International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 05 - Issue 04 || April 2019 || PP. 18-36

Prediction of Concentrations of Suspended Particle Levels of 10 micrometers (PM10) in México City with Probability Distribution Functions, Trend 2010-2018.

M.Sc. Zenteno Jimenez José Roberto
Geophysical Engineering, National Polytechnic Institute, Mexico City,
ESIA-Unit Ticoman Delegation Gustavo A. Madero
jzenteno@ipn.mx

Abstract: The study includes an analysis of data from 2010 to 2018, it was proposed to obtain the best or better features probability distribution model the concentrations of PM10 in México City using the following pdf, probability distribution function gama, probability density function of extreme value, probability distribution function gumbel and probability distribution function weibull, to obtain estimators by method maximum likelihood and moments was used and helped the Matlab 2017 program, assessment forecasting model RMSE, MSE, coefficient of determination and Index of Approximation, at the same time an analysis is made to observe its tendency within the period to data of concentrations of daily maximum after corroborating with the official page of air of méxico city, the trend analysis is done with Bayesian Inference.

Keywords: Particulate Matter of 10 micrometers, probability distributions, adjustment indicators, Extreme Value Theory, Bayesian Inference

The particles come from smoke trucks and factories, fire, plant pollens, spores of fungi, skin emerges from body, oxides and metals are contaminants breathed by people of México City.

The classification of particulate material. In general, and for purposes of environmental pollution and health effects, Particulate material is classified according to its size expressed in micrometers (one micrometer is one thousandth of a millimeter, the diameter of a hair is between 70 and 80 micrometers). Particulate material with an aerodynamic diameter less than or equal to 10 micrometers is referred to as PM10.

From the toxicological point of view, particles greater than 10 micrometers are efficiently retained by the upper respiratory tract, including the nose, larynx, larynx and trachea, but smaller particles (PM4.7- PM0.65) can penetrate up to the bronchi and even the alveoli (PM 0.65-PM0.43); the deepest part of the respiratory system.

Particulate coarse material (the largest particles) usually contains soil and dust derived from the action of the wind that results from agricultural activities, unpaved roads, buildings, some industrial activities or simply from the action of wind on the bare ground. It also includes pollen particles, mold spores and parts of plants and insects and, near the coasts, particles produced by the marine aerosol.

In general, contamination with particulate material has effects on the respiratory and cardiovascular system. Some estimates indicate that particulate matter (PM) pollution is the cause of around 2.1 million deaths per year on the planet; approximately four times more than deaths attributable to ozone pollution.

Given the stochastic nature of atmospheric processes, concentrations of air pollutants can be treated as random variables with measurable statistical properties. If certain conditions are the statistical characteristics of pollutant concentrations, they can be described by probability density functions. Probability density functions (pdf) have been widely used in recent years in a variety of applications, where smoothing data.

Interpolation or extrapolation is needed (Wilks, 1995). Specifically, in the atmospheric sciences the most characteristic applications include the approximation of the frequency of exceedances of the critical levels of concentration and the estimation of the reduction of emissions, required for the standard of air quality objectives (Georgopoulos and Seinfend, 1982; Abatzoglou et al., 1996; Burkehardt et al., 1998; Morel et al., 1999).

The maximum likelihood method is considered advantageous for estimating parameters compared to moment methods (which is also occasionally used). On the other hand, the maximum likelihood method requires a great processing power due to the complex numerical calculations involved, when large data sets are analyzed, computational time increases substantially.

Probability Distribution Functions and Methodology

Four probability distribution functions were used, which are the function of gamma distribution, distribution function GEV, Gumbel distribution function maximum, and the Weibull distribution function.

Table 1. Probability Distribution Functions and Their Parameters.

Distribution	Probability density function	parameters
GEV	$f(x) = \left(\frac{1}{\sigma}\right) exp^{-\left((1+kz)^{(-\frac{1}{k})}\right)(1+kz)^{(-1-\frac{1}{k})}}$	***K* shape** **σ scale** **μ location**
Gumbel	$f(x) = \left(\frac{1}{\sigma}\right) exp^{(-z-exp^{-z})}$ $z = \frac{x-\mu}{\sigma}$	**σ scale** **μ location**
Weibull	$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} exp^{\left(-\frac{x^{\alpha}}{\beta}\right)}$	**The α form** **The scale β**
Spectrum	$f(x) = \frac{Beta^{alfa}}{\Gamma(alfa)} x^{alfa-1} e^{-Betax}$	$Beta = \frac{\sum x_i^2}{\sum x_i} - \frac{\sum x_i}{N}$ $Alfa = \frac{(\sum x_i)^2}{N\sum x_i^2 - (\sum x_i)^2}$

Statistical Adjustment Estimators

Indicators deviation of a group of data relative to a model can be used to assess the goodness of fit between the two. Among the most common indicators they are as follows: RMSE, MAE, NRMSE, CV-MRSE, SDR, and. Those who were used to determine the distribution that best fit the data gave. Are the mean square error (RMSE), mean square error (MSE), prediction accuracy (AP), IA and determination coefficient (R^2)

Table 2. Adjustment Estimator

Estimator	Equation
Error Measures **Root Mean Square Error**	$RMSE = \sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(Pi-Oi)^2}$
Error Measures **Mean Square Error**	$MSE = \left(\frac{1}{N}\right)\sum_{i=1}^{N}(Pi-Oi)^2$
Accuracy Measures **Coefficient of Determination**	$R^2 = \left(\frac{\sum_{i=1}^{N}(Pi-P)(Oi-O)}{NS_pS_o}\right)^2$
Accuracy Measures **Index of Accuracy**	$IA = 1 - \frac{\sum_{i=1}^{N}(Pi-Oi)^2}{\sum_{i=1}^{N}(\lvert Pi-O\rvert - \lvert(Oi-O)\rvert)^2}$

Notation: N = number of observations, P_i= predictive valúes, O_i= observed values, P = average of predicted values, **O** = average of the observed values, S_p= Standard Deviation of Predicted valúes, S_o = Standard deviation of the observed valúes.

Study Área

The México city in its geographical location is located in a closed or almost closed basin, which in all directions is north, south, east or west, adjoins a mountain range or mountain pass, which is the highest altitude with volcanoes to the east the Popocatepetl and the Iztaccihualt, which the circulation of wind and the dispersion of pollutants makes it difficult, both for suspended particles and for other pollutants.

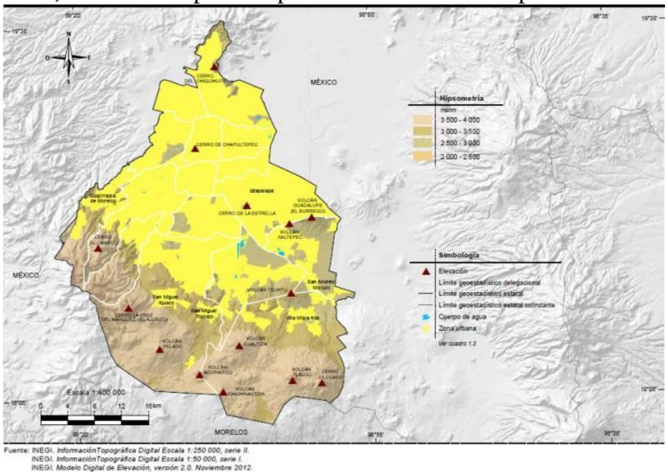


Figure 1. Relieve of Mexico City (Source: https://www.paratodomexico.com/)

Statistical Data Description

In the table below we can see the features of the database which show a 6% or unread null values.

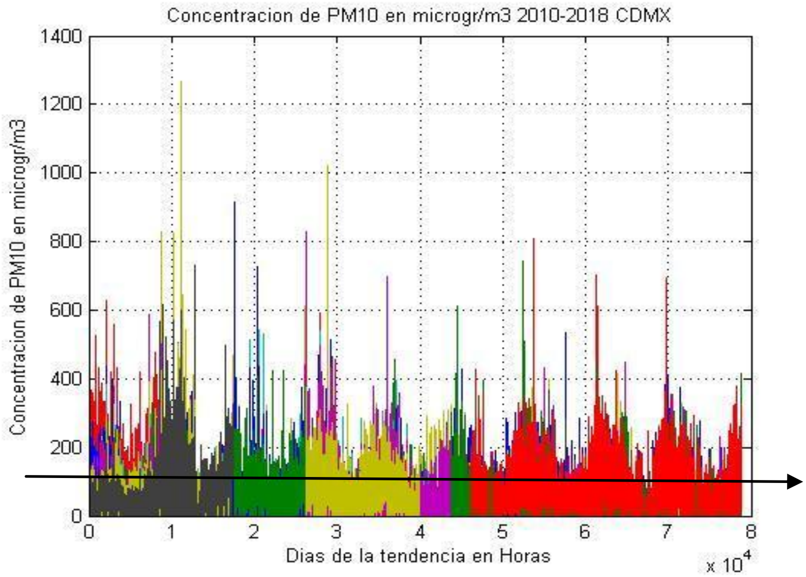


Figure 2. Concentrations of PM10 Daily Mexico City 2010-2018

Table 3. Description of PM10 Data 2010-2018 Trend of Media Concentration Statistics

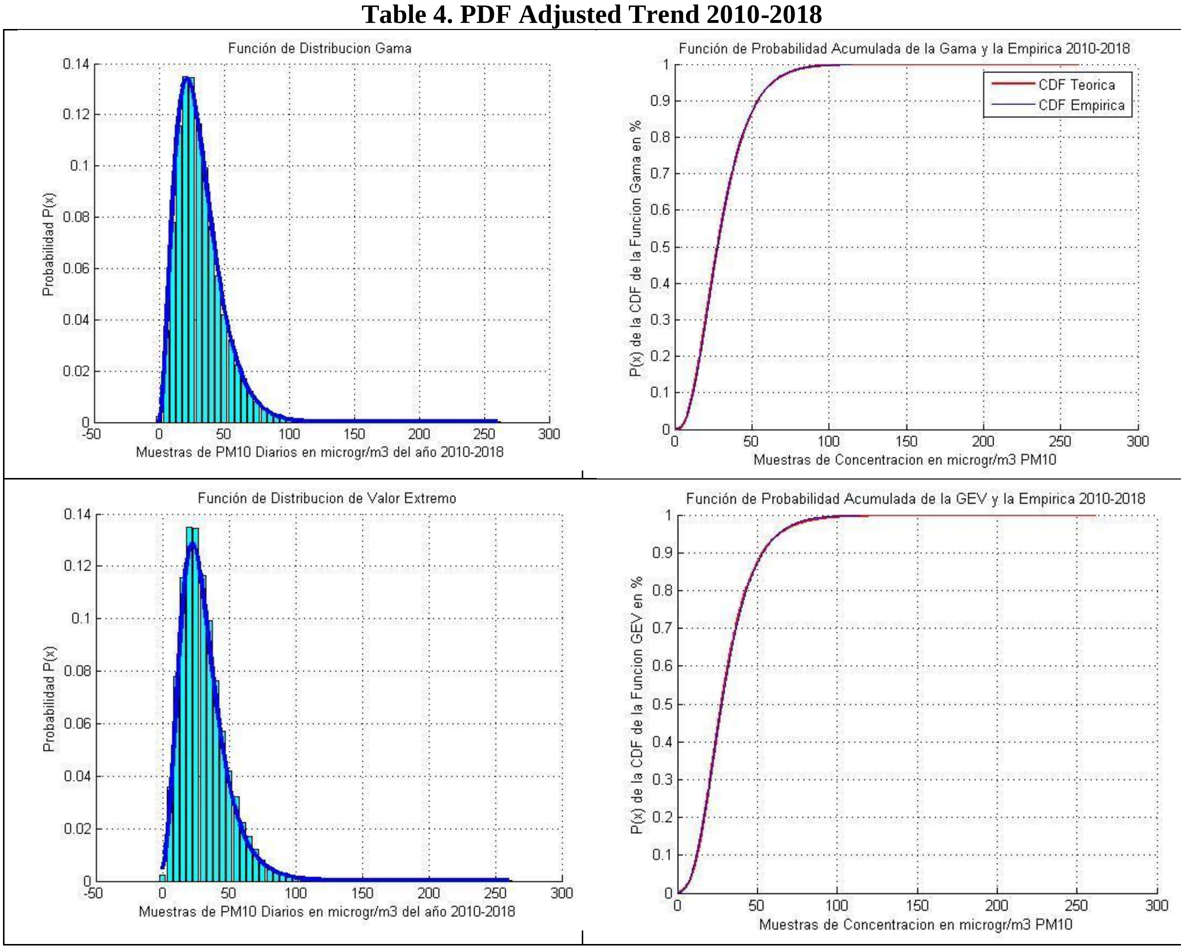
Number of Data	78888
Minimum	0.3750**µ**gr $/m^3$
Maximum	261.95**µ**gr $/m^3$
Mean	30.96**µ**gr $/m^3$
Variance	289.00**µ**gr $/m^3$
Standard deviation	17.23**µ**gr $/m^3$
Median	27.83**µ**gr $/m^3$

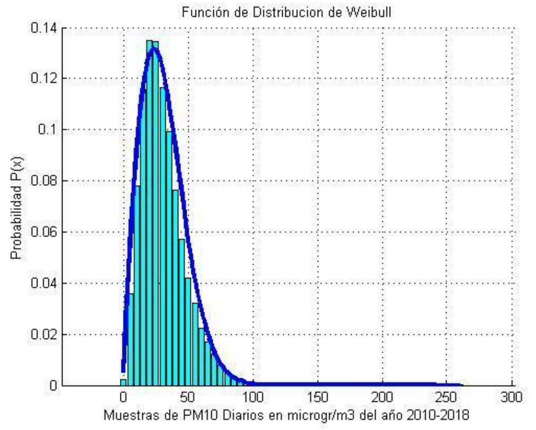
Results

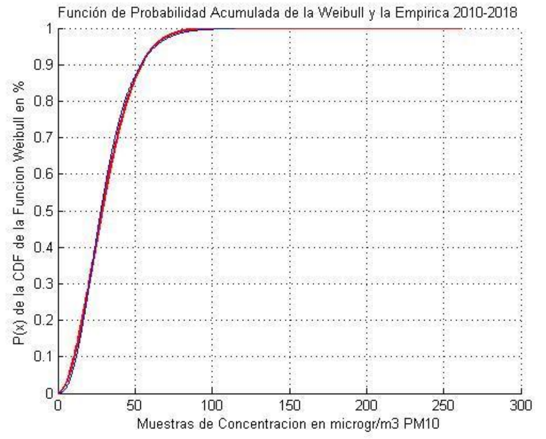
Table 4. Parameters of Estimation and Indicators Adjustment Trend 2010-2018

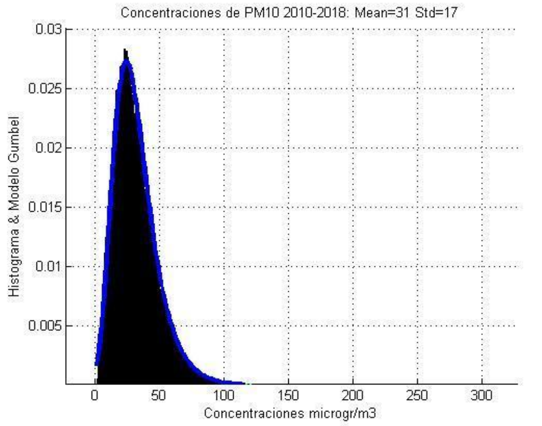
Distribution	Dear parameters	RMSE	MSE	R^2	IA	Kolmogorov-Smirnov	Chi Test
GEV	K = 0.0422 sigma = 8.12 mu = 22.97	.4462	.1991	.8108	.6517	0	h = 0 p = 0.4971
Gumbel	mu = 23.21 sigma = 0.074	.7528	0.5667	.5727	.3465	---	h = 0 p = 0.4834
Weibull	Alpha Beta = 1.90 = 35	.4513	.2037	.8020	.6486	0	h = 0 p = 0.2385
Spectrum	Alpha = 3.242beta = 9,552	.4439	.1970	0.8150	.6555	0	h = 0 p = 0.4966

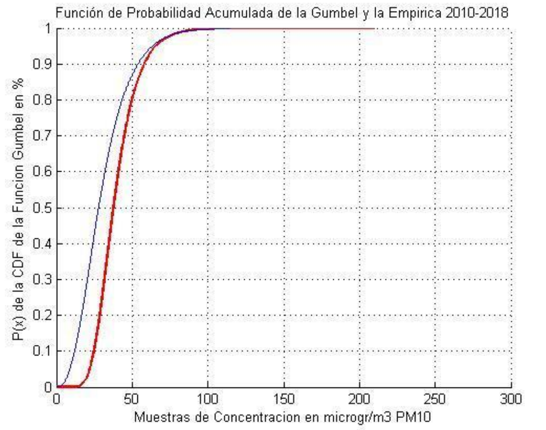
Table 4. PDF Adjusted Trend 2010-2018

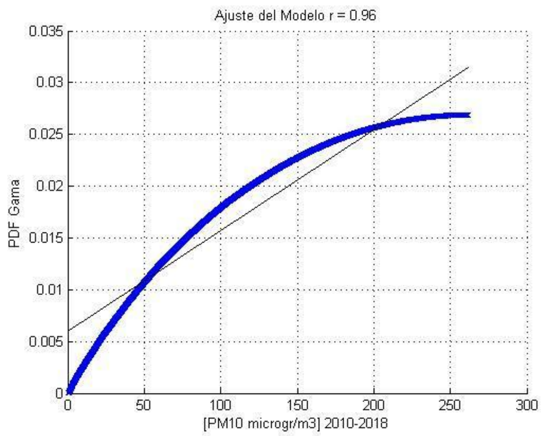


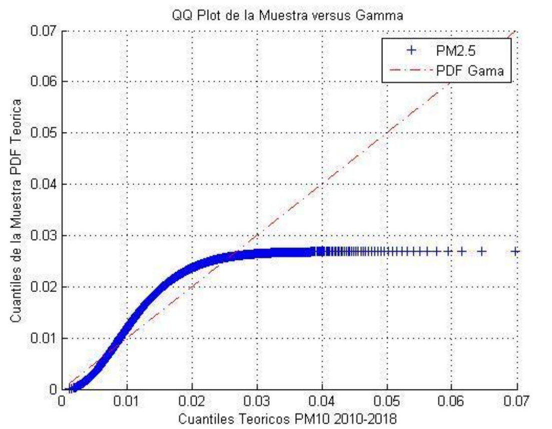

Función de Distribucion de Weibull
Probabilidad P(x)
Muestras de PM10 Diarios en microgr/m3 del año 2010-2018


Función de Probabilidad Acumulada de la Weibull y la Empirica 2010-2018
P(x) de la CDF de la Funcion Weibull en %
Muestras de Concentracion en microgr/m3 PM10

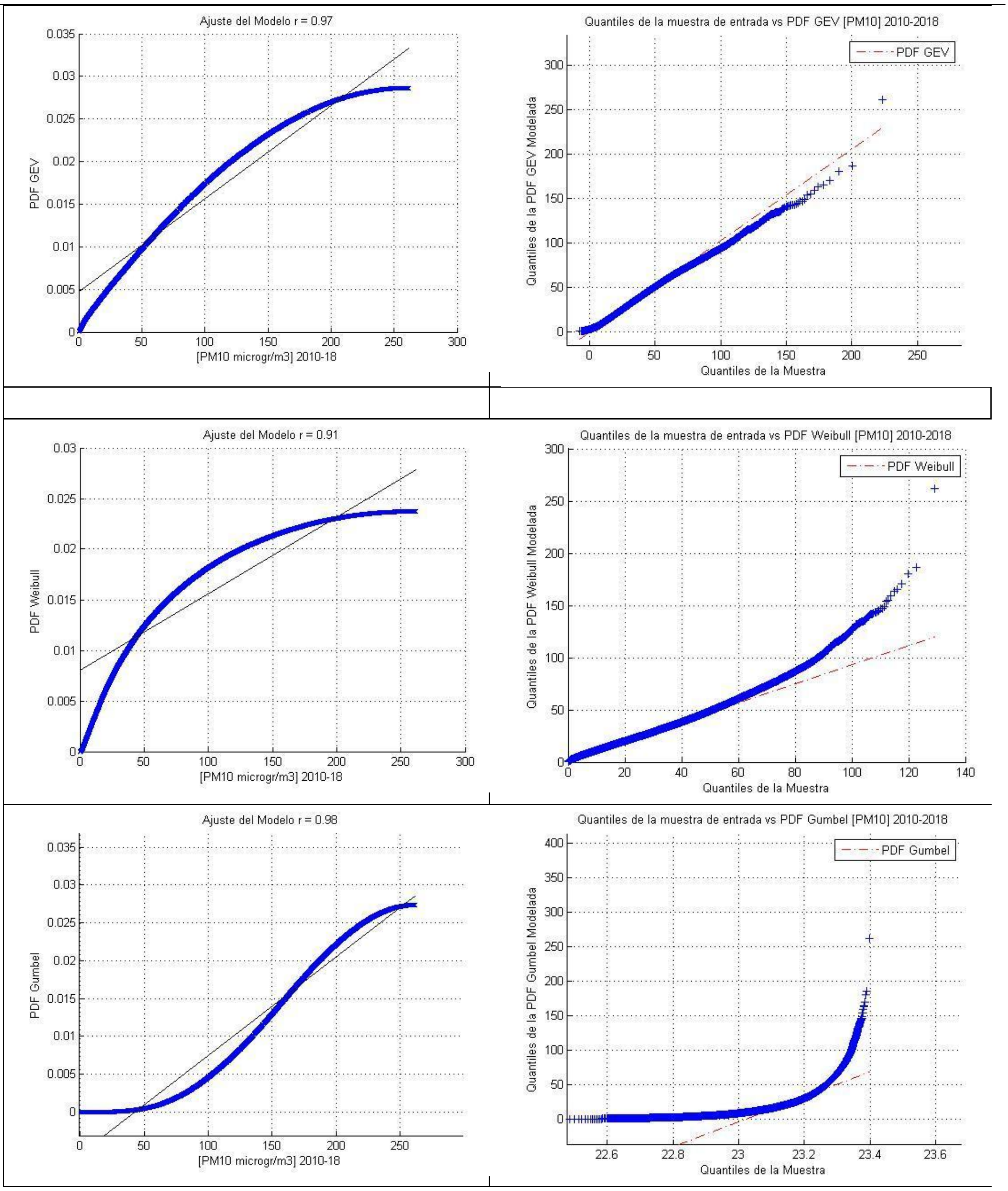

Concentraciones de PM10 2010-2018: Mean=31 Std=17
Histograma & Modelo Gumbel
Concentraciones microgr/m3


Función de Probabilidad Acumulada de la Gumbel y la Empirica 2010-2018
P(x) de la CDF de la Funcion Gumbel en %
Muestras de Concentracion en microgr/m3 PM10


Ajuste del Modelo r = 0.96
PDF Gama
[PM10 microgr/m3] 2010-2018


QQ Plot de la Muestra versus Gamma
PM2.5
PDF Gama
Cuantiles de la Muestra PDF Teorica
Cuantiles Teoricos PM10 2010-2018

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 05 - Issue 04 || April 2019 || PP. 18-36



The best pdf modeling the concentration of PM10 are the GEV, Gama and Weibull, most closely the first two pdf and pdf Weibull like third option, of trend 2010-2018

Table 5. Concentrations 2010-2018 of PM10

Year	GEV Mean of PM10 µgr $/m^3$
2010	45.2440
2011	44.6359
2012	42.0050
2013	43.2191
2014	36.4488
2015	32.8680
2016	34.9588
2017	33.9663
2018	29.9577

Year	Weibull Mean of PM10 µgr $/m^3$
2010	45.3768
2011	44.6995
2012	42.1212
2013	43.3425
2014	36.5970
2015	32.9692
2016	35.0874
2017	34.0939
2018	30.0508

Year	Gama Mean of PM10 µgr $/m^3$
2010	45.2532
2011	44.6331
2012	42.0153
2013	43.2324
2014	36.4795
2015	32.8798
2016	34.9660
2017	34.0331
2018	29.9632

Mean Weibull 31.05 µgr $/m^3$
Mean GEV 30.95 µgr $/m^3$
Mean Gama 30.96 µgr $/m^3$

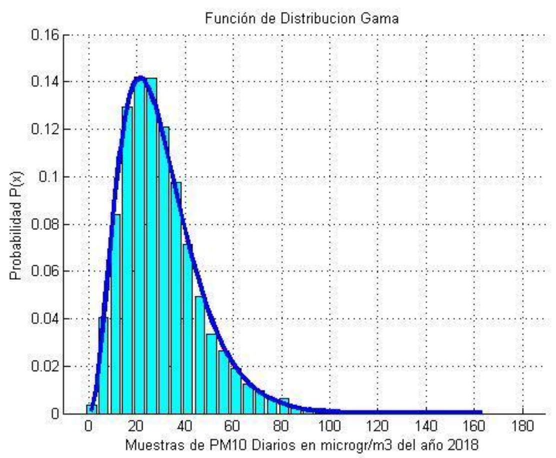
Table 6. Adjustment of the Graphics 2018

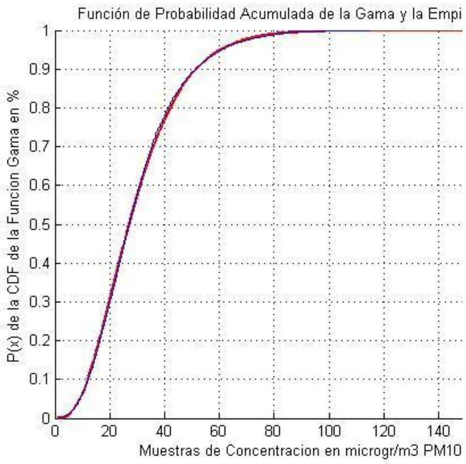
Gama pdf

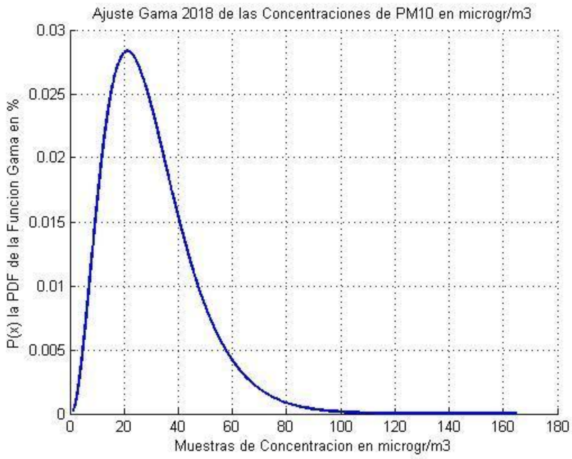
2018

Alpha = 3.4395
Beta = 8.7114

RMSE = 0.3792
MSE = 0.1438
IA = 0.7549
R2 = 0.8818



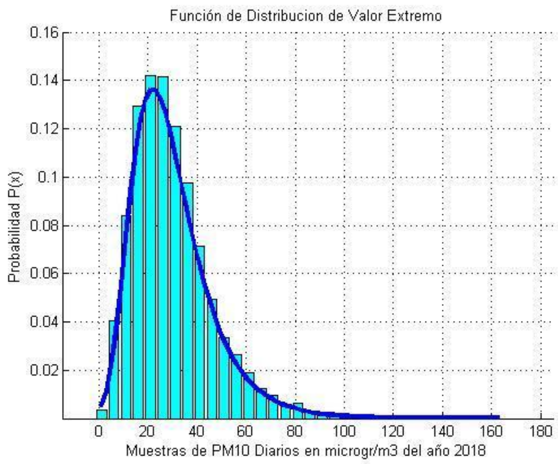


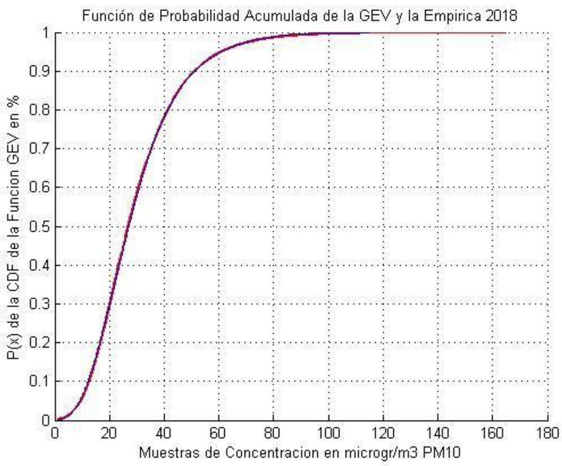


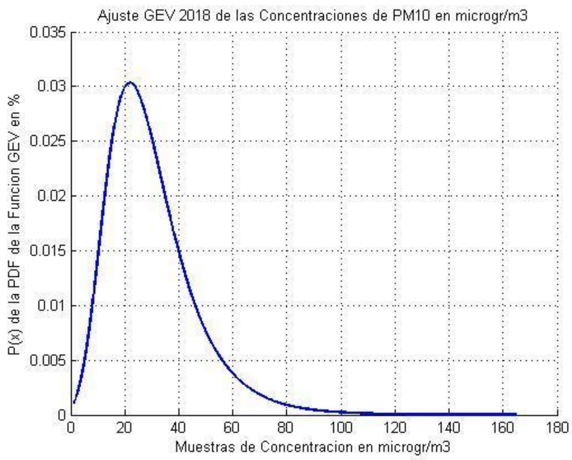
GEV 2018

K = 0.0431
Sigma = 12.1313
Mu = 22.4173

RMSE = 0.3786
MSE = 0.1433
IA= 0.7554
R2 = 0.8820







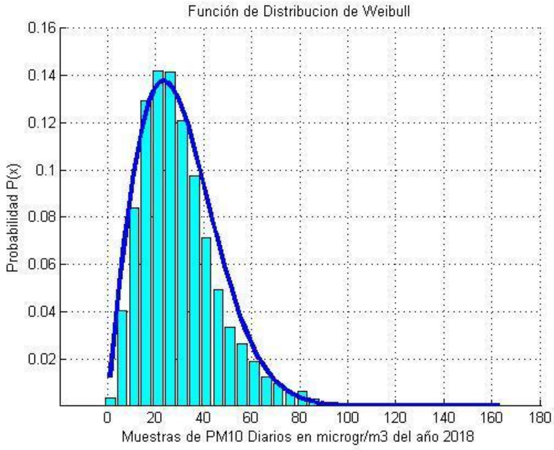
International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 05 - Issue 04 || April 2019 || PP. 18-36

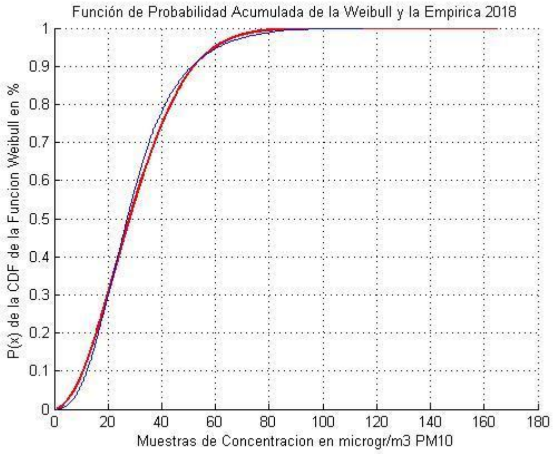
Weibull

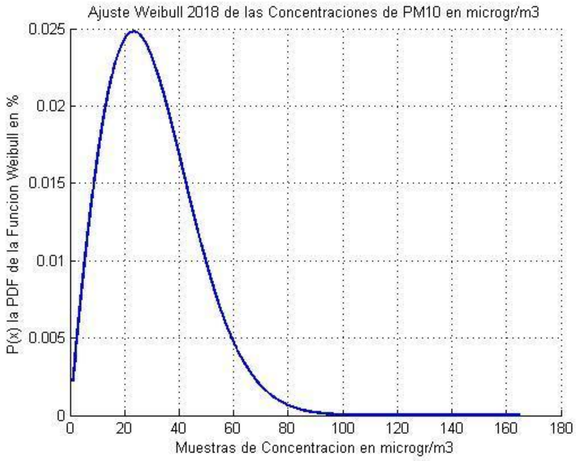
2018

Alpha = 33.8854
Beta = 1.9391

RMSE = 0.3884
MSE = 0.1509
IA = 0.7450
R2 = 0.8683







In Table 5 we can see the trend is decreasing and comparison with the data obtained from the official website of the México City is made. Now an analysis is done to observe the trend within the period using the method of obtaining new functions Distribution Probability Normal and Extreme Value for Bayesian Inference Data of Maximum Daily, which have a type behavior Gaussian biased, see in [20] then we can also observe whether a function of the type GEV one more or less variance than the other anger concentration obtained by decreasing or increasing.

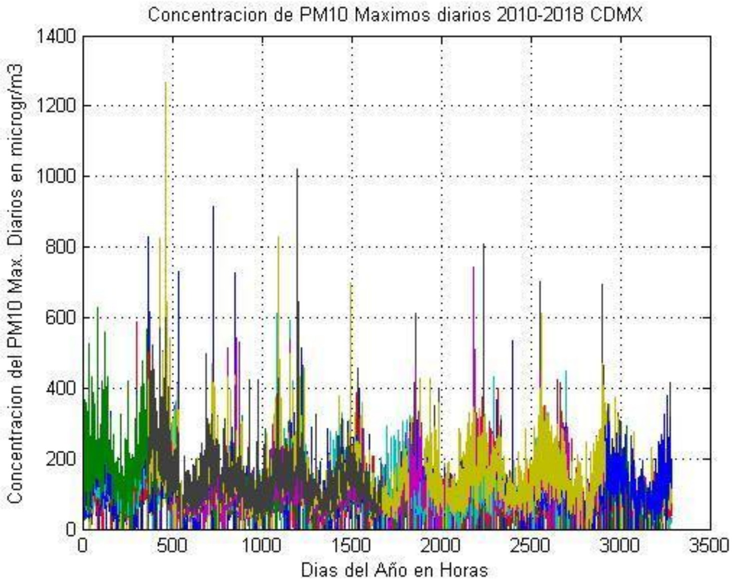


Figure 3. Concentrations of PM10 Maximum daily Mexico City 2010-2018

Statistics Maximun Daily of PM10
min: 2708**µ**gr /m^3
max: 412.29**µ**gr /m^3
mean: 74.90**µ**gr /m^3
median: 67.91**µ**gr /m^3
std: 40.91**µ**gr /m^3

The News GEV or GEV One we find them with the following expressions:

$GEV(\sum_{i=1}^{n}\frac{\mu_i}{n},\frac{1}{n-1}\sum_{i=1}^{n-1}\sigma_i\ ,k)$	(1)
With $k > 0\ x \in \left[\mu - \frac{\sigma}{k}, +\infty\right]$	$k < 0\ x \in \left[-\infty, \mu - \frac{\sigma}{k}\right]$

New GEV

$k = \left(\frac{GEVk + GEVkA}{\sum_{i=1}^{2} pn}\right)$	(2)
$Sigma = \left(\frac{GEVsd + PostSD}{\sum_{i=1}^{2} pn}\right)$	**(3)**
$Mu = \left(\frac{GEVmu + Postmean}{\sum_{i=1}^{2} pn}\right)$	**(4)**

This expression was best worked, approaching the input distribution function by Bayesian Inference are looking for values above the official standard annual average concentration. We obtain the following results:

Table 6. Adjust Trends of Daily Maximum PM10 New GEV

PDF	Estimators
PDF GEV New 2010 k = -0.2360 Sigma = 17272 Mu = 40.37	MSE = 0.00090098 RMSE = 0.0311 AP = 0.7038 R2 = 0.94 AI = 0.9928
PDF GEV New 2011 k = -0.2309 Sigma = 14.55 Mu = 33.64	MSE = 0.00086797 RMSE = 0.0298 AP = 0.7327 R2 = 0.88 AI = 0.9938
PDF GEV New 2012 k = -0.2165 Sigma = 14.1353 Mu = 35.38	MSE = 0.0015 RMSE = 0.0400 AP = 0.7650 R2 = 0.89 AI = 0.9877
PDF GEV New 2013 k = -0.2191 Sigma = 17.4188 Mu = 38,721	MSE = 0.0012 RMSE = 0.0358 AP = 0.99 R2 = 0.89 AI = 0.9814
PDF GEV New 2014 k = -0.2662 Sigma = 15.1624 Mu = 39.0227	MSE = 0.00085351 RMSE = 0.0296 AP = 0.8112 R2 = 0.99 AI = 0.9903
New PDF GEV 2015 k = -0.2511 Sigma = 12.81 Mu = 34.54	MSE = 4.4678e-04 RMSE = 0.0214 AP = 0.8621 R2 = 0.9001 IA = 0.9955
New PDF GEV 2016 k = -0.2254 Sigma = 14.3886 Mu = 35.05	RMSE = 0.00091391 MSE = 0.0306 AP = 0.9728 R2 = 0.9600 IA = 0.9905
New PDF GEV 2017 k = -0.2913 Sigma = 15.84 Mu = 38.49	RMSE = 0.0393 MSE = 0.0015 AP = 0.9989 R2 = 0.9763 IA = 0.9978
PDF GEV New 2018 k = -0.2186 Sigma = 12.22 Mu = 31.3285	RMSE = 0.0249 MSE = 0.0005953 AP = 0.9820 R2 = 0.9838 IA = 0.9990

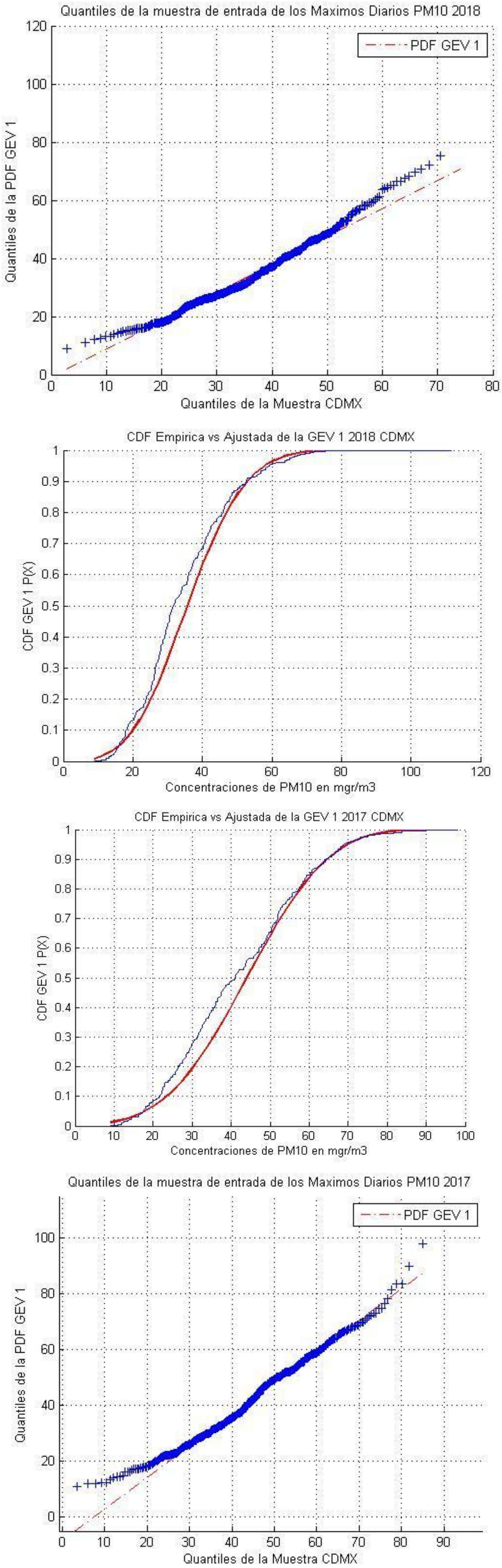
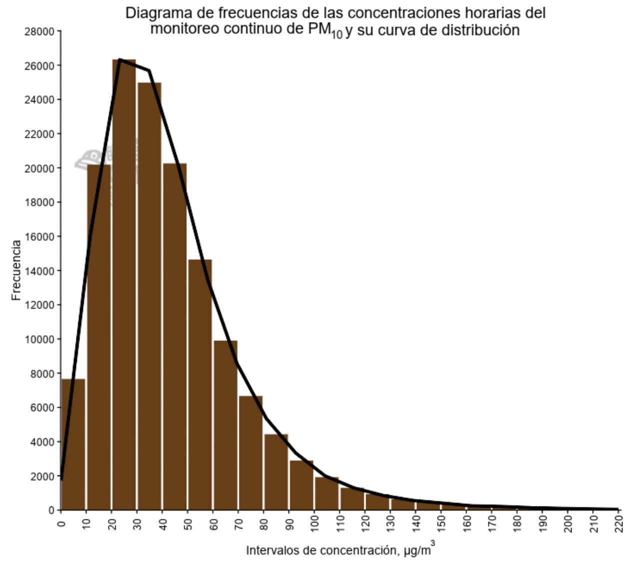
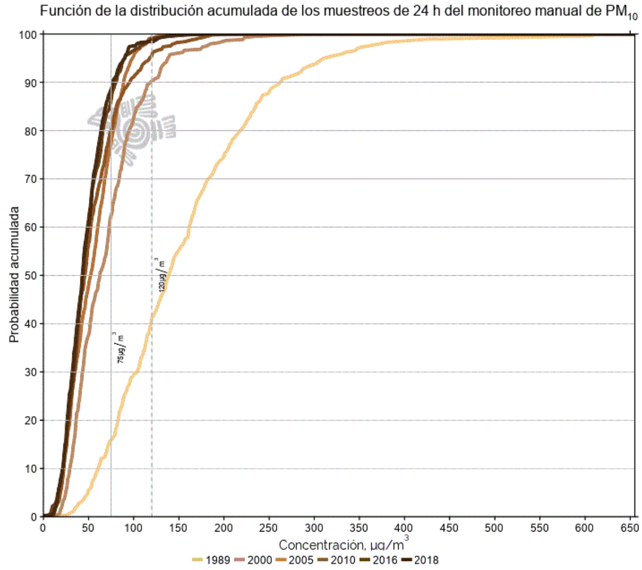
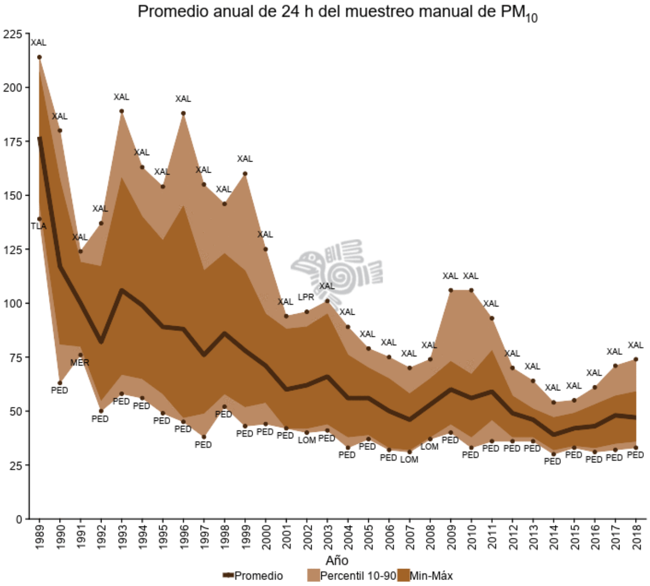


Figure 4. Concentrations of PM10 Daily Mexico City 2017 and 2018 the New GEV

Graph of the Official Website of México City







Trend μgr $/m^3$	Average CDMX Air web	New GEV maximums daily
2010	55	48
2011	60	40
2012	53	41
2013	50	47
2014	45	45
2015	48	40
2016	48	41
2017	53	44
2018	51	36

Figure 5. PM10 Maximum Daily Mexico City (Source: http://www.aire.cdmx.gob.mx/)

We are looking the trend of maximum daily of PM10, and we can observe that approximate the graph officer of mexico city and both can see that the concentration decreases slowly.

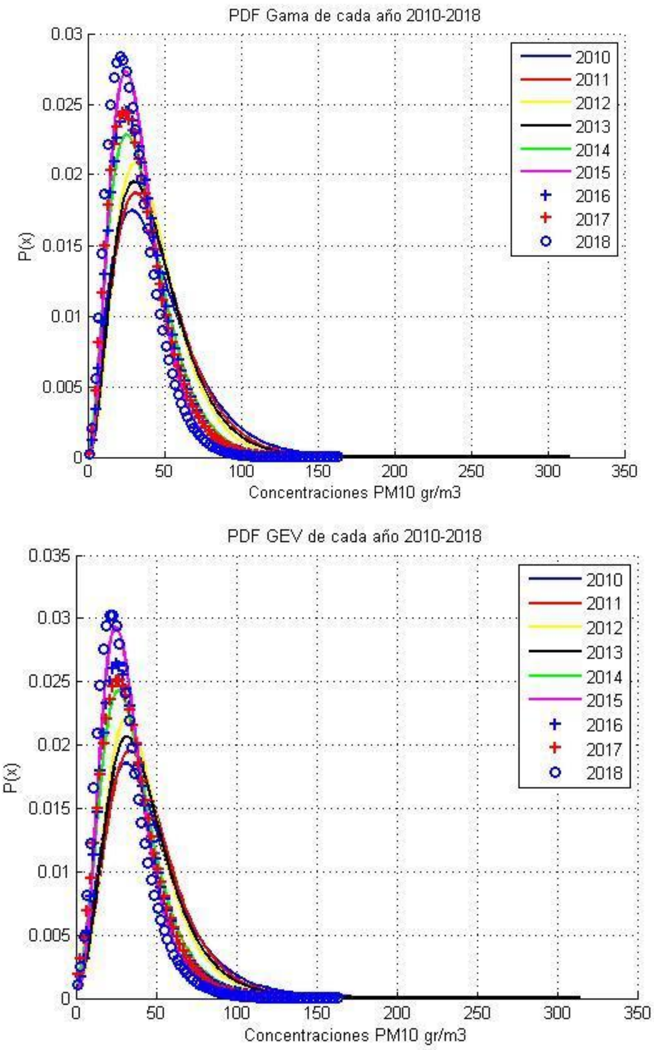


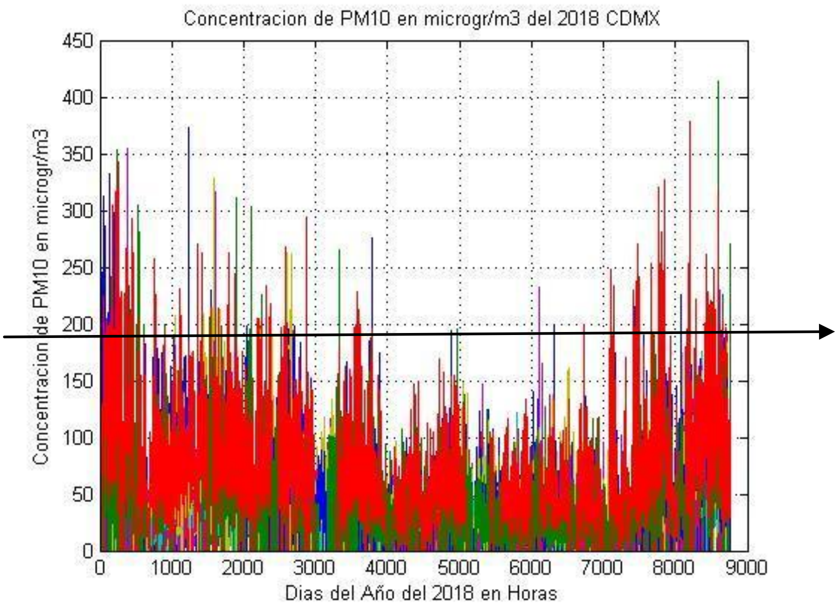
Figure 6. Concentrations of PM10 Daily Mexico City both pdf GEV and Gama

The result was compared against the Bayesian model for the average and standard deviation unknown taking as forecast model pdf Inverse Gamma [21] to observe the new Max Daily Means of PM10 using Gibbs sampling, with very excellent results , corroborating the approach of Method [20] for the News GEV were found for these nearly Gaussian data.

Trend µgr $/m^3$	**Average CDMX Air Web**	**New GEV Maximum Daily**	**IGamma Forecasted averages**	**IGamma Parameters**
2010	55	48	48,02	Alpha = 2.03 Beta = 49.57
2011	60	40	40.10	Alpha = 2.03 Beta = 41.35
2012	53	41	40.93	Alpha = 2.03 Beta = 42.31
2013	50	47	53.68	Alpha = 2.03 Beta = 55.39
2014	45	45	44.35	Alpha = 2.03 Beta = 46.08
2015	48	40	40.35	Alpha = 2.04 Beta = 42.09
2016	48	41	42.45	Alpha = 2.03 Beta = 43.95
2017	53	44	42.12	Alpha = 2.03 Beta = 43.61
2018	51	36	34.60	Alpha = 2.03 Beta = 35.83

Figure 7. Maximum Daily of PM10 of Mexico City Comparative Measures of the News GEV and the predicted averages

Now just we calculate the days of exceedances for 2018 as an example, having the maximum concentration of the Daily Maximum Concentrations and PM10, we have the following;



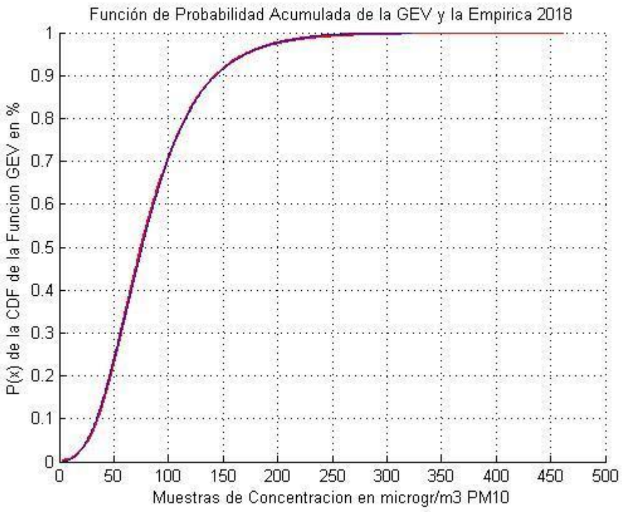


Figure 7. 2018 Daily PM10 Concentrations in Mexico City and CDF

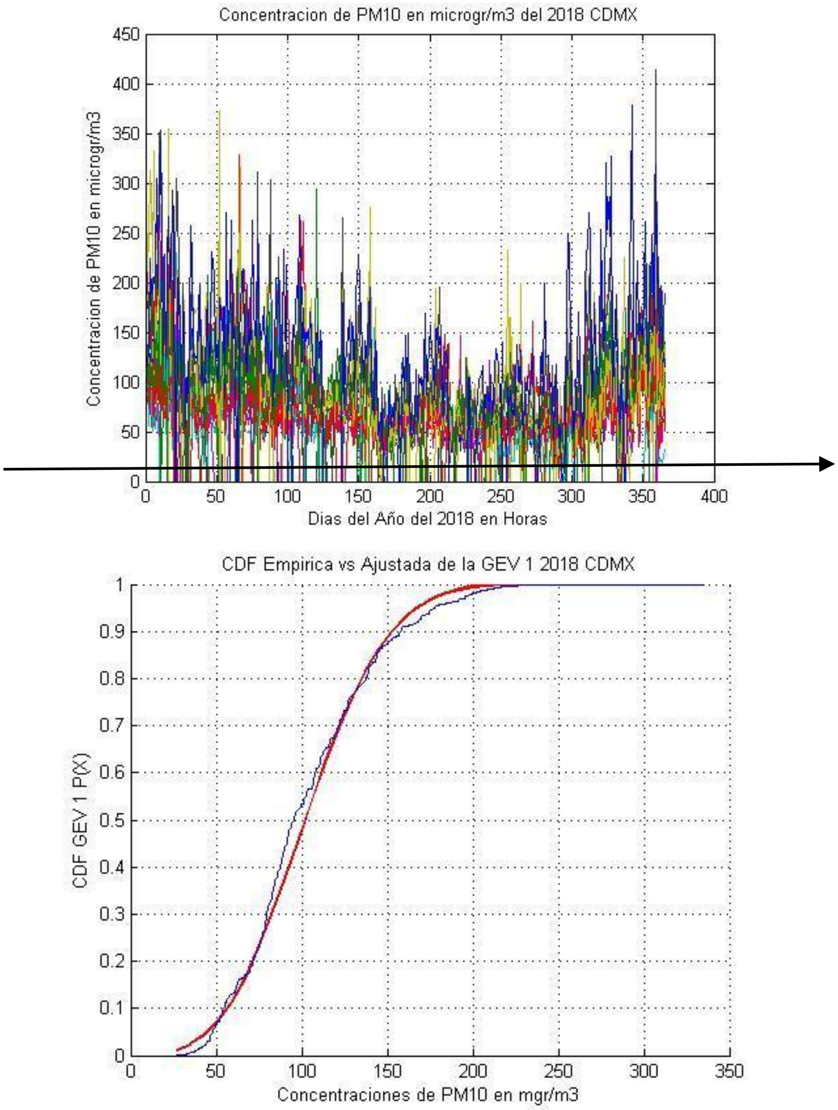


Figure 8. PM10 Concentrations Maximum Daily CDF 2018 and Mexico City


Consultas

259 días cumplen con el criterio.

Índice de calidad del aire mayores a 75					
Fecha	NOPM10	NEPM10	CEPM10	SOPM10	SEPM10
2018-01-01	115	112	105	101	113
2018-01-02	103	101	96	90	110
2018-01-03	110	119	97	70	87
2018-01-04	106	117	103	86	93
2018-01-05	110	122	105	93	96
2018-01-06	107	117	102	99	93
2018-01-07	101	111	77	71	73
2018-01-08	113	125	102	73	101
2018-01-09	105	119	90	73	91
2018-01-10	120	138	103	70	105

Figure 9. Consultation of PM10 Exceedance above 75 µgr $/m^3$ of 2018 México City (Source:http://www.aire.cdmx.gob.mx/)

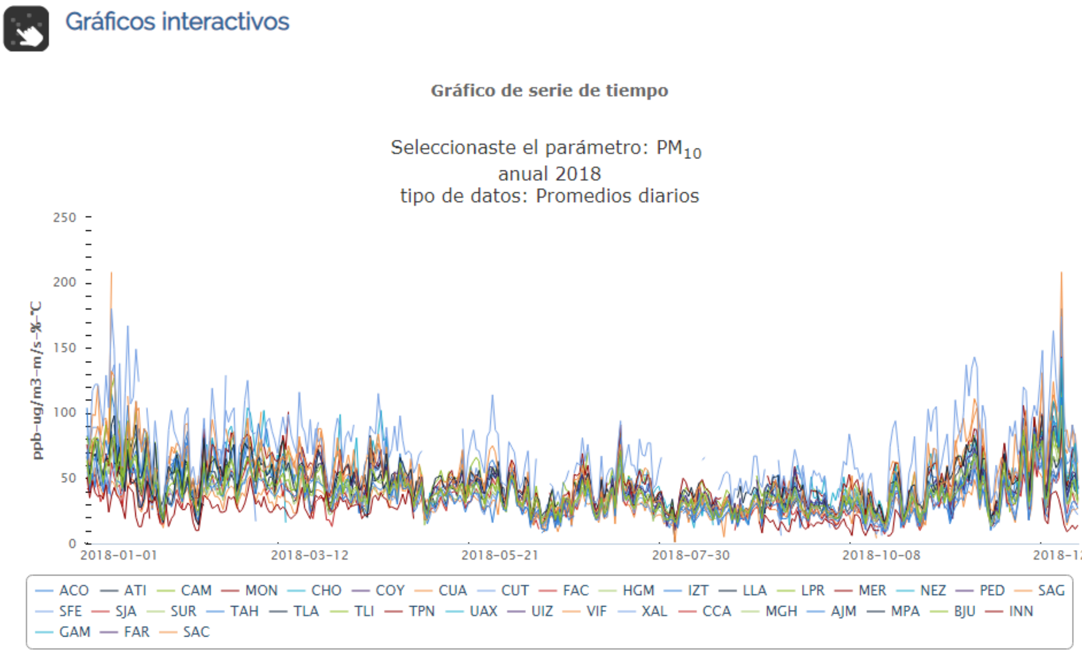


Figure 10. Consultation Daily averages of PM10 2018 Mexico City (Source: http://www.aire.cdmx.gob.mx/)

Calculate the CDF New GEV to maximum concentration is approximately (1-0.4500) * 365 = 200.8 days excendencia which takes to the true value

Now with daily highs always have one or more the following as we are looking above the norm of 75, it is considered a maximum, The New GEV give us (1-0.2700) * 365 = 266.45 days, which is very good and the value is exceeded. See also the daily averages of Fig 10, the average without a maximum concentration is 120**µ**gr $/m^3$ as at the beginning of the study.

The New Air Quality Standard is also given to the City of Mexico, in the following link: NADP-009-AIR-2017 (http://www.aire.cdmx.gob.mx/default.php?ref=Z2Q=)

Conclusions

With this study it was found that the probability distribution function was the most adequate for the behavior of daily data of PM10 which were with the best adjustment the pdf GEV and the pdf Gama, is comparative with the adjustment given by the official page of the Mexico City, these distribution functions are part of the Extreme Value Theory.

With the trend analysis we used the methodology proposed in [20], for data with Gaussian behavior, which for maximum ozone data is perfectly coupled, in this case it was used for the maximum data of PM10 which also adjusted but biased, almost Gaussian data as the case of PM2.5 Particles, adjust the maximum data and give a good approximation of the parameters that we want to look for as the mean for the functions generated by the GEV giving us approximate results and comparisons with the graph of the page of the of Mexico City. We can also see in the QQ plot that the pdf New GEV or GEV 1 are adjusted to more extreme concentrations can be seen as the CDF is better adjusted in higher concentrations to observe the desired trend, could not adjust a second GEV as the case of Gaussian data given the nature of the data which allows only a New GEV function and for exceedance calculations with this data it is possible to obtain it. Now making an adjustment of the complete trend from 2010 to 2018 can be inferred in the possible concentration that can be obtained for this year, but that will continue to be studied later.

It was also found that the tendency of the PM10 concentration is above the average, we can observe it from the trend graph, the average is above and in others low, if we apply at maximum concentrations according to the observed data without the atypical data, but the tendency of the concentration in general goes downwards.

References

[1]. AJ Jakeman, JA Taylor, RW Simpson, Modeling of air pollutant distributions Concentrations - II. Estimation of one and two parameters statistical distributions, Atmos. Environ., 20 (1986) 2435-2447.

[2]. Berger, A., Melice JL and Demuth, CL (1982) of distributions Statistical daily and high atmospheric SO2 - Concentrations. Atmospheric Environment. 16 (5), 2863-2877

[3]. PM10 data base of Mexico City website ofhttp://www.aire.cdmx.gob.mx/

[4]. Georgopoulos, PG and Seinfeld, JH (1982) 'Statistical distribution of air pollutant concentration', Environmental Science & Technology, Vol. 16, pp.401A-416A.

[5]. Gumbel, EJ, 1958. Statistics of Extremes. Columbia University Press, New York, p. 164.

[6]. Kambezidis HD Tulleken, R., Amanatidis, GT, Paliatsos, AG and Asimakopoulos, DN (1995) 'Statistical evaluation of selected air pollutants in Athens, Greece', Environmetrics, Vol. 6, pp.349-361.

[7]. Kao, AS and Friedlander, SK (1995) Frequency distributions of PM10 chemical components and Their sources. Environmental Science and Technology. 29 (5), 19-28

[8]. Lu, H., Fang, G., 2003. Predicting the exceedances of a critical concentration PM10 - a case study in Taiwan. Atmospheric Environment 37, 3491-3499.

[9]. Morel, B., Yeh, S. and Cifuentes, L. (1999) 'Statistical distribution for air pollutants applied for the study of the problem in particulate Santiago', Atmospheric Environment, Vol. 33, pp.2575-2585.

[10]. PG Georgopoulous, JH Seinfeld, Statistical distributions of air pollutant Concentrations, Environ. Sci. Technol., 16 (1982) 401A-416A.

[11]. Roberts, MS, 1979. Review of statistics of extreme values with applications to air quality data, part II. Applications. Journal of Air Pollution Control Association 29, 733-740.

[12]. Samet, J., Domonici, F., Curriero, FC, Coursac, I. and Zeger, SL (2000) "Fine particulate air pollution and mortality in 20 US cities', The New England Journal of Medicine, Vol. 343, pp. 1742-1749.

[13]. Paper presented at the Congress of the Mexican Geophysical Union 2017 forecast of ozone concentrations by probability distributions for CDMX https://www.raugm.org.mx/2017/pdf/constancia.php?clave=809

[14]. Mijić Z., M. Tasić, S. Rajsic, V. Novakovic The statistical characters of PM10 in Belgrade area, Atmos. Res., 92 (2009) 420-426.

[15]. Berger, JO, Statistical Decision Theory and Bayesian Analysis, Springer Ser. Stat., 2nd ed., Springer-Verlag, New York, 1985.

[16]. Christopeit, N., Estimating parameters of an extreme value distribution by the method of moments, J. Stat. Plann. Inference, 41, 173-186, 1994.

[17]. Prescott, P., and Walden AT, Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples, J. Stat. Comput. Simul., 6, 241-250, 1983.

[18]. Otten, A., and MAJ Van Montfort, Maximum-likelihood estimation of overall the extreme-value distribution parameters, J. Hydrol., 47, 187-192, 1980.

[19]. Jose Jimenez Roberto Zenteno, Prediction of Concentrations of Ozone Levels in Mexico City using Probability Distribution Functions, International Journal of Latest Research in Engineering and Technology (IJLRET) || Volume 04 - Issue 07 || July 2018 || PP. 35-45

[20]. Jose Jimenez Roberto Zenteno. A Methodology for Obtaining news Probability Distributions Functions Normal and Extreme Value for Bayesian Inference and Stochastic Gaussian Mixed Case One: For Daily Maximum Ozone Concentration Data. International Journal of Research in Latest Engineering and Technology (IJLRET) || Volume 04 - Issue 11 || November 2018 || PP. 15-35

[21]. Statistical Modeling and Computation, Dirk P. Kroese - Joshua Chan CC, Springer Ed 2014, Bayesian Inference Chapter 8, page 236.