International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 07 - Issue 06 || June 2021 || PP. 07-15

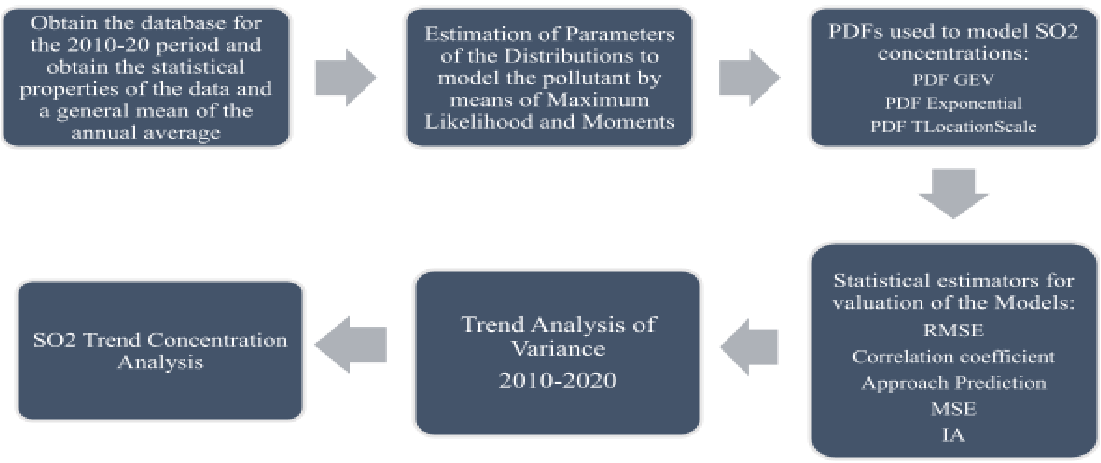
Analysis of Sulfur Dioxide Concentrations in México City, trend 2010 - 2020

M. Sc. Zenteno Jiménez José Roberto
Geophysical Engineering, National Polytechnic Institute, México City, ESIA-Ticóman Unit Mayor Gustavo A. Madero
jzenteno@ipn.mx

Abstract: The study comprises an analysis of data from 2010 to 2020, it was proposed to obtain the best or best probability distribution functions that model SO2 concentrations in México City, using the following pdf, T location scale distribution function, extreme value distribution function and exponential distribution function, to obtain the estimators the method of maximum likelihood and moments was used and aided by the Matlab program, for valuation of the forecast model, RMSE, MSE, coefficient of determination, approximation of prediction and approximation index, in turn an analysis is made to observe its trend with an analysis of variance, the daily concentration data is downloaded from the official monitoring page and corroborating with the official air page of México City.

Keywords: Sulfur Dioxide, Probability Distributions, Adjustment Indicators, Analysis of Variance.

Probability Distribution Functions and Methodology



Four probability distribution functions were used, which are the GEV distribution function, the exponential distribution function and the tlocation scale distribution function.

Table 1 Probability Distribution Functions and their Parameters.

Distribution	Probability density function	Parameters
GEV	$f(x) = \left(\frac{1}{\sigma}\right) exp^{-\left((1+kz)^{\left(-\frac{1}{k}\right)}\right)(1+kz)^{\left(-1-\frac{1}{k}\right)}}$	***K* shape** **σ scale** **μ location**
TLocation Scale	$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sigma\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(\frac{v+\left(\frac{x-\mu}{\sigma}\right)^2}{v}\right)^{-\frac{(v+1)}{2}}$	**σ scale** **μ location** **v shape**

Exponential	$f(x) = \left(\frac{1}{\mu}\right) exp^{-\left(\frac{x}{\mu}\right)}$	**µ mean**

Statistical Fit Estimators

The deviation indicators of a group of data in relation to a model can be used to assess the goodness of fit between the two. Among the most common indicators are the following: RMSE, MAE, NRMSE, CV-MRSE, SDR, and R ^ 2. Those that were used to determine the distribution that best fit the data. They are the mean square error (RMSE), mean square error (MSE), prediction precision (AP) and coefficient of determination (R ^ 2) Table2 gives the equations for the adjustment indicators that have been used by Lu (2003) and Junninen et al. (2002).

Table 2 Fit 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 **Coefficiente of Determination**	$R^2 = \left(\frac{\sum_{i=1}^{N}(Pi-P)(Oi-O)}{NS_pS_o}\right)^2$
Accuracy Measures **Prediction Accuracy**	$AP = \frac{\sum_{i=1}^{N}(Pi-O)^2}{\sum_{i=1}^{N}(Oi-O)^2}$
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, Pi = Predictive Values, Oi = Observed Values, P = Average of Predicted Values, O = Average of Observed Values, Sp = Standard Deviation of Predicted Values, So = Standard Deviation of Values Observed.

Study área

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

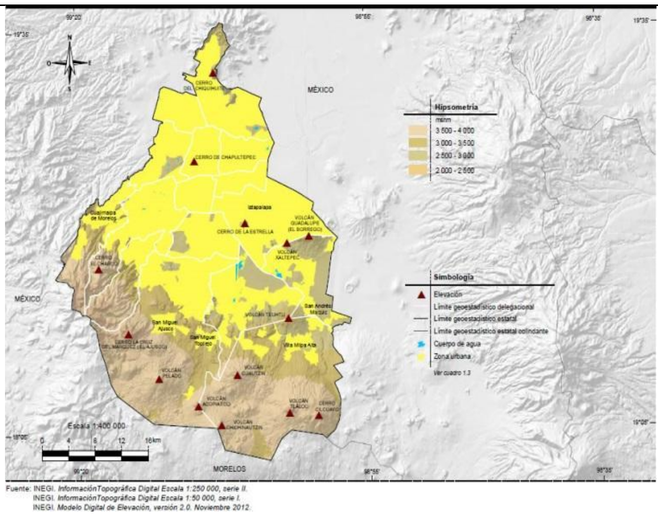


Figure 1. Relief of Mexico City (Source: https://www.paratodomexico.com/)

Statistical Description of the Data

In the following table we can see the characteristics of the database which shows 5% of null or unread values.

Table 3 Statistical Description of the Average Concentration of SO2 Data Trend 2010-2020

Data Number	105408
Minimum	0.1 ppb
Maximum	40 ppb
Mean	3.2 ppb
Variance	18.10 ppb
Standard deviation	4.25 ppb
Median	1.66 ppb

Results

Table 4 Estimation Parameters and Trend Adjustment Indicators 2010-2020 in ppb

Distributión	Parameters	Mean and Variance	RMSE	MSE	R^2	IA
GEV	k 0.1139 sigma 0.1822 mu 0.2931	Mean=0.4235 Var=0.0806	0.3819	0.1458	0.8768	0.7487
TLocation	mu= 0.36427 sigma= 0.17782 nu= 2.981	Mean= 0.3643 Var= 0.0960	0.3708	0.1374	0.8490	0.7583
Exponential	Mu=0.4236	Mean= 0.4236 Var= 0.1794	0.3094	0.0957	0.9002	0.8047

Fit Testing

Kolmogorov– Smirnov	Chi Test
Dmax= 0.5926 KSSTAT = 0.0176	h = 0 p = 0.8027
Dmax= 0.5941	h = 0 p = 0.8200
Dmax= 0.4412 KSSTAT = 0.2033	h = 0 p = 0.3900

SO2 Trend Time Series 2010-2020

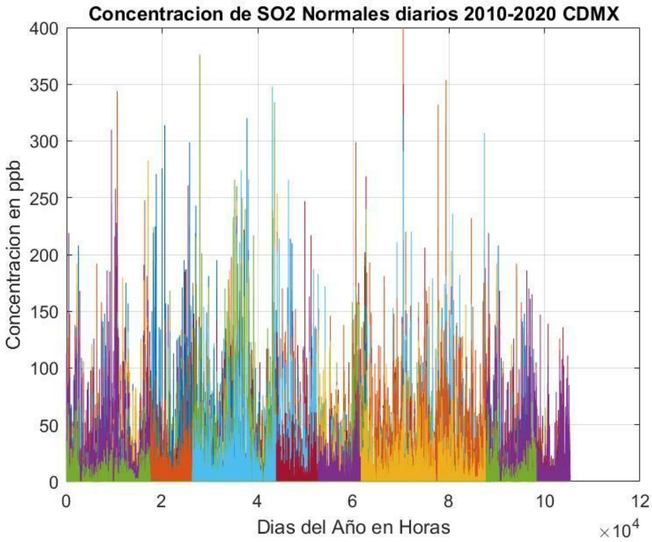


Figure 2 Time Series of SO2 Concentrations of México City, the Matrix

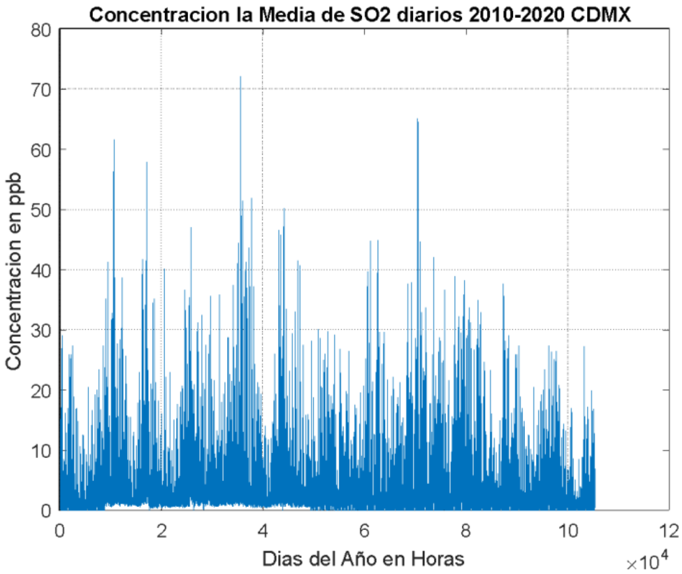
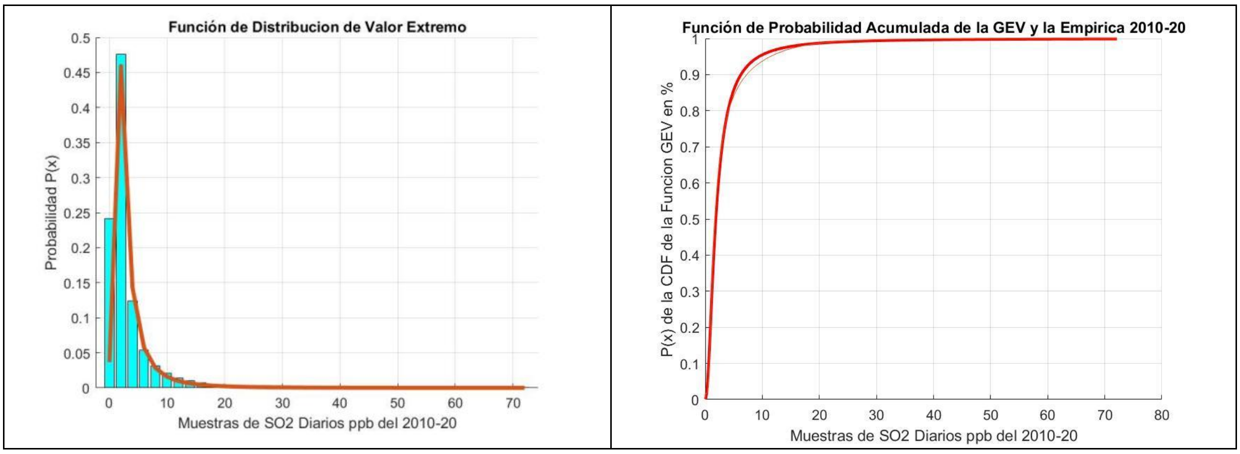
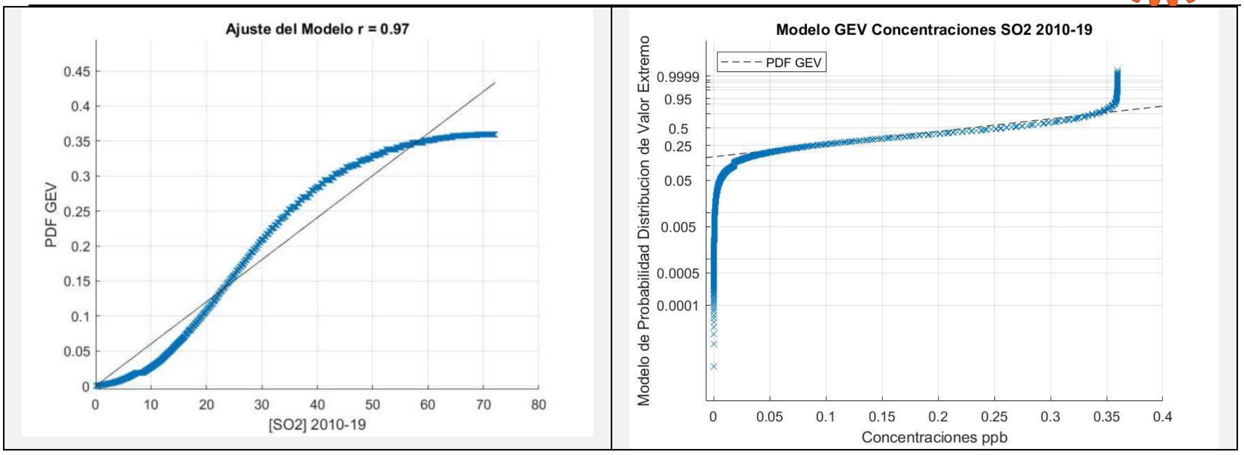


Figure 3 Time Series of the SO2 Mean of México City.

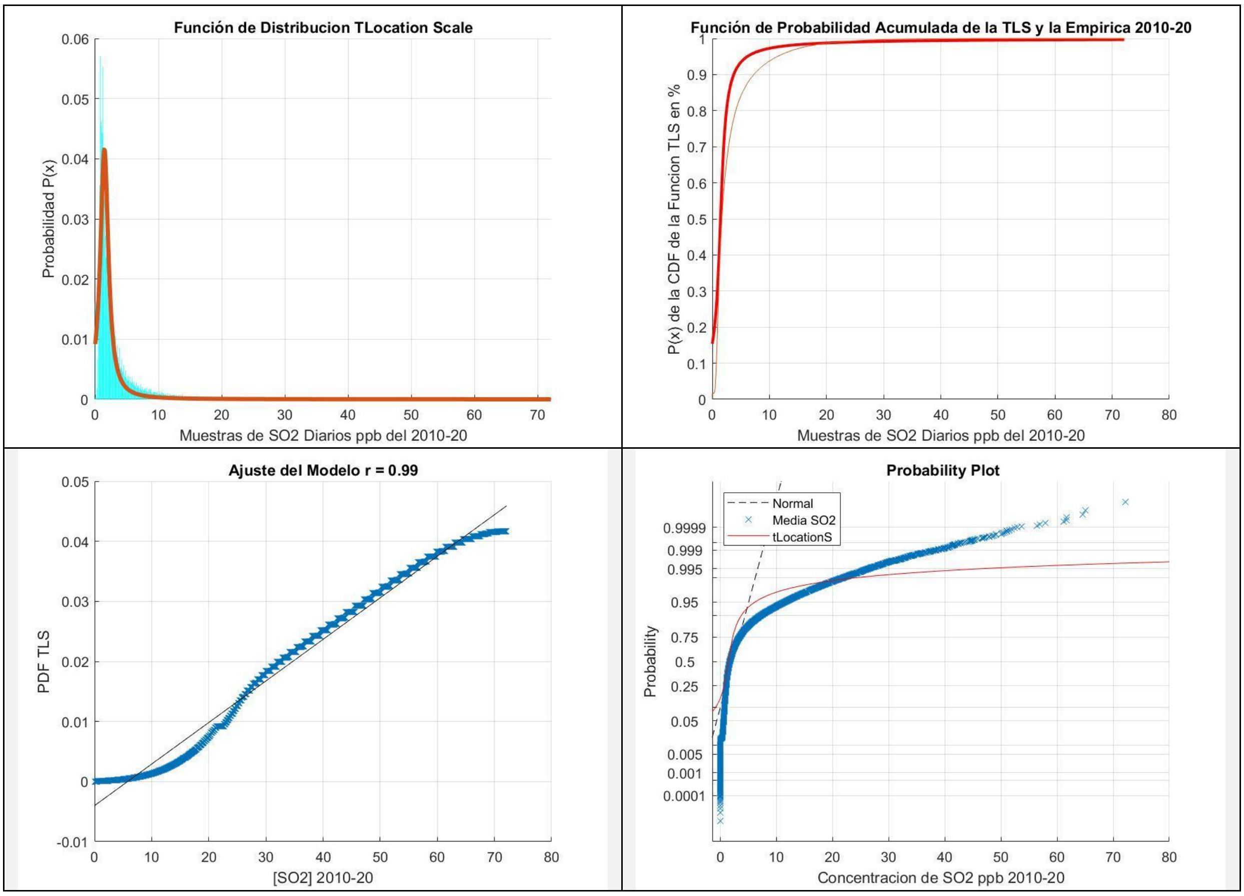
Analysis by probability distribution function

GEV

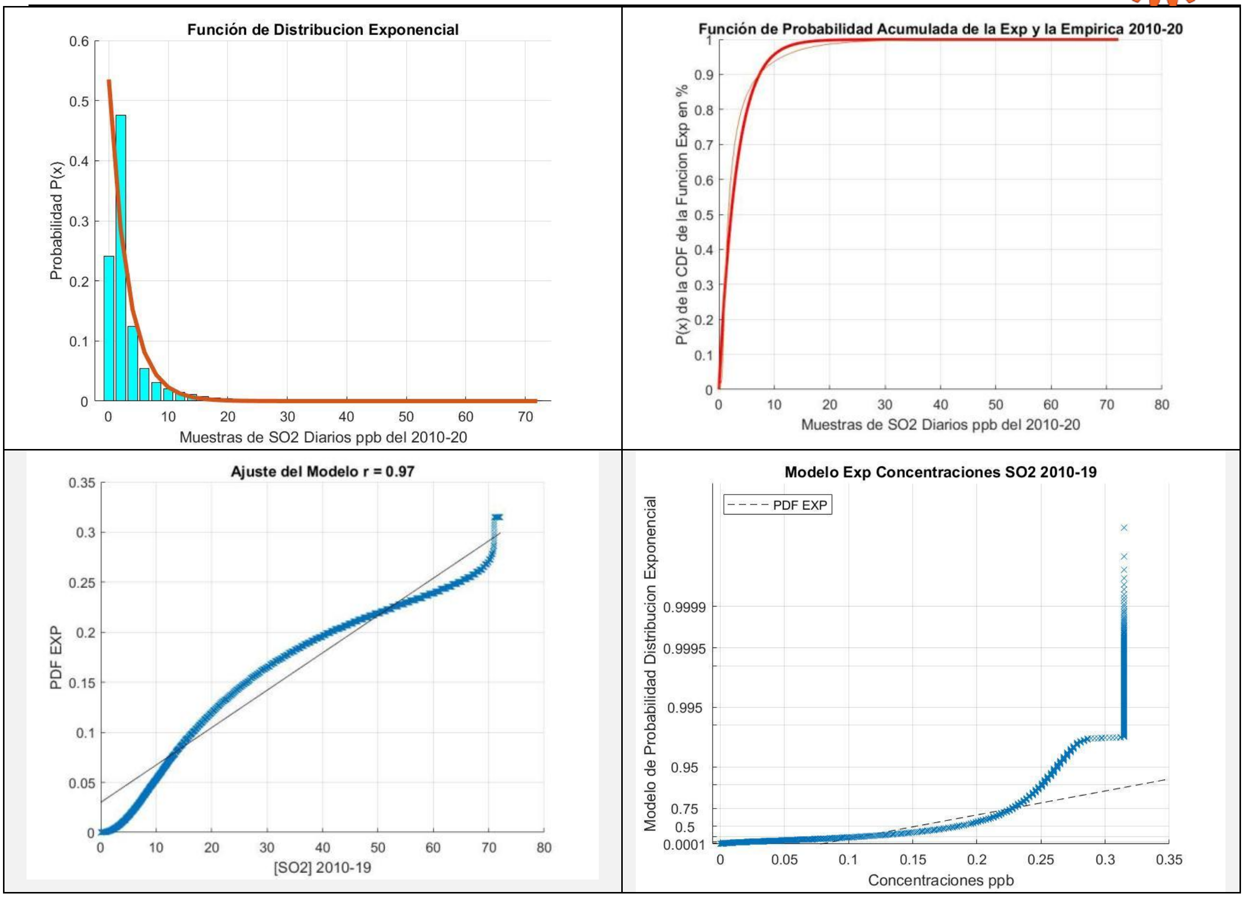




TLS



Exponential



Analysis of Variance of the SO2 Trend 2010-2020

An analysis of variance (ANOVA) tests the hypothesis that the means of two or more populations are equal. ANOVAs assess the importance of one or more factors by comparing the means of the response variable at different levels of the factors. The null hypothesis states that all the population means (means of the factor levels) are equal while the alternative hypothesis states that at least one is different.

To run an ANOVA, you must have a continuous response variable and at least one categorical factor with two or more levels. ANOVA analyzes require population data that follow an approximately normal distribution with equal variances between factor levels. However, ANOVA procedures work quite well even when the assumption of normality is violated, unless one or more of the distributions are highly skewed or the variances are quite different. Transformations of the original dataset can correct these violations.

The name "analysis of variance" is based on the approach in which the procedure uses the variances to determine if the means are different. The procedure works by comparing the variance between the group means and the within-group variance as a way to determine whether the groups are all part of a larger population or separate populations with different characteristics.

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	71019.8	10	7101.98	275.52	0
Error	2490319.7	96613	25.78		
Total	2561339.5	96623			

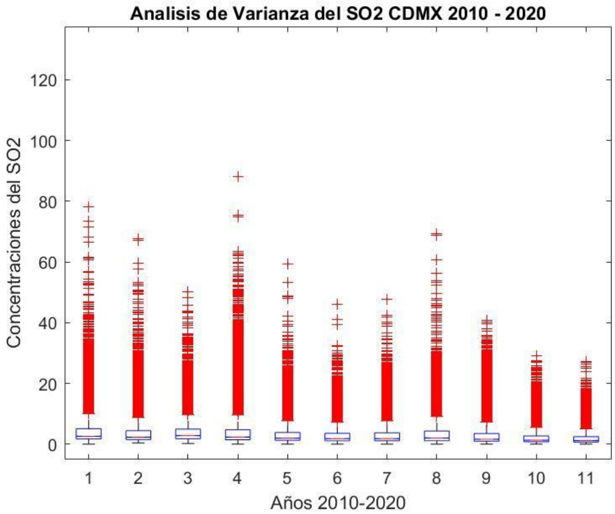


Figure 4 Analysis of Variance of the Mean of SO2 of Mexico City, where (Mean 1 = 2010 to Mean 11 = 2020)

Now with the year 2021

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	139659.4	11	12696.3	528.43	0
Error	2532286.6	105396	24		
Total	2671945.9	105407			

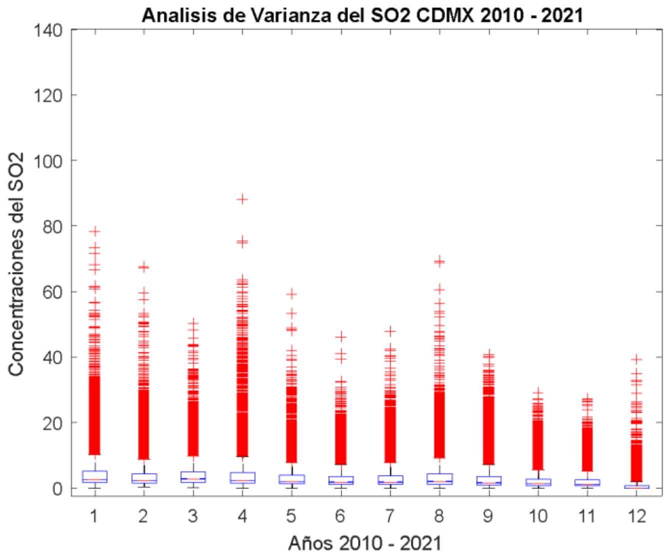
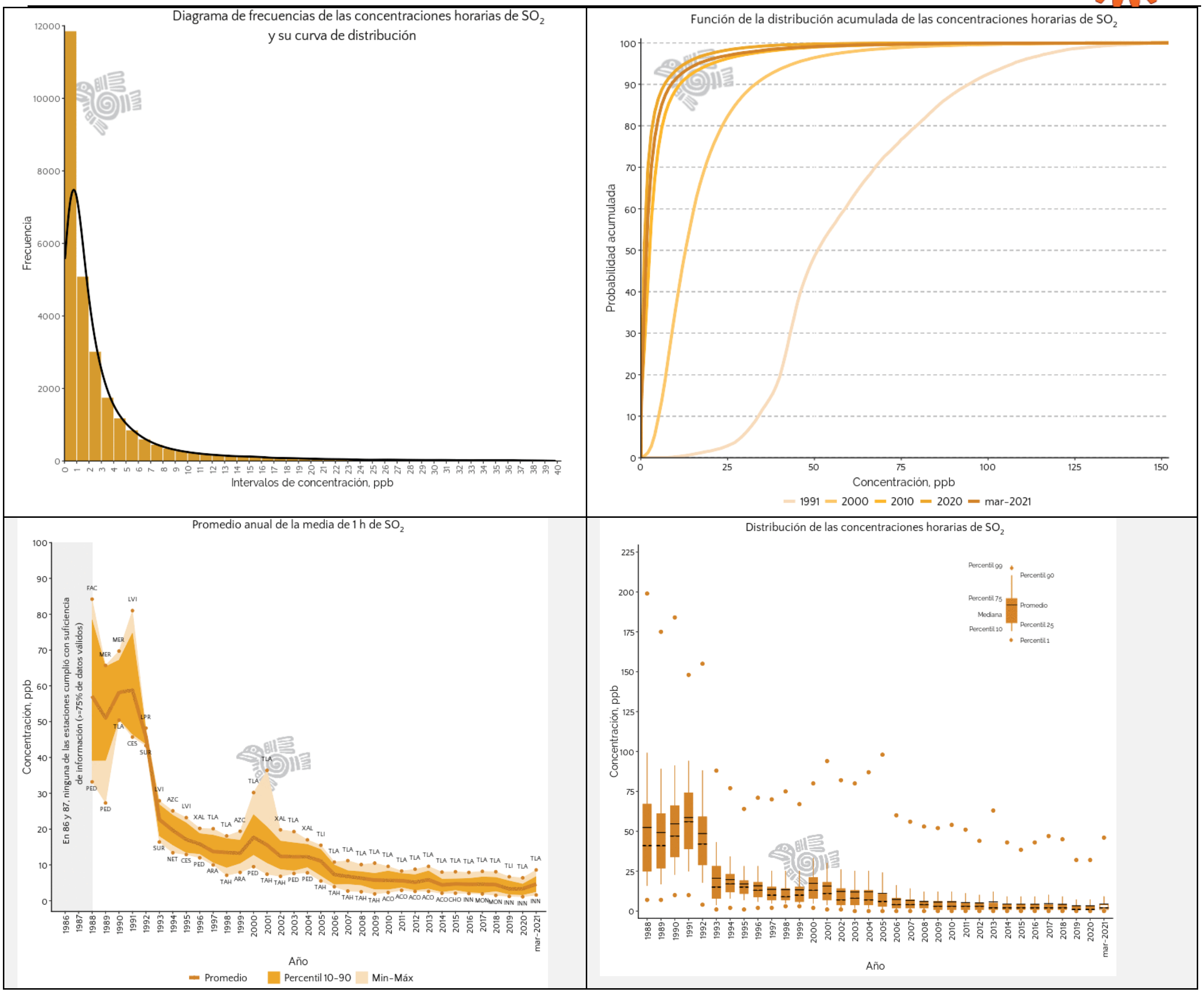


Figure 5 Analysis of Variance of the Mean of SO2 of Mexico City, where (Mean 1 = 2010 to Mean 12 = 2021)

Comparing with the Trend Charts of the Official Website of Mexico City



Conclusions

With this study it was verified which type of probability distribution function was the most adequate for the behavior of daily SO2 data, the GEV pdf and the Exponential pdf were with the best fit, in comparison with the adjustment given by the official website of the Mexico City.

With the trend analysis, the Analysis of Variance was used to see the trend of the concentration of SO2 in Mexico City, which can be seen that the trend is downward although with a slight increase for this year of 2021, It would be necessary to see in more detail where that source comes from, whether internal or external to Mexico City.

References

[1]. A.J. Jakeman, J.A. Taylor, R.W. Simpson, Modeling distributions of air pollutant concentrations - II. Estimation of one and two parameters statistical distributions, Atmos. Environ., 20 (1986) 2435-2447.

[2]. Berger, A., Melice, J. L. and Demuth, C. L. (1982) Statistical distributions of daily and high atmospheric SO2 – concentrations. Atmospheric Environment. 16 (5), 2863 – 2877

[3]. Data base of PM2.5 website of México City http://www.aire.cdmx.gob.mx/

[4]. Georgopoulos, P.G. and Seinfeld, J.H. (1982) 'Statistical distribution of air pollutant concentration', Environmental Science Technology, Vol. 16, pp.401A–416A.

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

[6]. Kambezidis, H.D., Tulleken, R., Amanatidis, G.T., Paliatsos, A.G. and Asimakopoulos, D.N. (1995) 'Statistical evaluation of selected air pollutants in Athens, Greece', Environmetrics, Vol. 6, pp.349–361.

[7]. Kao, A. S. and Friedlander, S. K. (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 PM10 concentration – 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 particulate problem in Santiago', Atmospheric Environment, Vol. 33, pp.2575–2585.
[10]. P.G. Georgopoulous, J.H. Seinfeld, Statistical distributions of air pollutant concentrations, Environ. Sci. Technol., 16 (1982) 401A-416A.
[11]. Roberts, E.M.,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, F.C., Coursac, I. and Zeger, S.L. (2000) 'Fine particulate air pollution and mortality in 20 US cities', New England Journal of Medicine, Vol. 343, pp.1742–1749.
[13]. Trabajo presentado en el Congreso de la Unión Geofísica Mexicana 2017 Pronostico de Concentraciones de Ozono por Distribuciones de Probabilidad para la CDMX https://www.raugm.org.mx/2017/pdf/constancia.php?clave=809
[14]. Berger, J. O., Statistical Decision Theory and Bayesian Analysis, Springer Ser. Stat., 2nd ed., Springer-Verlag, New York, 1985.
[15]. Prescott, P., and A. T. Walden, 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.

[16]. Otten, A., and M. A. J. Van Montfort, Maximum-likelihood estimation of the general extreme-value distribution parameters, J. Hydrol., 47, 187–192, 1980.
[17]. Zenteno Jiménez José Roberto, Prediction of Concentrations of Ozone Levels in México City using Probability Distribution Functions, International Journal of Latest Research in Engineering and Technology (IJLRET) || Volume 04 - Issue 07 || July 2018 || PP. 35-45
[18]. Zenteno Jiménez José Roberto. A Methodology for Obtaining news Probability Distributions Functions Normal and Extreme Value for Bayesian Inference and Stochastic Mixed Gaussian Case One: For Daily Concentration Data Maximum Ozone. International Journal of Latest Research in Engineering and Technology (IJLRET)|| Volume 04 - Issue 11 || November 2018 || PP. 15-35
[19]. Statistical Modeling and Computation, Dirk P. Kroese – Joshua C.C. Chan, Springer Ed. 2014, Bayesian Inference Chapter 8, 236 page.
[20]. Zenteno Jiménez José Roberto, Analysis of Concentrations of Carbon Monoxide Levels in Mexico City with Probability Distribution Functions and its Analysis of Variance International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031 www.ijlret.com || Volume 06 - Issue 12 || December 2020 || PP. 01-11