive. Because of the smoothing term in Equation (6) and the standardization of the covariates used in the model, it is not possible to establish a direct relation between the parameters estimated in Table 1 and the parameters of the GEV distribution in model Equation (6), except through the terms $\beta_{(1)0}$ and $\beta_{(2)0}$, which can be interpreted as the values of the location and scale parameters, respectively, of the GEV distribution in the middle of the studied period and in the center of the geographic area considered in the study.

Table 3. Estimates and 95% credible intervals of the nonstationary GEV model for the PM₁₀ maxima.

% Parameter	Mean	95% CI
$\beta_{(1)0}$	235.6905	(235.367, 236.0144)
$\beta_{(1)t}$	-46.1495	(-46.4354, -45.8724)
$\beta_{(1)s1}$	23.609	(23.3706, 23.8377)
$\beta_{(1)s2}$	25.6023	(25.3444, 25.8610)
$\beta_{(2)0}$	4.6134	(4.6112, 4.6157)
$\beta_{(2)t}$	-0.3365	(-0.3387, -0.3344)
$\beta_{(2)s1}$	0.2200	(0.2179, 0.2220)
$\beta_{(2)s2}$	0.1687	(0.1667, 0.1707)
κ	0.1715	(0.1694, 0.1737)

The spatial smoothing for the location and scale functions is shown in Figure 5. The estimation function for the location parameter in the latitude–longitude coordinate system shows that the PM₁₀ maxima tend to increase in the northwesterly direction. The scale parameter also increases in the northwesterly direction, but unlike for the location parameter, a slight decrease is observed in the region close to Iztacalco. The estimate of the shape parameter is 0.1715, and its 95% credible interval is (0.1694, 0.1737).

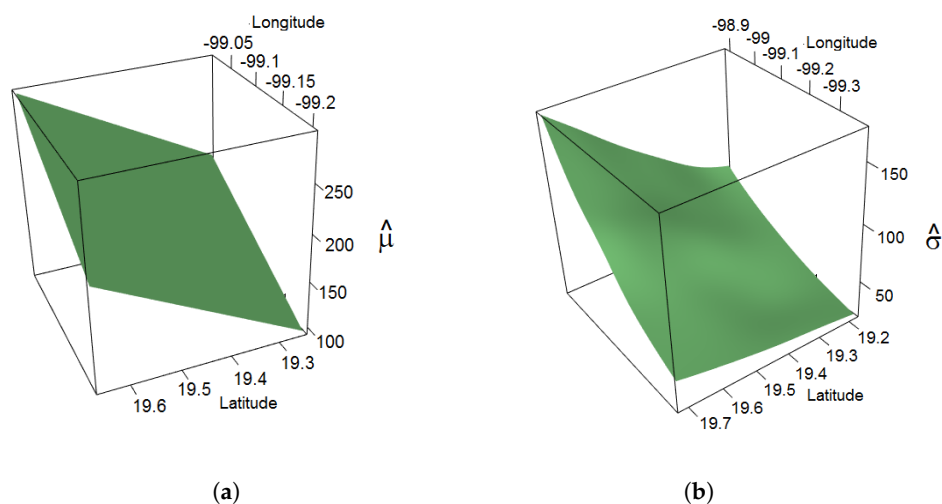


Figure 5. Estimated spatial smoothing of the (a) location parameter and (b) scale parameter in the year 2016.

We have developed and validated a model of extreme values using a robust and solidly-based theory, an important application of the results of our study is through a risk map or return level map. The return level Z_p is the threshold at which an extreme value is exceeded with probability p , which is expected to occur once every $1/p$ years [34]. Figure 6 shows the return levels for a return period of 25 years. Accordingly, the trend of increased risk in the northwest of the study area is preserved, with the greatest levels of risk in areas close to VIF = Villa Flores, SAG = San Agustín and ACO = Acolman and a lower level of risk in the area surrounding Pedregal.

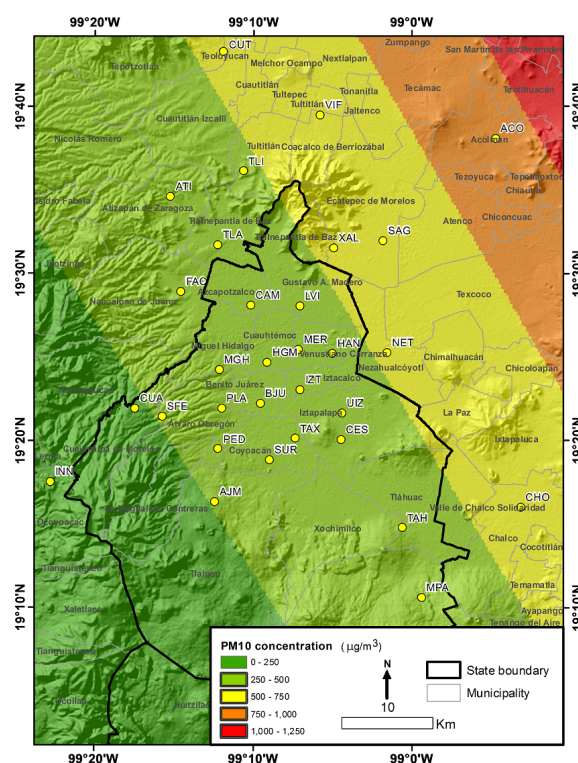


Figure 6. Return level surface for a return period of 25 years for the study region.

Previous studies have analysed extremes of air pollutants using the generalized extreme value distribution [30,36]. However, the majority of these investigations have used information from a single site and consequently have ignored the aspect of spatial variability, which may result in underestimation of the GEV parameters. Several approaches to the spatio-temporal modeling of extreme values [19,28] have been implemented for environmental extreme data analysis. Therefore, on the basis of these studies, we have proposed a model that offers structural advantages for the modeling of extreme values.

Similar to the findings reported by studies in other countries [3], the extreme PM_{10} concentrations studied here exhibited spatial variations within the study area.

These results were consistent with the environmental characteristics of each monitored region: the locations with the highest industrial and population densities showed higher concentrations of PM_{10} , and consequently, in these locations, the GEV model yielded high estimates of the location and scale parameters and the 25-year return map showed greater risks of extreme future events.

Most PM_{10} modeling studies on short-term dynamics have found that weather covariables such as wind speed or temperature are significant in the model [37]. This is not always the case in studies with a longer time horizon, where the conclusions for the long term are more robust. One of the reasons is due to the temporality of the data, daily maxima present a different association with

respect to the meteorological covariates than the maxima obtained in a longer time window, such as quarterly maxima.

One of the findings of our study is the negative contribution of the effect of time on the trend of PM_{10} maxima; an explanation of this phenomenon is the continuous implementation of new emission control policies and the continuous revision of environmental norms in the Mexico City, which have recently become increasingly strict, mainly by regulating the traffic and restricting the automotive vehicles that can be used i.e., vehicle's model year. Additionally, the Mexico City government have implemented a website for monitoring air quality (<http://www.aire.cdmx.gob.mx/default.php>), thus environmental contingency alerts are activated to mitigate the adverse effect of air pollutants on public health.

4. Conclusions

Recent years have seen a growing interest in the monitoring of PM_{10} air pollution because of the multiple health problems it causes in densely populated areas. In the Mexico City metropolitan area, PM_{10} air pollution often originates from dust from roadways and organic black carbon formed during combustion processes. This has led to is an ongoing public health problem over the past several years affecting a large section of the population. Recently, due to population growth and the increase in the number of combustion vehicles and industries, the PM_{10} air pollution problem has worsened. Therefore, it is extremely important to study the spatio-temporal trend of the PM_{10} maxima and provide this information to the management authorities so that prevention standards and policies can be reviewed and adapted to prevent extreme cases of PM_{10} air pollution.

Because wind speed (*ws*) and relative humidity (*rh*) are statistically significant covariates in most models for the short-term forecast for PM_{10} concentration, we consider including these covariates in the non-stationary GEV model. However by using the deviance statistic and the chi-square test, we found that at a credible level of 95%, the model with these covariates was not statistically better than the model that did not include these variables. Therefore, wind speed and relative humidity were not statistically significant in the non-stationary GEV model with quarterly block maxima of PM_{10} levels. The return levels for a return period of 25 years revealed a clear spatial trend of increased levels of PM_{10} in the northwest study area and a decreasing trend in the extreme values over time.

In this study, we implemented a methodology to analyse the nonstationary extreme data and to perform a spatio-temporal study of the maxima of breathable particulate matter pollution (PM_{10}) levels in the Mexico City metropolitan area. These PM_{10} levels usually vary in space and time, and can potentially include significant spatio-temporal interactions. Therefore, the commonly used models can underperform the GEV estimates and consequently result in inaccuracies. We achieved this using an analysis of non-stationary extreme values, in which we used a combination of existing traditional statistical techniques. We used the maximum likelihood method to perform a variable selection step, the penalized maximum likelihood estimators to obtain initial values and reduce the convergence time of the MCMC algorithm and fitted the estimators using a Bayesian approach to eliminate potential invalid parameter values. The combination of these statistical techniques gives support and solidity to our results.

We proposed a modification to the Bayesian estimation methods used in the previous studies related to the analysis of extreme values applied to environmental areas. In our model, the covariates are included in the generalized extreme value distribution through multivariate spline functions and, therefore, interactions between the covariates are also considered in the model. We observed that this approach can be easily extended to the modeling of extreme events and the generation of risk maps for air pollution, rainfall, heat waves, wind speed, etc. The results of the simulations conducted led us to conclude that the methodology is adequate and reliable for this type of study. Additionally, as a first attempt, time has been treated as a covariate. Extensions of this work should consider a more general model for spatio-temporal analysis, a specific analysis of the prior distributions of the parameters, and a method for determining the correct number of nodes in the multivariate spline functions.

Acknowledgments: This work was supported by the Program for Professional Teaching Development (PRODEP) under Grant Number DSA/103.5/16/10481.

Author Contributions: Alejandro Ivan Aguirre-Salado and Humberto Vaquera-Huerta conceived and designed the experiment and, wrote some parts of the paper. Alejandro Ivan Aguirre-Salado designed the programming code for processing the data and wrote the paper. Carlos Arturo Aguirre-Salado and Carlos Soubervielle-Montalvo helped in some spatial analyses and wrote some parts of the paper. Silvia Reyes-Mora, Ana Delia Olvera-Cervantes and Guillermo Arturo Lancho-Romero contributed in the mathematical rationale of the model and wrote some parts of the paper. All authors revised and approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Joseph, A.; Sawant, A.; Srivastava, A. PM₁₀ and its impacts on health—A case study in Mumbai. *Int. J. Environ. Health Res.* **2003**, *13*, 207–214.
- Kumar, P.; Jain, S.; Gurjar, B.; Sharma, P.; Khare, M.; Morawska, L.; Britter, R. New Directions: Can a “blue sky” return to Indian megacities? *Atmos. Environ.* **2013**, *71*, 198–201.
- Wang, X.; Guo, Y.; Li, G.; Zhang, Y.; Westerdahl, D.; Jin, X.; Pan, X.; Chen, L. Spatiotemporal analysis for the effect of ambient particulate matter on cause-specific respiratory mortality in Beijing, China. *Environ. Sci. Pollut. Res.* **2016**, *23*, 10946–10956.
- Kim, K.H.; Kabir, E.; Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **2015**, *74*, 136–143.
- Garcia, J.; Teodoro, F.; Cerdeira, R.; Coelho, L.; Kumar, P.; Carvalho, M. Developing a methodology to predict PM₁₀ concentrations in urban areas using generalized linear models. *Environ. Technol.* **2016**, *37*, 2316–2325.
- Edgerton, S.A.; Bian, X.; Doran, J.; Fast, J.D.; Hubbe, J.M.; Malone, E.L.; Shaw, W.J.; Whiteman, C.D.; Zhong, S.; Arriaga, J.; et al. Particulate air pollution in Mexico City: A collaborative research project. *J. Air Waste Manag. Assoc.* **1999**, *49*, 1221–1229.
- Thishan Dharshana, K.; Coowanitwong, N. Ambient PM₁₀ and respiratory illnesses in Colombo city, Sri Lanka. *J. Environ. Sci. Health A* **2008**, *43*, 1064–1070.
- Elbayoumi, M.; Ramli, N.A.; Yusof, N.F.F.M.; Al Madhoun, W. Spatial and seasonal variation of particulate matter (PM₁₀ and PM_{2.5}) in Middle Eastern classrooms. *Atmos. Environ.* **2013**, *80*, 389–397.
- Paschalidou, A.K.; Karakitsios, S.; Kleanthous, S.; Kassomenos, P.A. Forecasting hourly PM₁₀ concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environ. Sci. Pollut. Res.* **2011**, *18*, 316–327.
- El Adlouni, S.; Ouarda, T.; Zhang, X.; Roy, R.; Bobée, B. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resour. Res.* **2007**, *43*, W03410.
- Sang, H.; Gelfand, A.E. Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **2010**, *15*, 49–65.
- Jenkinson, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **1955**, *81*, 158–171.
- Smith, R.L. Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **1985**, *72*, 67–92.
- Hosking, J.R.M.; Wallis, J.R.; Wood, E.F. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **1985**, *27*, 251–261.
- Hosking, J.R. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B* **1990**, *52*, 105–124.
- Madsen, H.; Rasmussen, P.F.; Rosbjerg, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water Resour. Res.* **1997**, *33*, 747–757.
- Gaetan, C.; Grigoletto, M. A hierarchical model for the analysis of spatial rainfall extremes. *J. Agric. Biol. Environ. Stat.* **2007**, *12*, 434–449.
- Reich, B.; Shaby, B.; Cooley, D. A Hierarchical model for serially-dependent Extremes: A study of heat waves in the Western US. *J. Agric. Biol. Environ. Stat.* **2014**, *19*, 119–135.
- Cooley, D.; Sain, S.R. Spatial hierarchical modeling of precipitation extremes from a regional climate model. *J. Agric. Biol. Environ. Stat.* **2010**, *15*, 381–402.
- Leadbetter, M.R.; Lindgren, G.; Rootzen, H. *Extremes and Related Properties of Random Sequences and Processes*; Springer: New York, NY, USA, 1983; p. 336.

21. Wang, X.L.; Zwiers, F.W.; Swail, V. North Atlantic Ocean wave climate scenarios for the 21st century. *J. Clim.* **2004**, *17*, 2368–2383.
22. Kharin, V.V.; Zwiers, F.W. Estimating extremes in transient climate change simulations. *J. Clim.* **2005**, *18*, 1156–1173.
23. Weissman, I. Estimation of parameters and large quantiles based on the k largest observations. *J. Am. Stat. Assoc.* **1978**, *73*, 812–815.
24. Tawn, J. Bivariate extreme value theory: Models and estimation. *Biometrika* **1988**, *75*, 397–415.
25. Scarf, P.A. Estimation for a four parameter generalized extreme value distribution. *Commun. Stat. Theory M* **1992**, *21*, 2185–2201.
26. Rosen, O.; Cohen, A. Extreme percentile regression. In *Statistical Theory and Computational Aspects of Smoothing*; Springer: New York, NY, USA, 1996; pp. 200–214.
27. Pauli, F.; Coles, S. Penalized likelihood inference in extreme value analyses. *J. Appl. Stat.* **2001**, *28*, 547–560.
28. Bocci, C.; Caporali, E.; Petrucci, A. Geoadditive modeling for extreme rainfall data. *ASTA Adv. Stat. Anal.* **2013**, *97*, 181–193.
29. Yee, T.W.; Stephenson, A.G. Vector generalized linear and additive extreme value models. *Extremes* **2007**, *10*, 1–19.
30. Rodriguez, S.; Reyes, H.; Perez, P.; Vaquera, H. Selection of a subset of meteorological variables for ozone analysis: Case study of pedregal station in Mexico City. *Environ. Sci. Eng. A* **2012**, *1*, 11–20.
31. Cannon, A.J. A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology. *Hydrol. Process.* **2010**, *24*, 673–685.
32. Jauregui, E. Local wind and air pollution interaction in the Mexico basin. *Atmósfera* **1988**, *1*, 131–140.
33. Fisher, R.A.; Tippett, L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Philos. Soc.* **1928**, *24*, 180–190.
34. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2001.
35. Figueiredo, M.A. On Gaussian radial basis function approximations: Interpretation, extensions, and learning strategies. In Proceedings of the 15th International Conference on Pattern Recognition (ICPR-2000), Barcelona, Spain, 3–7 September 2000; Volume 2, pp. 618–621.
36. Küttchenhoff, H.; Thamerus, M. Extreme value analysis of Munich air pollution data. *Environ. Ecol. Stat.* **1996**, *3*, 127–141.
37. Lin, C.Y.; Chiang, M.L.; Lin, C.Y. Empirical model for evaluating PM₁₀ concentration caused by river dust episodes. *Int. J. Environ. Res. Public Health* **2016**, *13*, 553.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).