International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 04 - Issue 07 || July 2018 || PP. 35-45

Prediction of Concentrations of Ozone Levels in México City using Probability Distribution Functions

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

Abstract: The study includes a data analysis from 2010 to 2017, which was proposed to obtain the best or best probability distribution functions that modeled ozone concentrations in México city, using the following pdf, exponential distribution function, gaussian reverse distribution function, normal log distribution function and gama distribution function, to obtain the estimators, the maximum verosimidity and moments method was used for the pdf range, for the estimation of the forecast model we used rmse, mse, coefficient of determination and prediction accuracy, in turn, a forecast is made for days of exceedance for this 2018 corroborating with the official air page of México City.
Keywords: Ozone, Probability Distributions Functions, Fit Indicators

México City has had a quite marked record with the evaluation of this atmospheric pollutant, the Ozone which decades ago the Government and Government and Scientific Institutions have given themselves to the task of trying to diminish the concentrations of this contaminant that overwhelms to the City and to the Capitalinos by means of established Legislations and technological means as well as investigations regarding the components of the gasolines that use the automotors and systems of transport within the City. We know that the temperature directly affects the Ozone concentration, and this year 2018 has been affected by such temperature rises so it is expected for the next months and represents a health risk, so the Environmental Commission of the Megalopolis (CAMe) recommended the population to take precautions this year.

The season of high concentrations of this pollutant starts around mid-February and ends with the first rains in June. Likewise, the exposure of high levels of ozone is associated with physiological and inflammatory effects in the lungs of healthy young adults, who exercise in the open air. Therefore, it is recommended to reduce the exposure time outdoors, especially to the most vulnerable population such as children, seniors, pregnant women and people with respiratory problems andcardiovascular diseases, there is a direct relationship between chronic exposure to the pollutant and the increase in cases of morbidity and mortality.

According to México, the Official Mexican Standard (NOM-020-SSA1-2014) recommends concentrations lower than 0.095 ppm for the average of 1 hour, and lower than 0.070 ppm for the average of 8 hours (annual maximum). Therefore, tropospheric ozone found at surface level in urban areas occurs when nitrogen oxides (NOX) and volatile organic compounds (VOCs) react in the atmosphere in the presence of sunlight. In high concentrations it can put human health and vegetation at risk.

There is a direct correlation with the increase in temperature, here are two graphs which shows the high tendency of ozone concentrations to the increase in temperature as well as the solar radiation and its maximum point is at noon.

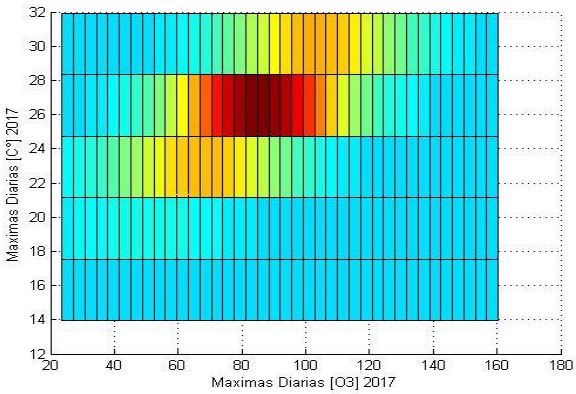


Figure 1. Relationship between the Temperature and the Concentration of Daily Ozone Maximes in ppb 2017 in México City (own elaboration)

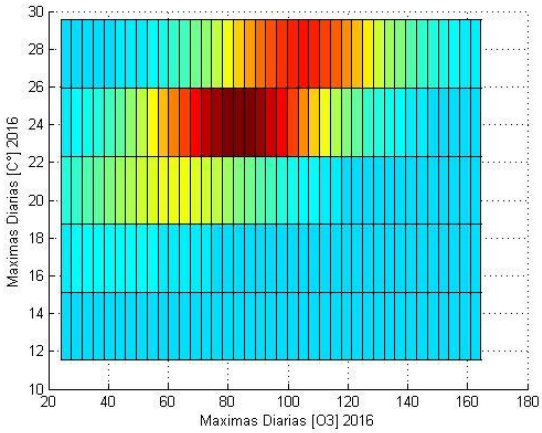


Figure2. Relationship between the Temperature and the Concentration of Daily Ozone Maximes in ppb 2016 in México City (own elaboration)

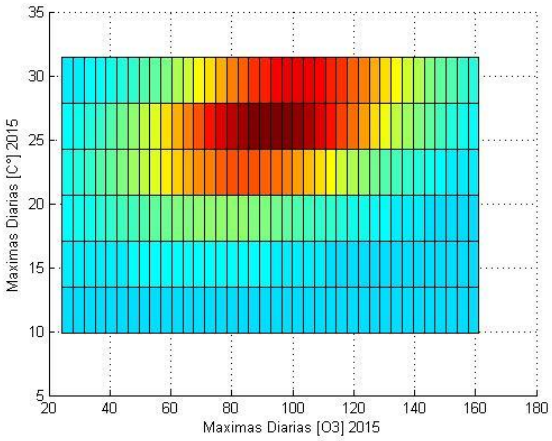


Figure3. Relationship between the Temperature and the Concentration of Daily Ozone Maximes in ppb 2015 in México City (own elaboration)

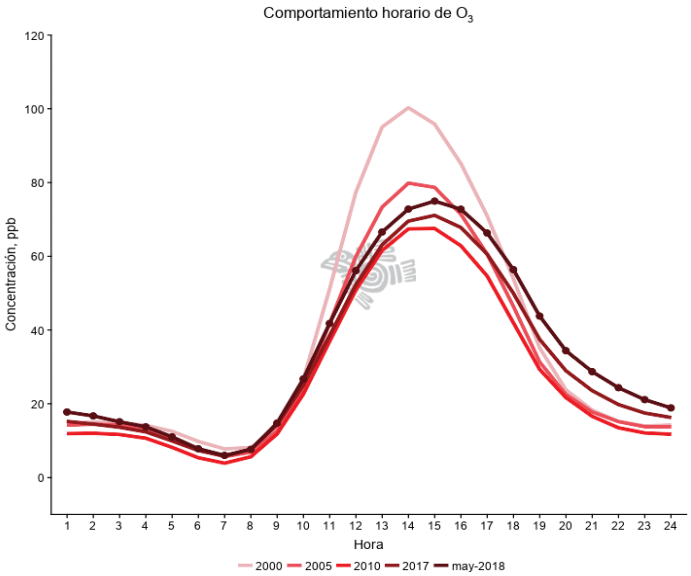
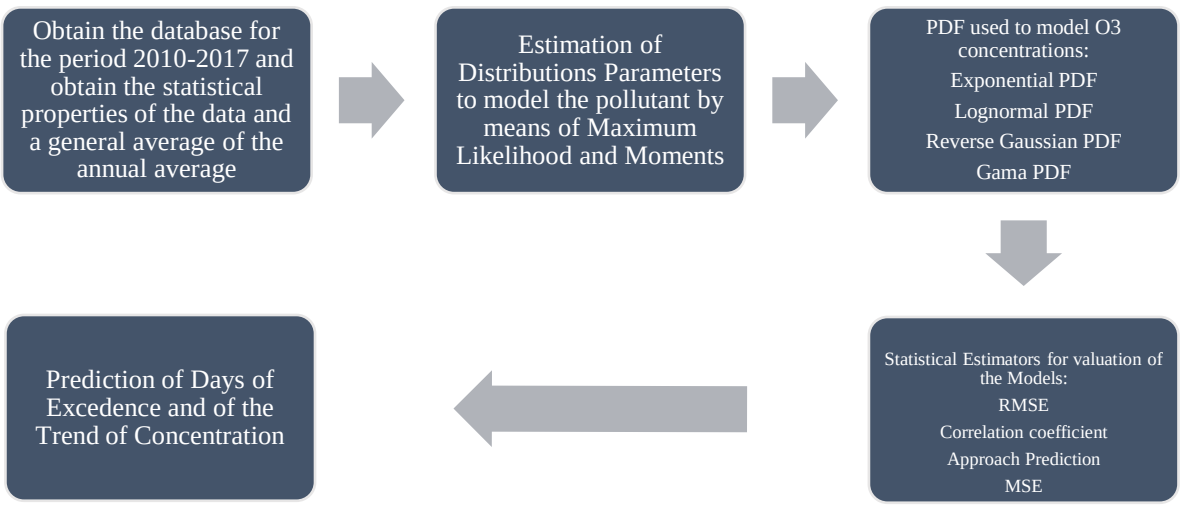


Figure4. Relationship between Daily Radiation and Ozone Concentration in ppb in México City (Source: http://www.aire.cdmx.gob.mx/)

Probability Distribution Functions and Methodology



Four probability distribution functions were used, which are the Exponential Distribution function, the LogNormal Distribution function (Kao and Friedlander, 1995), the Gaussian Reverse Distribution function (Chhikara and Folks, 1989), and the Distribution function Gama (Berger et al., 1982).

Table 1. Probability Distribution Functions and their Parameters.

Distribution	Probability density function	Parameters
Exponencial	$f(x) = \left(\frac{1}{mu}\right) exp^{-\left(\frac{x}{mu}\right)}$	$mu = \left(\frac{1}{N}\right)\sum_{i=1}^{n} x_i$
LogNormal	$f(x) = \left(\frac{1}{xRho\sqrt{2\pi}}\right) exp^{\left(\frac{-(ln(x)-lambda)^2}{2Rho^2}\right)}$	$Rho = \left(\frac{1}{N}\right)\sum_{i=1}^{n} ln(x_i)$ $Lambda = \left(\frac{1}{N-1}\right)\sum_{i=1}^{n}(ln(x_i) - Rho)^2$
Reverse Gaussian	$f(x) = \left(\frac{Rho}{2\pi x^3}\right)^{\frac{1}{2}} exp^{\left(\frac{-Rho(x-mu)^2}{2mu^2 x}\right)}$	$mu = Mean$ $Rho = \frac{(N-1)}{\sum_{i=1}^{n}\left(\frac{1}{x_i} - \frac{1}{mu}\right)}$
Gama	$f(x) = \frac{Beta^{alfa}}{\Gamma(alfa)} x^{alfa-1} e^{-Betax}$	$Beta = \frac{\sum x_i^2}{\sum x_i} - \frac{\sum x_i}{N}$ $Alfa = \frac{(\sum x_i)^2}{N\sum x_i^2 - (\sum x_i)^2}$

Adjustment indicators

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 gave the data. They are the mean square error (RMSE), mean square error (MSE), prediction accuracy (AP) and coefficient of determination (R2) Table2 gives the equations for the adjustment indicators that have been used by Lu (2003) and Junninen et al. (2002).

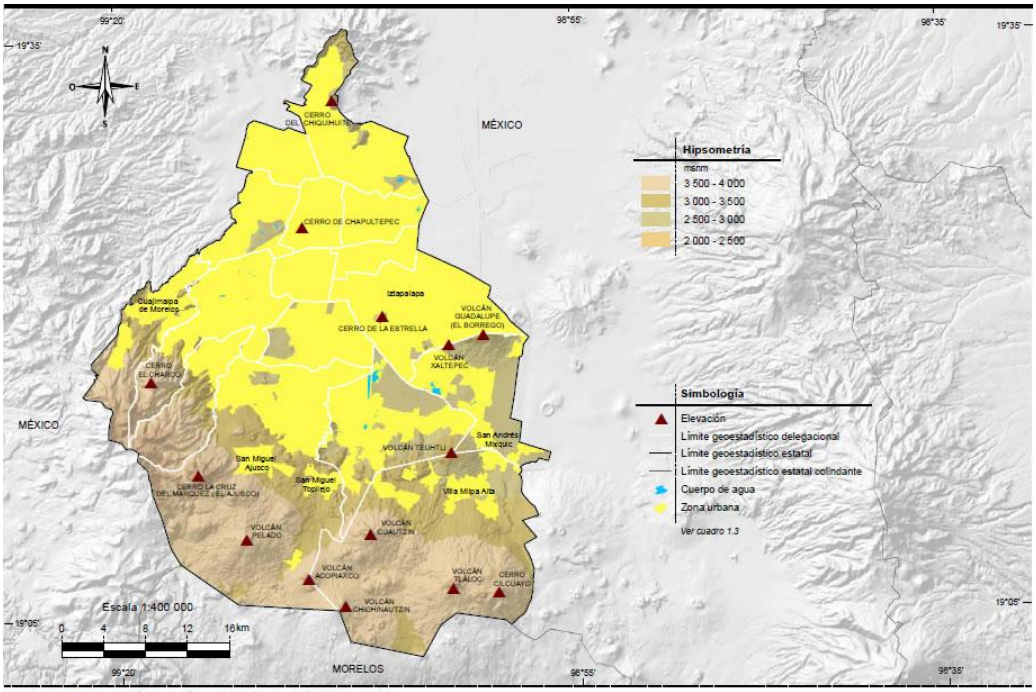
Table 2. Adjustment Indicators

Indicator	Equation
Root Mean Square Error	$$RMSE = \sqrt{\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(Pi - Oi)^2}$$
Mean Square Error	$$RMSE = \left(\frac{1}{N}\right)\sum_{i=1}^{N}(Pi - Oi)^2$$
Coefficiente of Determination	$$R^2 = \left(\frac{\sum_{i=1}^{N}(Pi - P)(Oi - O)}{NS_pS_o}\right)^2$$
Prediction Accuracy	$$AP = \frac{\sum_{i=1}^{N}(Pi - Oi)^2}{\sum_{i=1}^{N}(Oi - Oi)^2}$$

Notation: N =Number of Observations, P_i=Predictive Values, O_i = Observed Values, P =Average of Predicted Values, O =Average of Observed Values, S_p =Standard Deviation of Predicted Values, S_o =Standard Deviation of Values Observed.

Study Area

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 wind circulation and the dispersion of pollutants makes it difficult.



Fuente: INEGI. *InformaciónTopográfica Digital Escala 1:250 000, serie II.*
INEGI. *InformaciónTopográfica Digital Escala 1:50 000, serie I.*
INEGI. *Modelo Digital de Elevación, versión 2.0. Noviembre 2012.*

Figure 5. Relief of the México City (Source: https://www.paratodomexico.com/)

Figure 6. Contamination of the México City (Source: http://www.garuyo.com/)

Statistical Description of the Data

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

Table3. Description of the Ozone Data Statistics

Number of Data	**61368**
Mínimum	2.0 ppb
Máximum	210 ppb
Mean	38.43 ppb
Variance	1291 ppb
Standard Deviation	35.93 ppb
Median	11 ppb
Skewness	0.772 ppb
Kurtosis	18.03 ppb
Mode	2.0 ppb

Results

In the following table we can see that the best adjusted probability distribution function was the exponential, followed by the log normal distribution function, with an R2 of almost 0.99 and that of the log normal with a 0.92, we can also see that the RMSE of both it is almost the same and the smallest of the other distribution functions, and the Prediction Accuracy is greater in the exponential distribution function than the others, although the adjustment of the exponential distribution function would not have been very good with an around 0.90 in R2 would have opted to use the exponential distribution function anyway shows the best values shown in the other adjustment indicators.

Table 4. Estimation Parameters and Adjustment Indicators for the 2010-2017 Trend

Distribution	**Parameters Estimation**	**RMSE**	**MSE**	R^2	**AP**
Exponencial	**mu =36.35**	**0.3596**	**0.1293**	**0.9982**	**0.7600**
LogNormal	**Rho=2.45 Lambda=1.27**	**0.3549**	**0.1259**	**0.9227**	**0.7575**
Reverse Gaussiana	**mu=19.21 Rho=6.06**	**0.4512**	**0.2036**	**0.9171**	**0.6678**
Gama	**Alfa=1.14 Beta=16.79**	**0.3902**	**0.1523**	**0.8874**	**0.7267**

Cumulative probability of the 2010-2017 trend

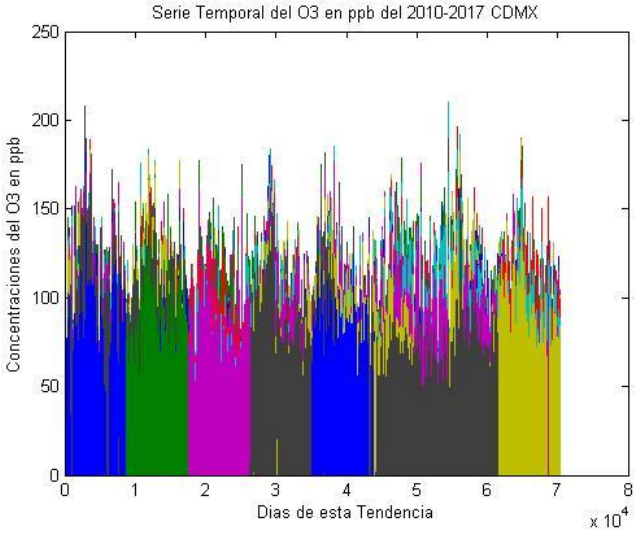


Figure 7. Temporal Series of Ozone Concentration in ppb in México City Trend 2010-2017 in days

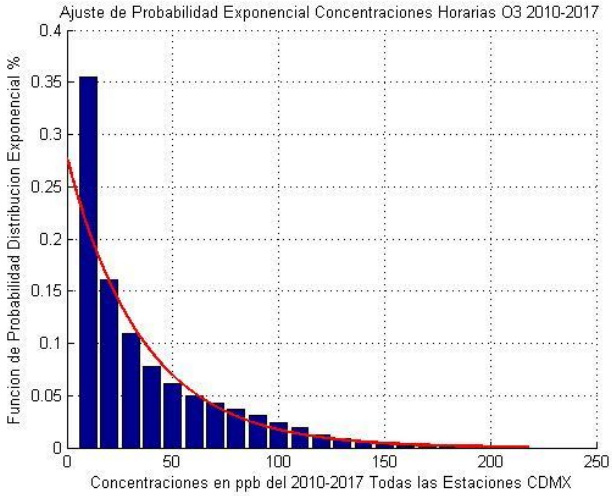


Figure 8. Histogram of the Ozone Concentration in ppb in México City and Exponential Adjustment adjusted to the Histogram 2010-2017.

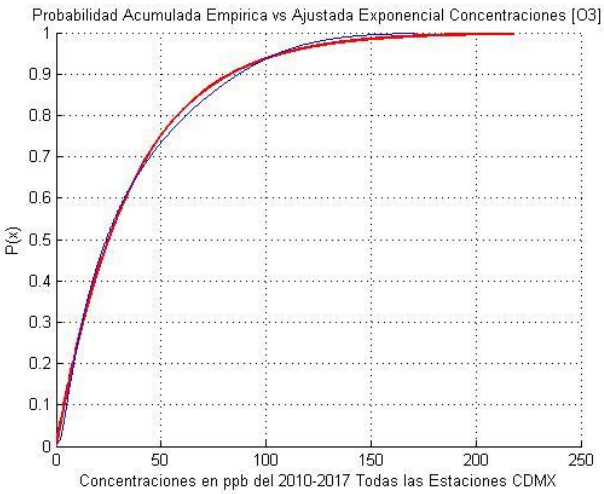


Figure 9. Exponential Distribution Function and Maximum Ozone Concentration in ppb in México City.

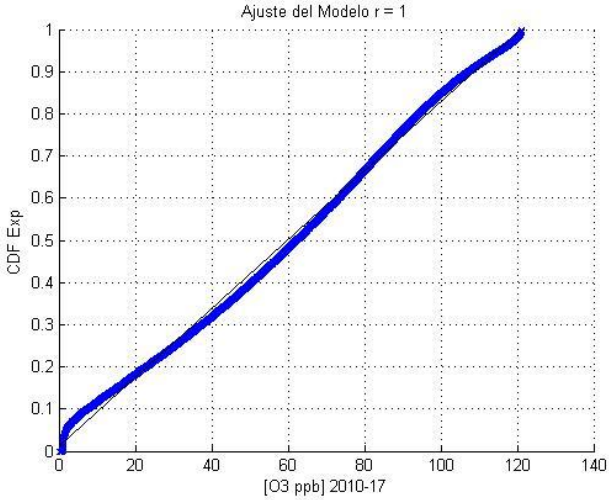


Figure 10. Adjustment of the Exponential Model as the best Adjustment of the exponential distribution function with the Average Ozone Concentration in ppb in México City.

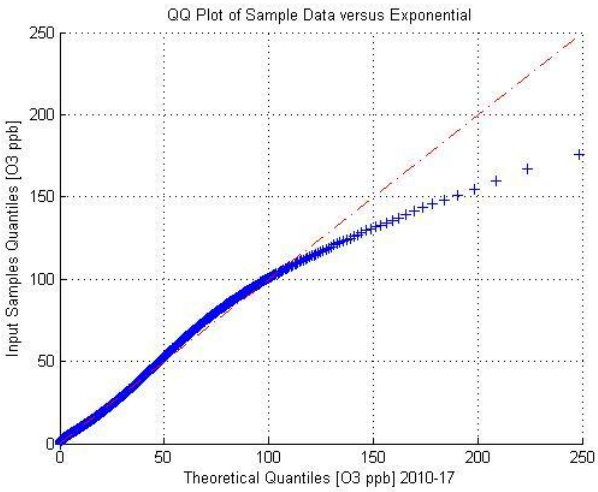


Figure 11. This QQ graph shows the behavior of the input data with respect to an exponential behavior which follows the trend, except for very high values of ozone concentration in ppb México City.

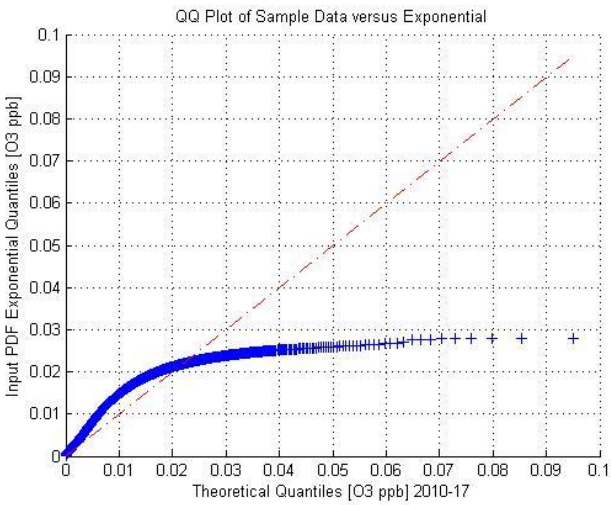


Figure 12. In this QQ graph the behavior of the input data is shown with respect to the exponential distribution function obtained from ozone concentration data in ppb México City.

Now taking the average of the Ozone Concentrations of the Tendency 2010-2017 we obtain the following graph, which we can observe the increase of the concentration of ozone in hot seasons from 90 days to 170 days of the year. Heat and Radiation is between March, April, May and June.

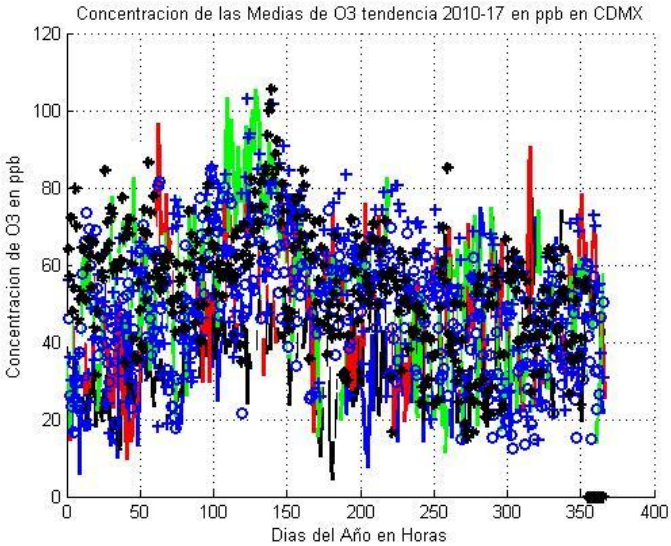


Figure 13. Concentration of Ozone Means in ppb in México City.

Forecast of ozone levels for 2018 in days of exceedance with the exponential pdf for the first 6 months.

Now the prognosis of days of leave for the year 2017 and for the year 2018 will be given with the official comparisons of the page of Mexico City in the days of Excedencies per year as well as per month, which further corroborates the study done with good approximation results. We can also observe that the variation is minimal in the means shown in figure 10, we can say that the measurements of ozone levels almost remain very similar every year, but with their differences between each given year, so we can make a forecast of concentration for the following year calculating the days of exceedance and we will check them with the real data of the official days.

It can also be noted that the average concentration of ozone varies with respect to the hot season and the non-hot season, which is half of the year on, that is when the trend of the concentrations drops, so the exponential distribution function changes in the maximum concentration with the trend going down but very little so we take two quantiles to make the forecast which one is at Maximum Ozone Concentration and the other lowest, in days that reverberates in about + - 4 days and averaging from 2010 onwards.

Table 5. Forecast of the Exponential Adjustment of the Tendency 2010-2017

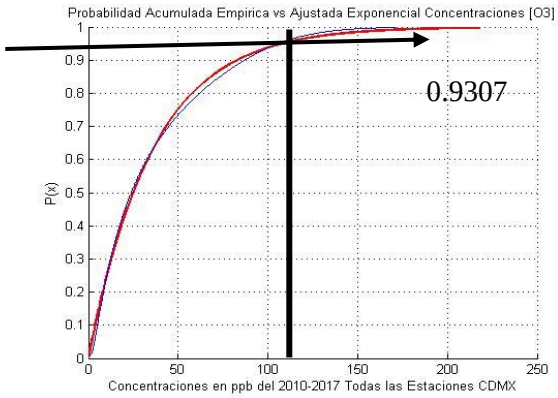
Trend 2010-2017 for this 2018 to Maximum Ozone Concentration	Forecast for 2018
	Above 95 ppb the Cuantil is 0.9307 now for 2018 it gives us 227 days maximum and minimum 221 days For this January 2018 by the same Cuantil gives us 19 days maximum. For this February 2018 by the same Cuantil gives us 18 days equal For the following months until August it may increase or decrease up to 19 days or less than 14 days. **Query Link:** **http://www.aire.cdmx.gob.mxyou can check the forecasts**

Table 6. Forecast of the Exponential Adjustment of the Tendency 2010-2017

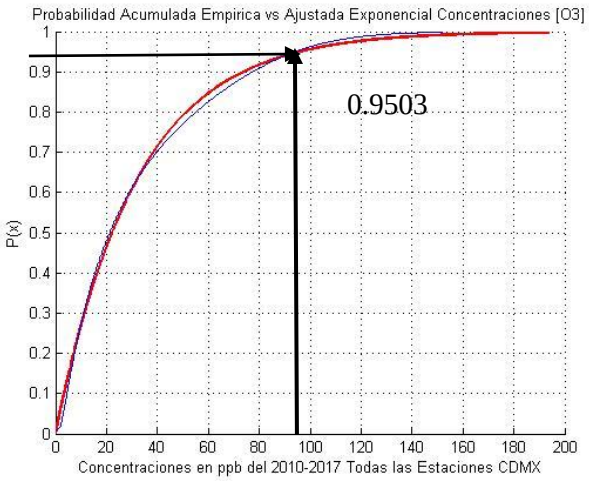
Tendency 2010-2017 for this 2018 with concentration of ozone going down	Forecast for 2018
	For this 2018 gives us the Cuantil of 0.9503. For this January 2018 for the same Cuantil gives us 14 days. For this February 2018 for the same Cuantil gives us 13 days For the following months until August it may increase up to 19 days or more For the months beginning August for the same Quantum gives us 14 to 9 days above 95 ppb of Ozone Exceedances **Query Link: http://www.aire.cdmx.gob.mxyou can check the forecasts**

Table 7. Forecast of the Exponential Adjustment of the Trend 2010-2017 for this year concentrations above 95 ppb

2018	Official Days	Cuantil 0.9307 in days	Cuantil 0.9503 in days	Forecast for the Year 2018 In days
January	**14**	**20**	**14**	**14 a 22**
February	**20**	**20**	**15**	**14 a 22**
March	**28**	**21**	**16**	**15 a 27**
April	**28**	**21**	**16**	**16 a 31**
May	**22**	**21**	**16**	**17 a 29**
June	**18**	**21**	**16**	**14 a 22**

Query Link: http://www.aire.cdmx.gob.mxyou can check the forecasts

Let's now make an observation having the quantile at maximum concentration obtained in accumulated days from 2010 to 2017 we get 202 days with 0.9307 and now with the calculation of days of exceeding one year we get 25 days adding both results gives us 227 days of ozone exceedances for 2017 and the official value in days of ozone exceedance for the city of Mexico in 2017 is 225 days.

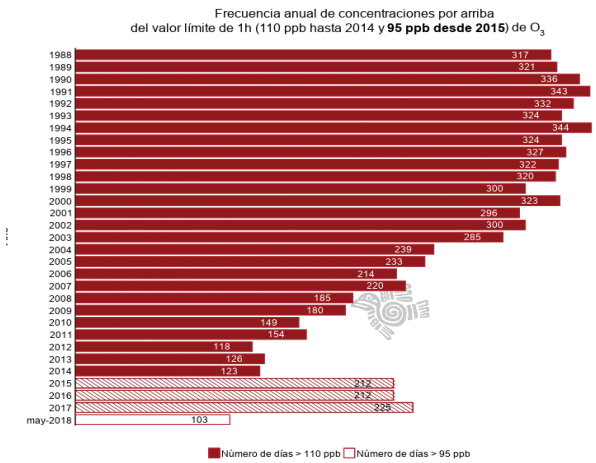


Figure 14. Annual frequency of Ozone concentrations above 95 in ppb in México City (Source: http://www.aire.cdmx.gob.mx)

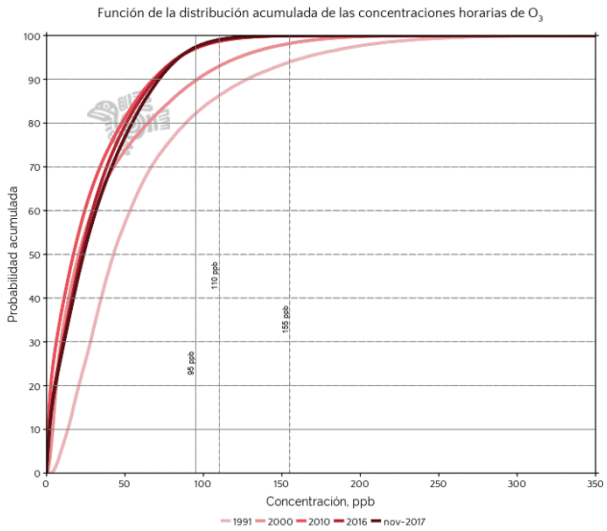


Figure 15. Cumulative distribution function of Ozone concentrations above 95 in ppb in México City (Source: http://www.aire.cdmx.gob.mx)

Let's see the graph of the official page of the City of Mexico the study includes from 1991-2017 are 9490 days now the norm is 95 ppb but not from that year of 1991, so taking as the general norm of all those years of Probability Accumulated of México City is 0.9763 which with that probability gives us an exceedance of leave days of 224.913.

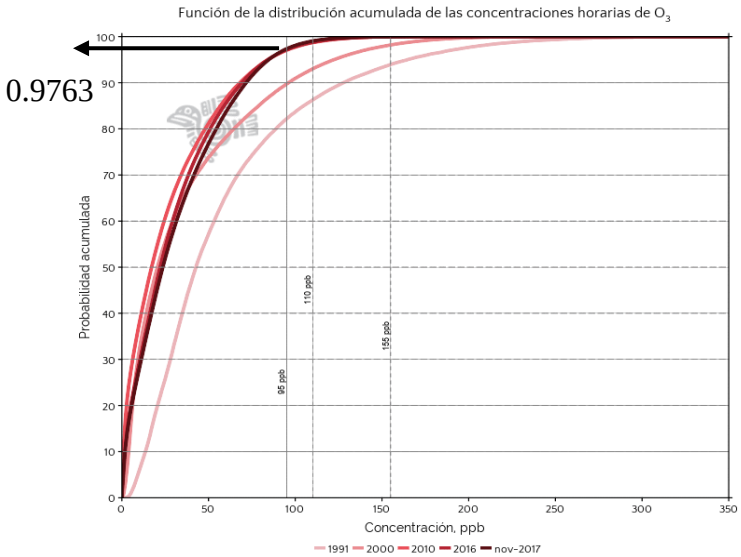


Figure 16. Accumulated distribution function of Ozone concentrations above 95 in ppb in México City (Source: http://www.aire.cdmx.gob.mx)

Conclusions

In this study we can say that of the 4 probability distribution functions that were proposed, two of them are quite close to the real data, being in this case the exponential distribution function the best of them, followed by the normal log, also we can see that the imposed average varies depending on the hot and unheated season which tends to lower the ozone concentrations and therefore the cumulative distribution function changes a bit and that corresponds to two measurements that give us the possible concentrations to forecast , we also obtain a simple method to calculate the days of exceedences of the ozone concentrations giving a good approximation to the real days and with that we verify that the concentration really changes from one season to another, we show it with the official measurements of the page of México City of atmospheric monitoring, it is left as a later study the possible modeling of the Ozone concentration with the second probability distribution function the normal log.

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]. Chhikara, R. S. and Folks, J. L. (1989) The inverse Gaussian distribution as a lifetime model. Thechnometrics. 19 (4), 461 – 468

[4]. Data base of Ozone website of México City http://www.aire.cdmx.gob.mx/

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

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

[7]. H.C. Lu, G.C. Fang, Y.S. Wu, Estimating the frequency distributions of particulate matter and their metal elements in a Temple, J.Air Waste Manag., 56 (2012) 1033-1040.

[8]. Heat waves and increased levels of ozone in the Cdmx: an obvious correlation, A. Sandoval-Villalbazo (2016).). http://ibero.mx/prensa-y-multimedios?sitio=detalle_comunicado.php&id_noticia=2984&foto_principal=1

[9]. 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.

[10]. 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

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

[12]. Mage, D.T., Ott,W.R., 1984. An evaluation of the method of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmospheric Environment 18, 163–171.

[13]. 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.

[14]. P.G. Georgopoulous, J.H. Seinfeld, Statistical distributions of air pollutant concentrations, Environ. Sci. Technol., 16 (1982) 401A-416A.

[15]. 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.

[16]. 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.

[17]. Work presented at the Congress of the Mexican Geophysical Union 2017 Prediction of Ozone Concentrations due to Probability Distributions for the CDMX.https://www.raugm.org.mx/2017/pdf/constancia.php?clave=809

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