International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 10 - Issue 04 || April 2024 || PP. 01-15

Gaussian Mixture of Several Components (Machine Learning) of Daily Ozone and Temperature Maximums in Mexico City Trend 2010-2023

M. Sc. Zenteno Jimenez Jose Roberto
Geophysical Engineering, National Polytechnic Institute, Mexico City,
ESIA-Ticoman Unit, Mayor Gustavo A. Madero
E-mail; jzenteno@ipn.mx

Abstract: In this study we study the trend of maximum Ozone concentrations in Mexico City and Maximum Temperatures, based on the Bivariate Analysis methodology, subsequently using algorithms related to the topic of Data Clustering, especially K-means to be able to observe the classification of groups and subsequently the trend of the different groups of each year of Ozone concentrations and Maximum Temperatures in the City, subsequently we use the Pattern Recognition Gaussian mixture model algorithm.

Keywords: Bivariate Analysis, K-means, Pattern Recognition Gaussian mixture model, Ozone, Maximum Temperatures

I. Introduction

The trend of maximum Ozone concentrations in Mexico City and Maximum Temperatures will be based on the Bivariate Analysis methodology, which, as we have seen in previous studies, will see the behavior for a couple of years of the Maximum Ozone concentrations. Daily ozone in Mexico City, a treatment will also be done with the K-means methodology to be able to observe the classification of groups if in any case the concentrations of each year are part of an already background of Ozone in the city at that same time. group or if it has quite notable variations each year, subsequently the trend of the different groups of each year of Ozone concentrations and Maximum Temperatures in the City, the Pattern Recognition Gaussian mixture model algorithm is used to see the trend or pattern of these concentration data and the maximum daily temperatures of Mexico City and their relationship between both

II. Bivariate Analysis

It is useful to determine if there is a correlation between variables and, if so, the strength of the connection, to find trends and patterns in the data. It is useful for making predictions about the value of a dependent variable based on changes in the value of an independent variable. in this case the behaviors of the daily Ozone Maximum against the Daily Temperature Maximum in Mexico City, [See http://www.ijlret.com/Papers/Vol-07-issue-03/1.B2021164.pdf] so the Analysis of its Bivariate probability distribution functions is used, this study has already been done in past articles to see the relationship between both variables, but now it will be done for each group of bone variable of only the Maximum Ozone concentrations in Mexico City.

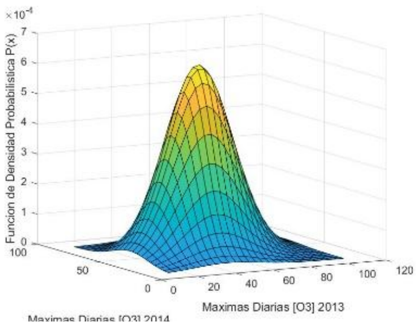


Figure (1) Gaussian Distribution 2D

III. K-means

K-means is also a partitional clustering method in which we need to specify the number of clusters before starting the clustering process. Suppose the number of clusters is m, then we can define an objective function as the sum of squared distances between a data point and its nearest cluster centers. We can follow a procedure to minimize the objective function iteratively by finding a new set of cluster centers that can reduce the value of the objective function in each iteration. Here we use this methodology to see if there are differences between the 2 groups formed or formed one united group.

K-means clustering (k-means for short), is one of the most well-known methods for data clustering. The goal of k-means is to find k points in a data set that can best represent the data set in a certain mathematical sense.

Data compression: We can use these cluster centers to represent the original data set. Since the number of centers is much smaller than the size of the original data set, the goal of data compression can be achieved.

Data Classification: We can use these cluster centers for data classification so that the calculation load and the influence of noisy data are reduced.

K-means is also a partitional clustering method in which we need to specify the number of clusters before starting the clustering process.

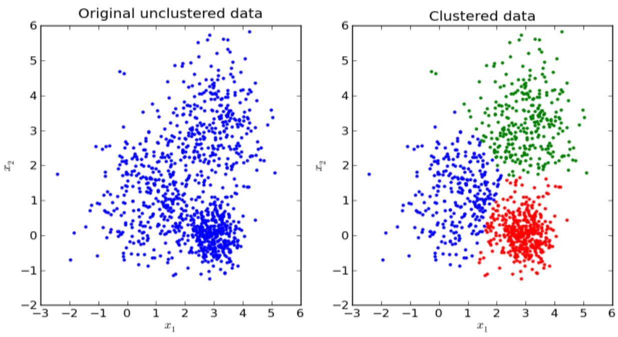


Figure (2)

There are some other k-medias facts that we should keep in mind:

K-means iteration can only guarantee non-increment of the objective function. However, it cannot guarantee the finding of the global minimum of the objective function, there are no efficient methods that can guarantee the finding of the global minimum of the objective function. Therefore, it is advisable to run k-means several times starting from different initial centers and then keep the best result.

There are some other k-medias facts that we should keep in mind:

A better set of initial centers will have a positive influence on the final results of the grouping.

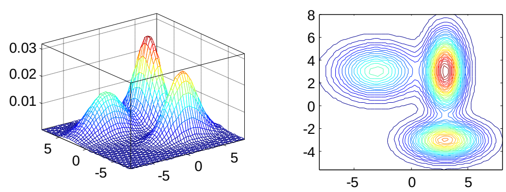
IV. Stochastic Gaussian Mixture

The Gaussian Mixture Model (GMM for short) is an effective tool for data modeling and pattern classification. GMM assumes that the modeled data are generated by a probability density distribution that is the weighted sum of a set of Gaussian PDFs. By using EM (expectation maximization), we can identify the optimal set of parameters for GMM iteratively.

Characteristics they have is that they are highly flexible for complicated data sets. There is no guarantee of convergence to the general optimum, we will use the 2D form.

$$gmm(x;\alpha_1,\mu_1,\Sigma_1,\alpha_2,\mu_2,\Sigma_2,\alpha_3,\mu_3,\Sigma_3)=$$

$$\alpha_1 g(x;\mu_1,\Sigma_1)+\alpha_2 g(x;\mu_2,\Sigma_2)+\alpha_3 g(x;\mu_3,\Sigma_3), \text{where} \sum_{i=1}^{3}\alpha_i=1.$$



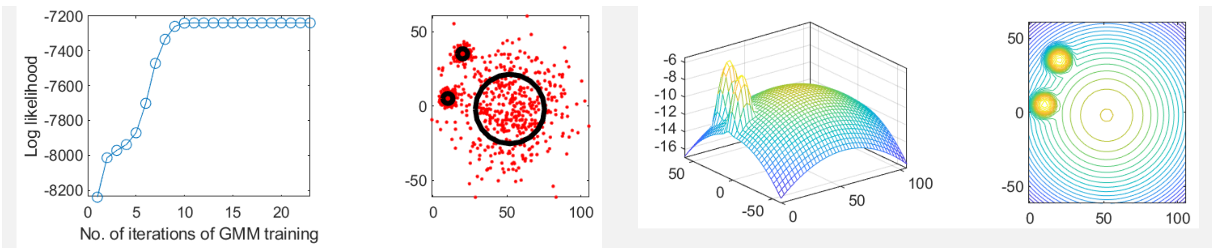


Figure (3)

Source (http://mirlab.org/jang/books/dcpr/)

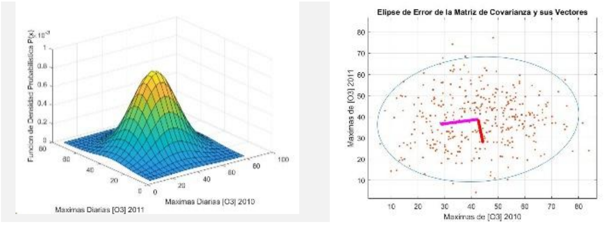
With full Covariance type

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12} & \rho_{13} \\ \rho_{21} & \sigma_2^2 & \rho_{23} \\ \rho_{31} & \rho_{32} & \sigma_3^3 \end{bmatrix}$$

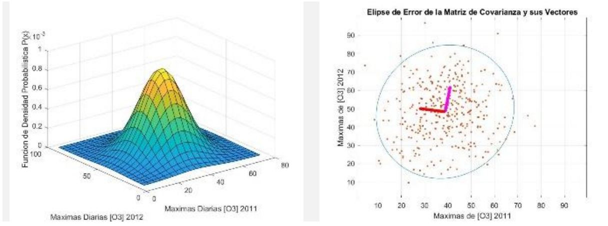
Let's see the following Results

Table 1. Bivariate Analysis of Ozone Maximum from year to year in Mexico City.

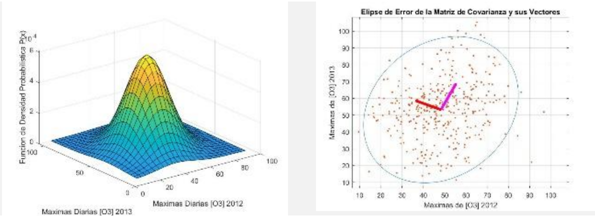
2010-2011



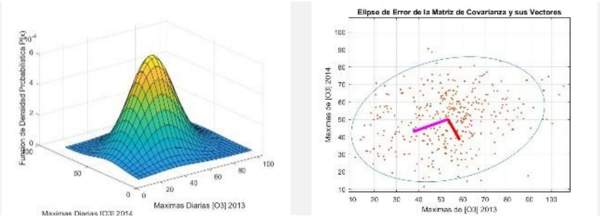
2011-2012



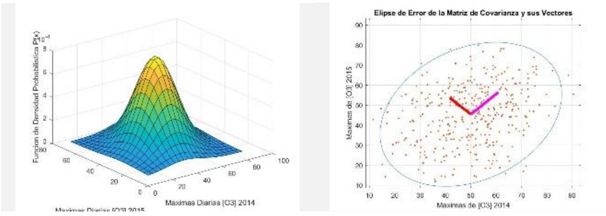
2012-2013



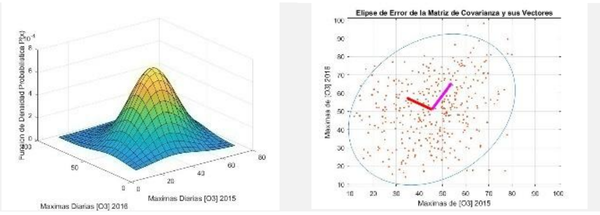
2013-2014



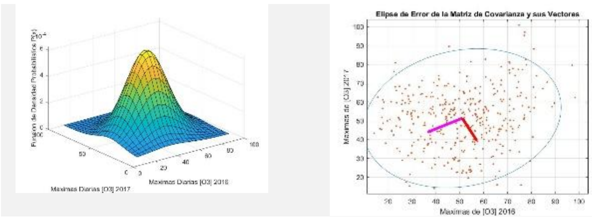
2014-2015



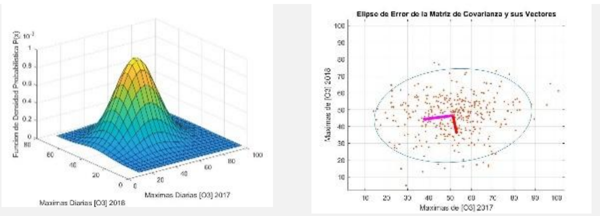
2015-2016



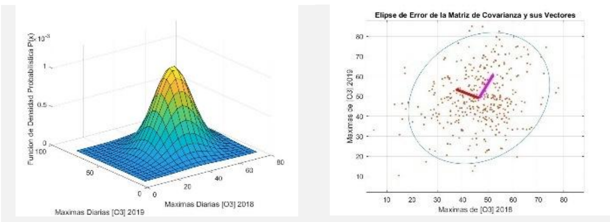
2016-2017



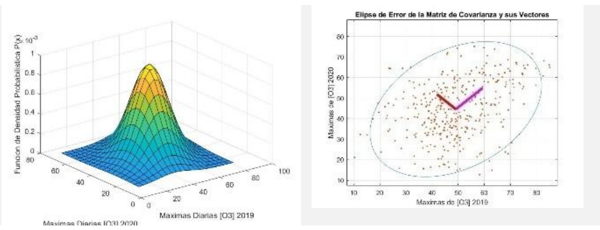
2017-2018



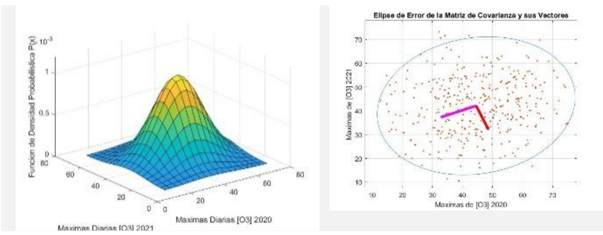
2018-2019



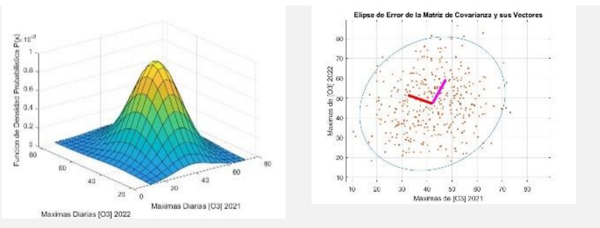
2019-2020



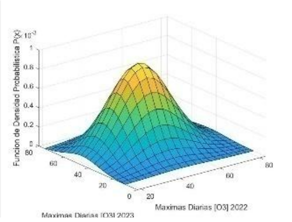
2020-2021



2021-2022



2022-2023



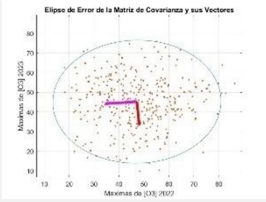


Table 2. Data from the Covariance Matrix and its Adjustment

2010-2011

Sigma =		R =	
236.7015	15.2581	1.0000	0.0820
15.2581	146.2184	0.0820	1.0000

2011-2012

Sigma =		R =	
146.2184	12.2280	1.0000	0.0680
12.2280	221.3347	0.0680	1.0000

2012-2013

Sigma =		R =	
221.3347	54.9333	1.0000	0.2073
54.9333	317.3343	0.2073	1.0000

2013-2014

Sigma =		R =	
317.3343	58.4664	1.0000	0.2241
58.4664	214.4227	0.2241	1.0000

2014-2015

Sigma =		R =	
214.4227	59.5023	1.0000	0.2762
59.5023	216.3687	0.2762	1.0000

2015-2016

Sigma =		R =	
216.3687	62.9052	1.0000	0.2525
62.9052	286.7594	0.2525	1.0000

2016-2017

Sigma =		R =	
286.7594	40.1880	1.0000	0.1571
40.1880	228.3284	0.1571	1.0000

2017-2018

Sigma =		R =	
228.3284	16.2849	1.0000	0.0943
16.2849	130.5380	0.0943	1.0000

2018-2019

Sigma =		R =	
130.5380	31.3080	1.0000	0.2040
31.3080	180.4835	0.2040	1.0000

2019-2020

Sigma =		R =	
180.4835	63.6961	1.0000	0.3526
63.6961	180.8123	0.3526	1.0000

2020-2021

Sigma =		R =	
180.8123	19.6424	1.0000	0.1231
19.6424	140.7614	0.1231	1.0000

2021-2022

Sigma =		R =	
140.7614	26.9536	1.0000	0.1649
26.9536	189.8730	0.1649	1.0000

2022-2023

Sigma =		R =	
189.8730	2.1219	1.0000	0.0121
2.1219	163.3078	0.0121	1.0000

By Clusters to be able to observe or classify which group the data belongs to

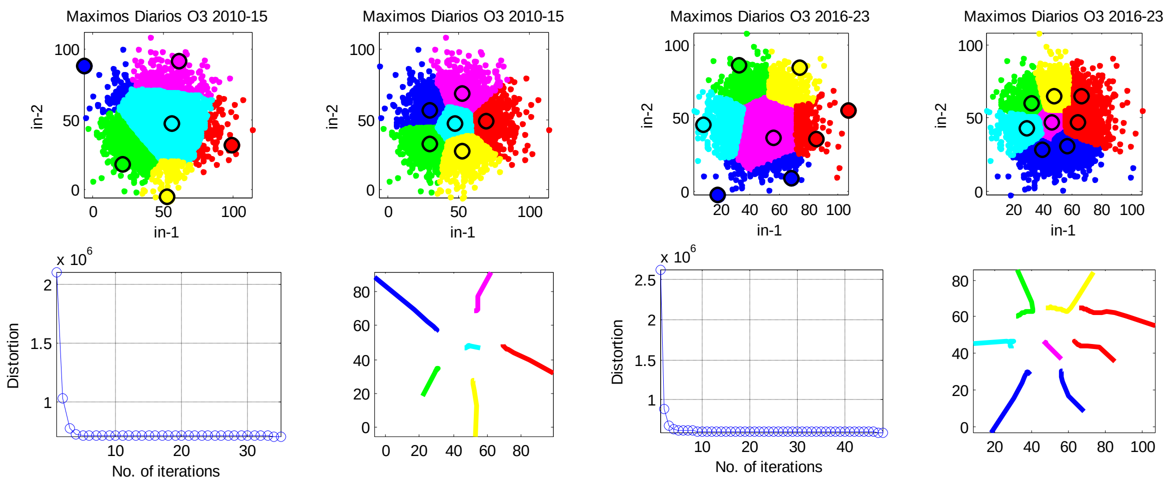
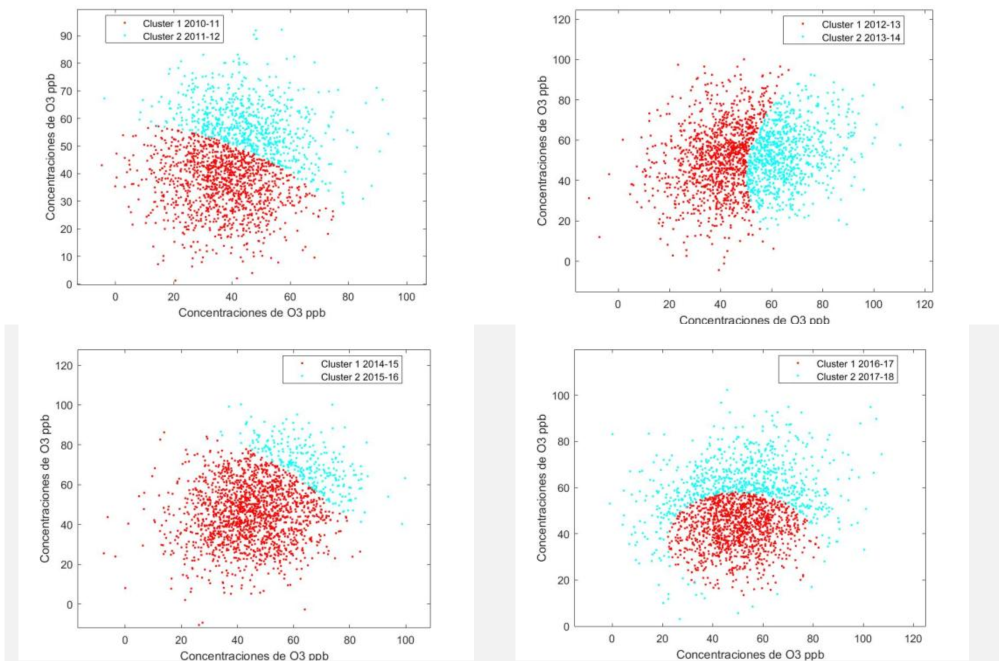


Figure (4)

The upper left graph is the initial centers and the corresponding groups, the top right graph is the final centers and corresponding groups. The bottom left graph is the distortion with respect to the number of iterations. The lower right graph is the trajectories of the centers during the grouping process.

In the image above, we can clearly identify the groups by visual inspection. If we set the number of groups to 6 to run k-means, the result is satisfactory. However, if there is no way to perform visual inspection (for example, when the data dimensions are more than 3), then we must use cluster validation methods to identify the optimal number of clusters.



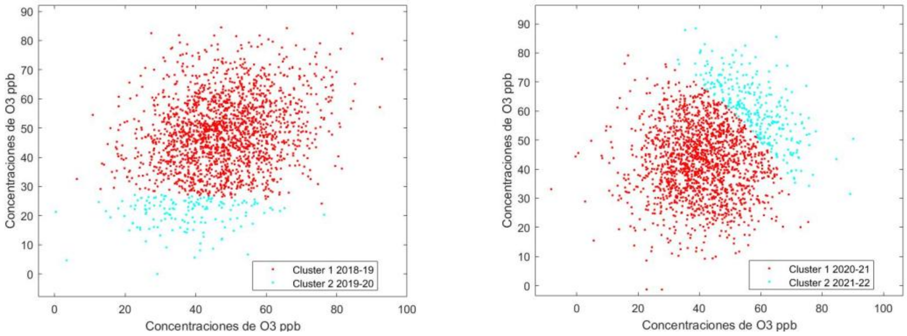
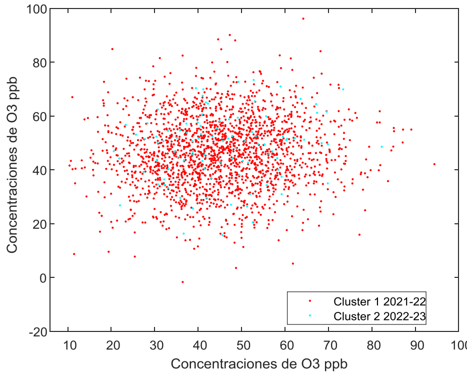
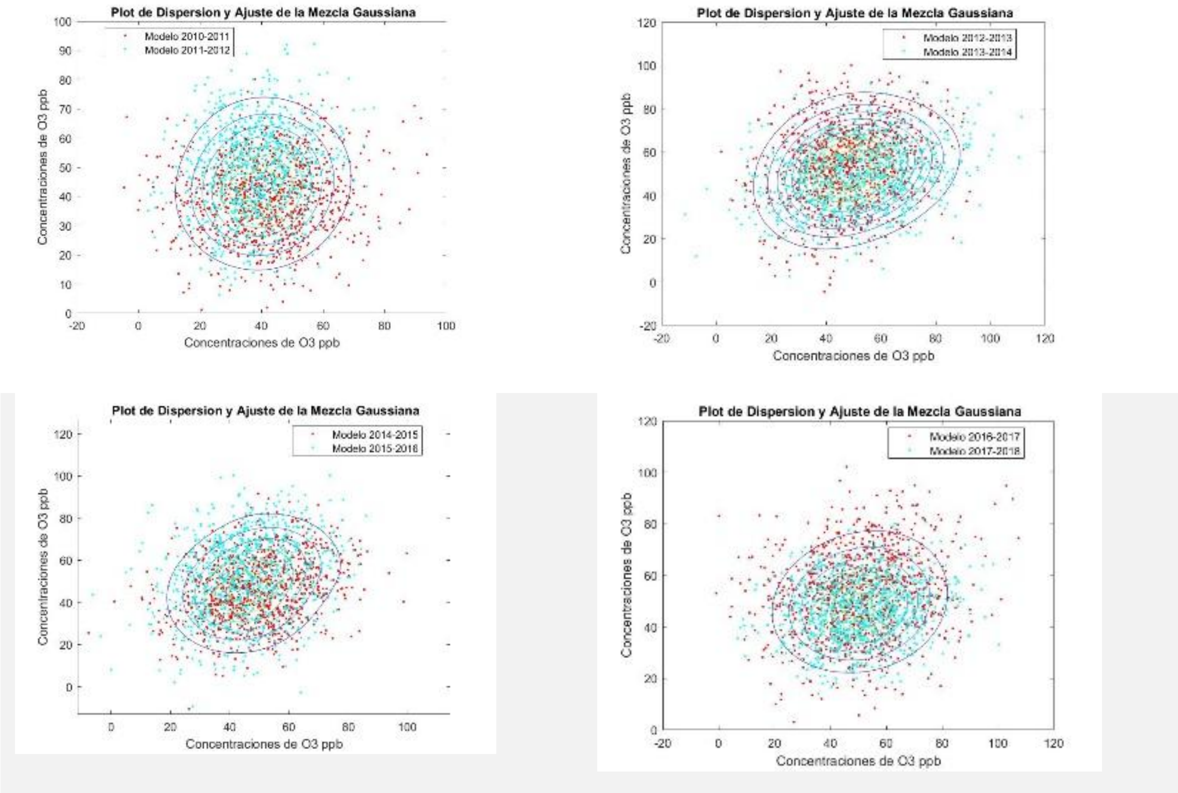


Figure (5) By Clusters by pairs of years of the Bivariate Distribution



Gaussian Mixture and its Scatter Plot



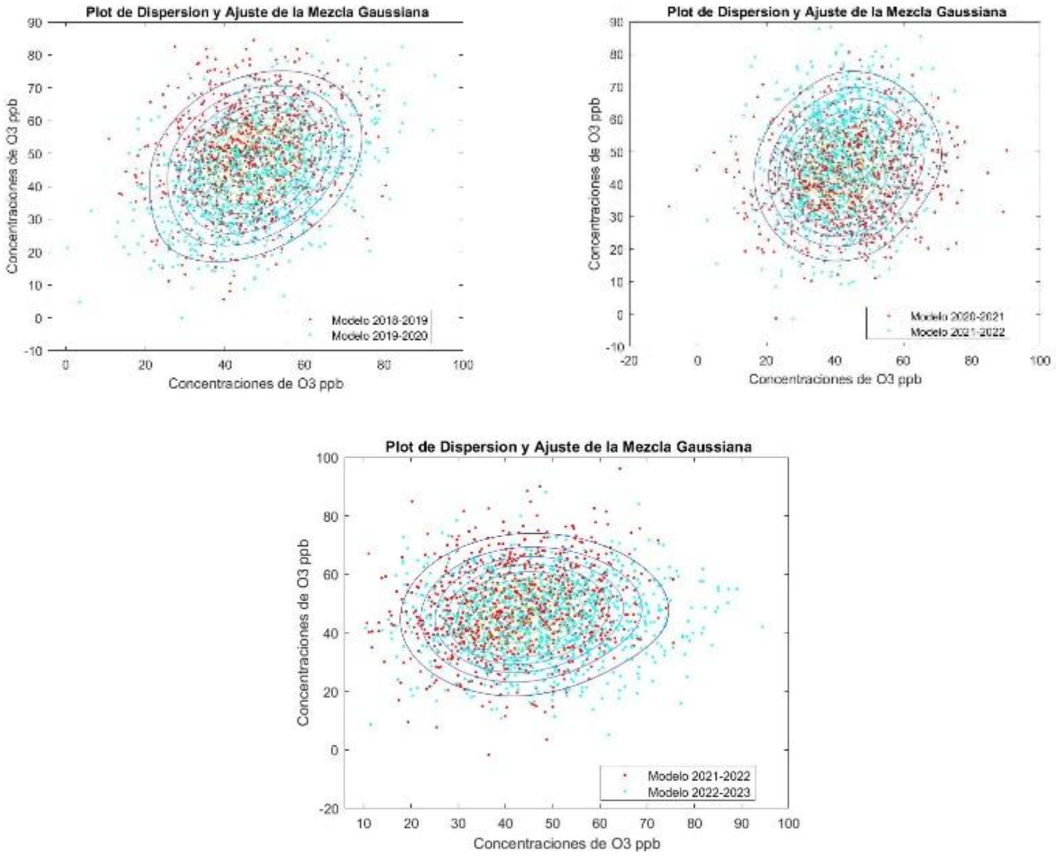
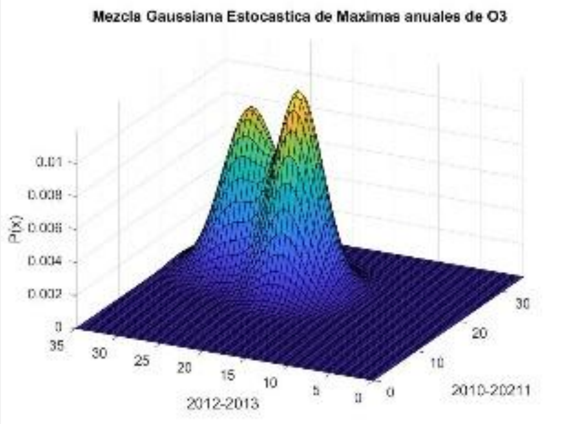
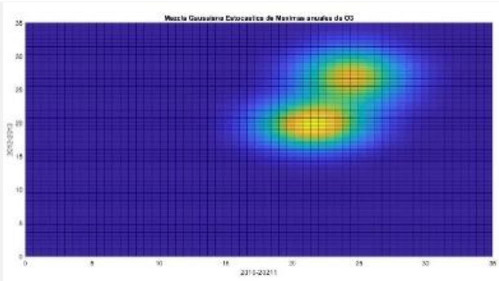


Figure (6)

Table 3. 2D Gaussian Mixture of Daily Ozone Maximum

2010-11 2012-13

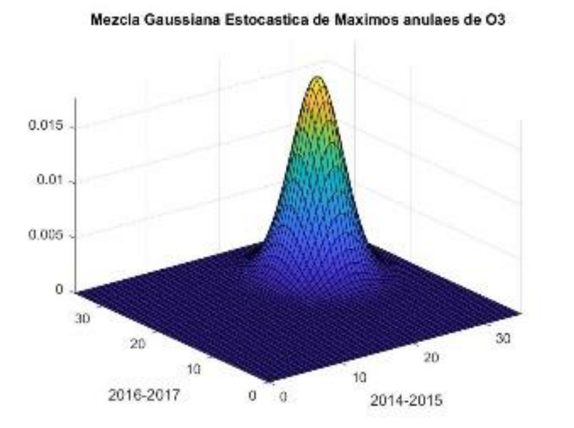


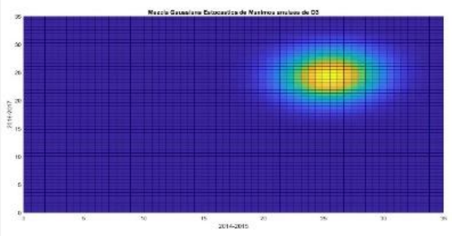


Component 1:
Mixing proportion: 0.500000
2010-2011
Mean: 21.2381 19.4059
Real Mean: 85 78

Component 2:
2012-2013
Mixing proportion: 0.500000
Mean: 24.1283 26.6116
Real Mean: 77.18 85.15

2014-15 2016-17

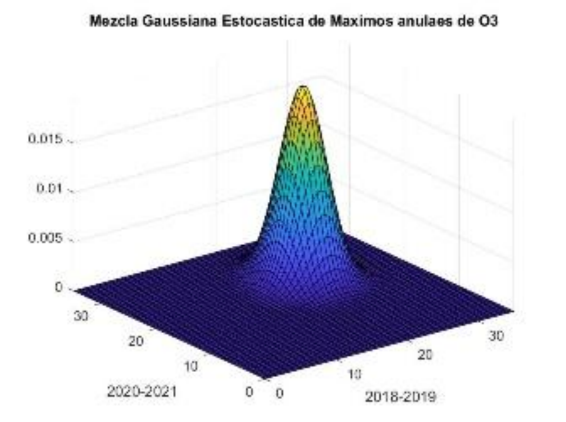


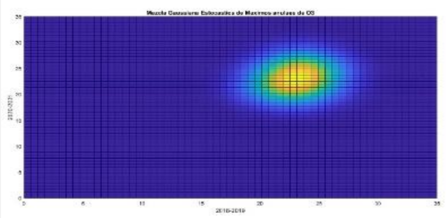


Component 1:
Mixing proportion: 0.500000
2014-2015
Mean: 25.0293 22.7009
Real Mean: 80 72.64

Component 2:
Mixing proportion: 0.500000
2016-2017
Mean: 25.5082 25.6616
Real Mean: 81.6 82.11

2018-19 2020-21

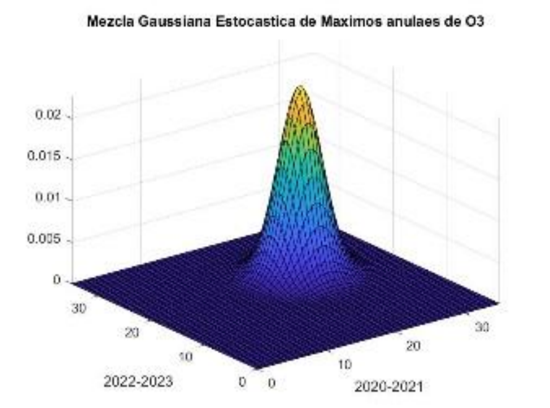


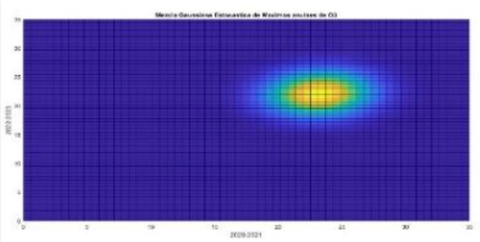


Component 1:
Mixing proportion: 0.500000
2018-2019
Mean: 23.2529 24.5596
Real Mean: 75 80

Component 2:
2020-2021
Mixing proportion: 0.500000
Mean: 22.2855 21.0674
Real Mean: 71 65

2020-21 2022-23





Component 1:
Mixing proportion: 0.500000
2020-2021
Mean: 22.2855 21.0674
Real Mean: 71.29 70

Component 2:
2022-2023
Mixing proportion: 0.500000
Mean: 23.5835 22.6229
Real MeanReal: 84.8 82

Gaussian mixture of n components

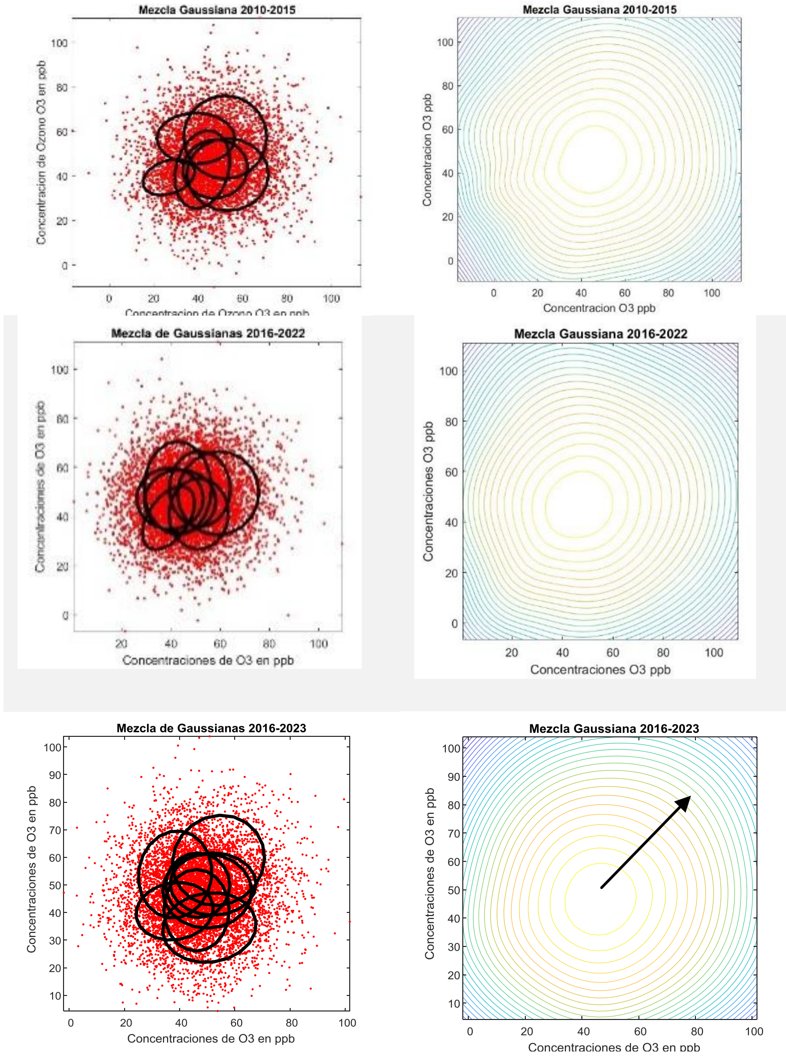


Figure (7)

In these results with this Gaussian mixture of various components for years, we can see a slight elevation pattern, but very slight and it is seen in the slightly more intense colors in the 2016-23 mixture, but with a downward trend of Ozone concentrations

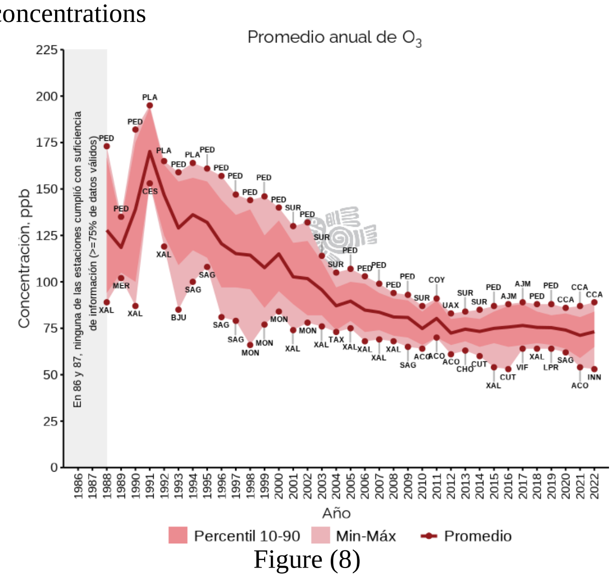


Figure (8)

In the official Ozone trend figure for Mexico City we can see that it only reaches 2022, it is not updated and is the one on the official portal, but we can see that it coincides with that slight trend shown in the Gaussian mixture of n components.

Maximum Daily Temperatures in México City.

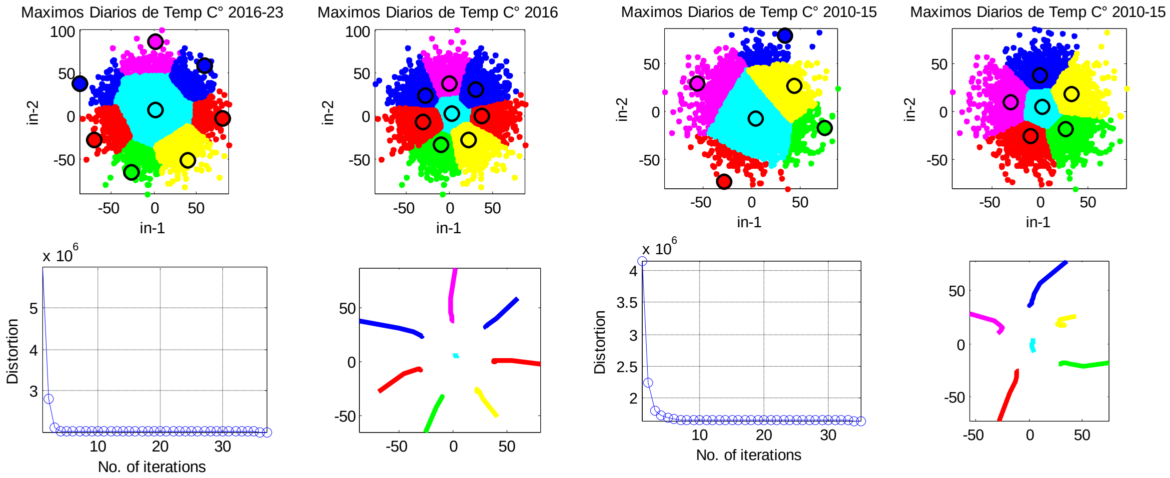


Figure (9)

By Clusters for Daily Maximum Temperatures in CDMX

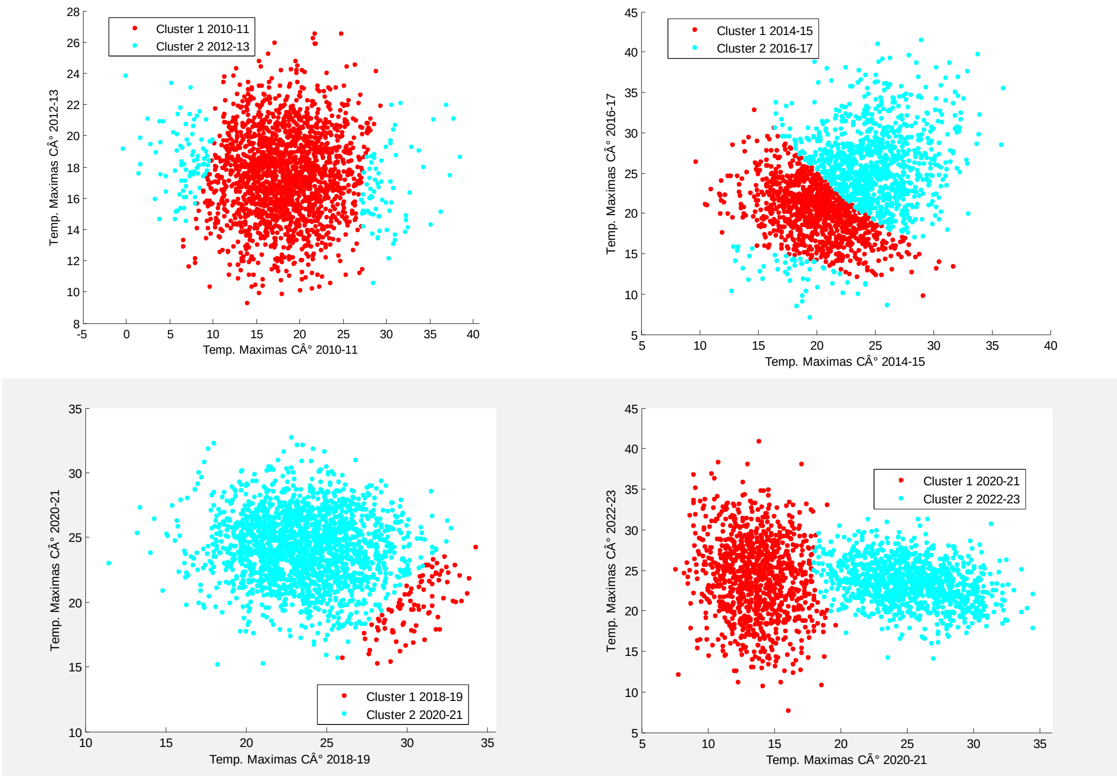


Figure (10)

By Cluster by Pairs of years

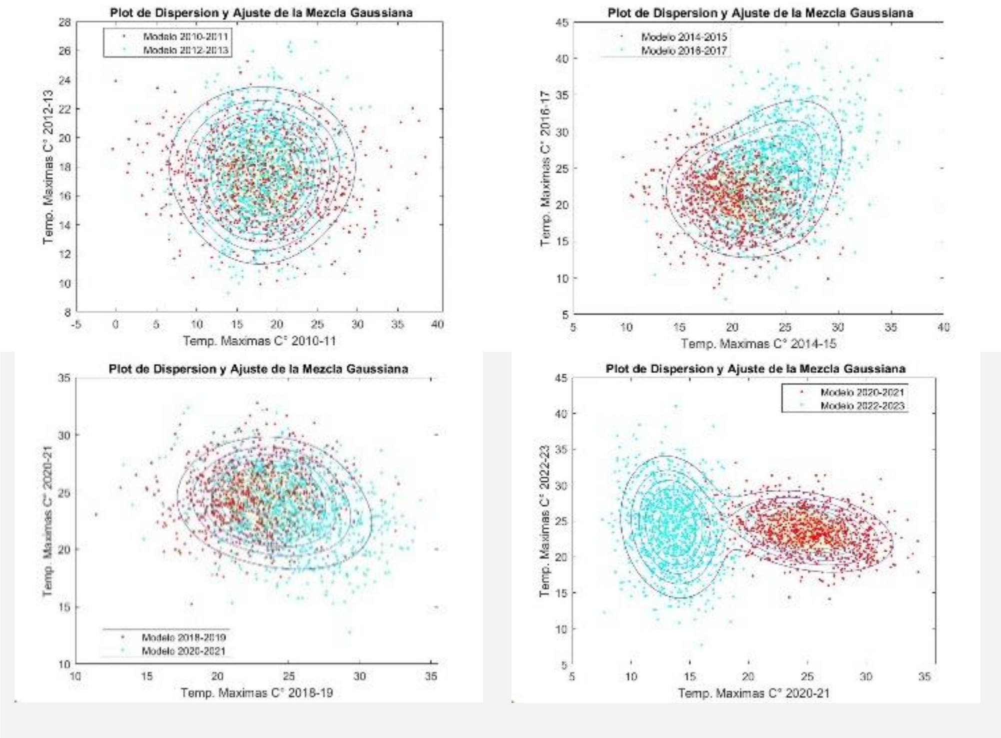
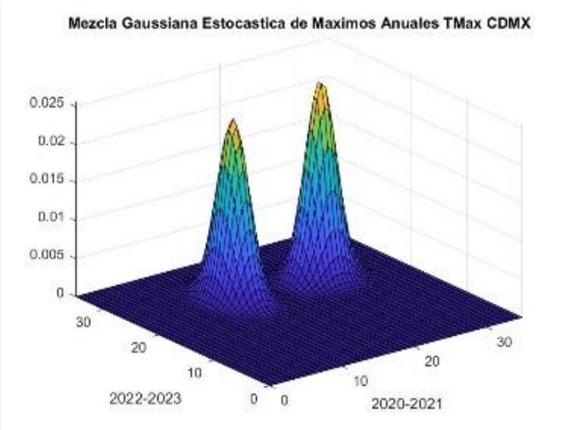


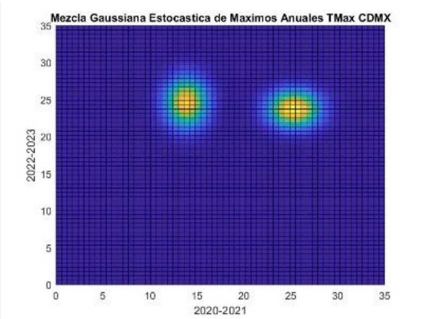
Figure (11)

Gaussian Mixture and its Scatter Plot

Table 4. 2D Gaussian Mixture of Daily Temperature Maximums

2020-21 2022-23

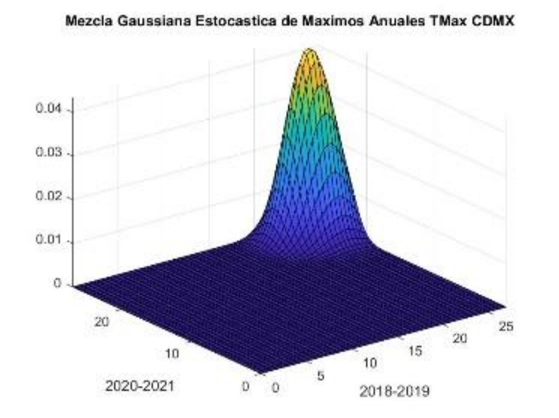


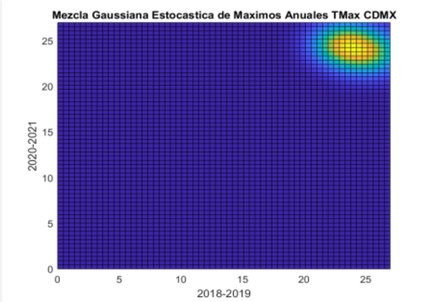


Component 1:
Mixing proportion: 0.500000
2020-2021
Mean: 25.0684 23.4198

Component 2:
Mixingproportion: 0.500000
2022-2023
Mean: 13.5900 24.3035

2018-19 2020-21

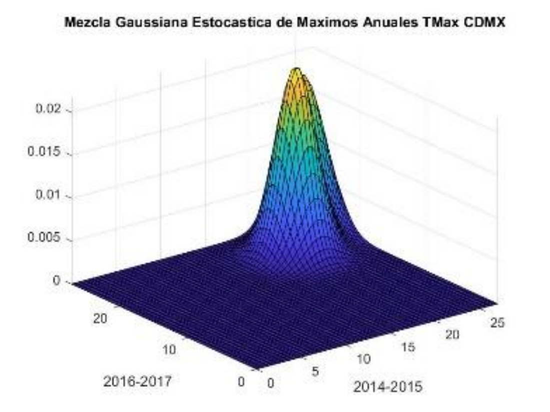


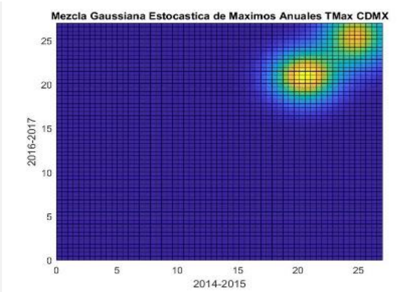


Component 1:
Mixing proportion: 0.500000
2018-2019
Mean: 22.8712 24.4113

Component 2:
2020-2021
Mixing proportion: 0.500000
Mean: 25.0684 23.4198

2014-15 2016-17

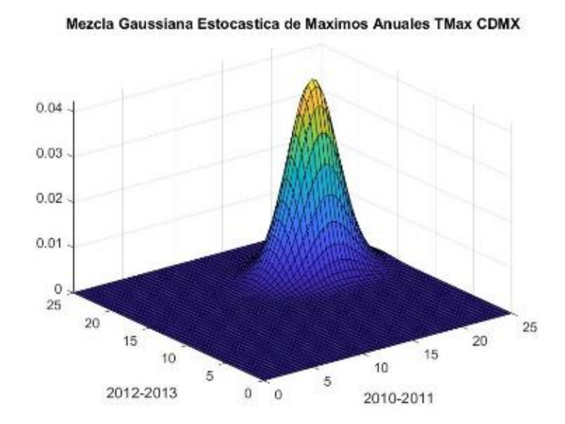


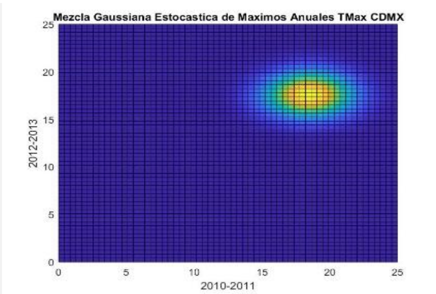


Component 1:
Mixing proportion: 0.500000
2014-2015
Mean: 20.3195 20.7225

Component 2:
Mixing proportion: 0.500000
2016-2017
Mean: 24.5593 25.2841

2010-11 2012-13





Component 1:
Mixing proportion: 0.500000
2010-2011
Mean: 18.1940 17.0605

Component 2:
Mixing proportion: 0.500000
2012-2013
Mean: 18.3277 17.7526

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 10 - Issue 04 || April 2024 || PP. 01-15

Gaussian mixture of n components

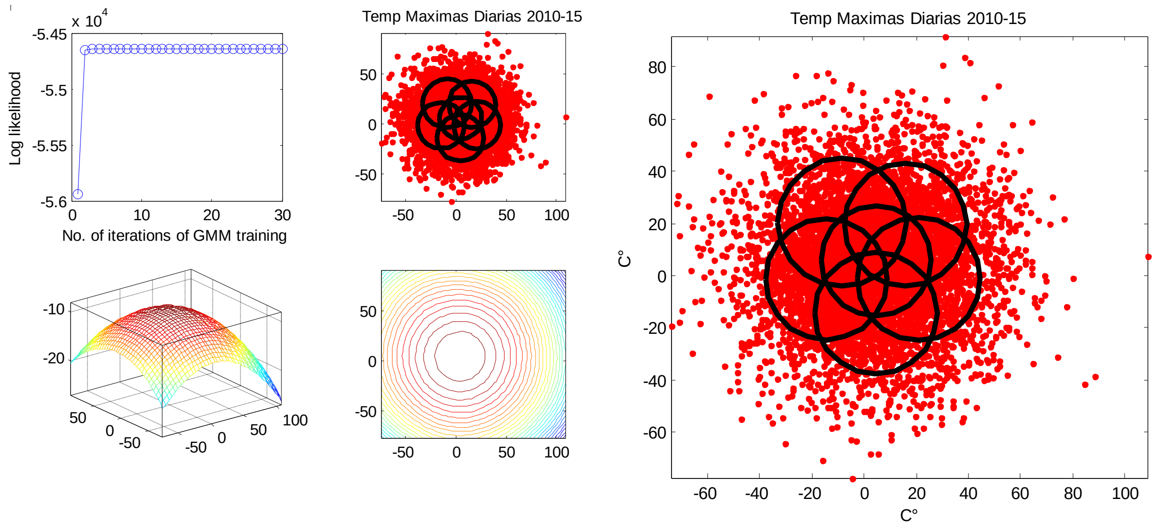


Figure (12)

GMM (Gaussian mixture model) for 2-D "Unequal" data. Number of Gaussians is 6, with K means = 6

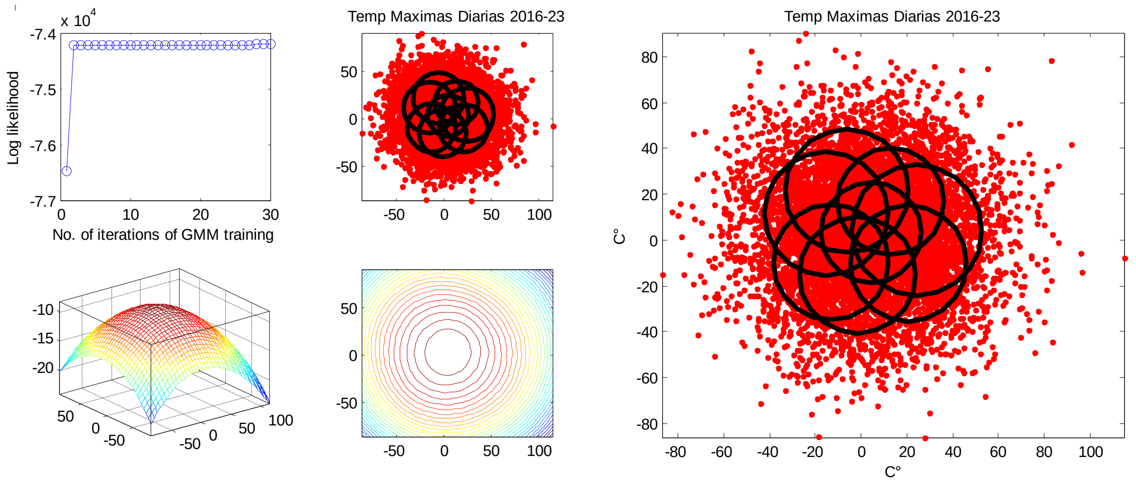


Figure (13)

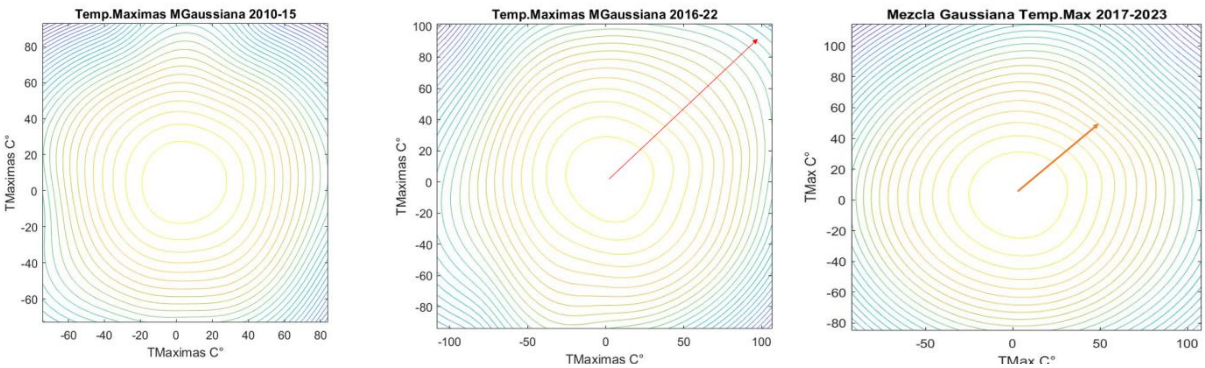


Figure (14)

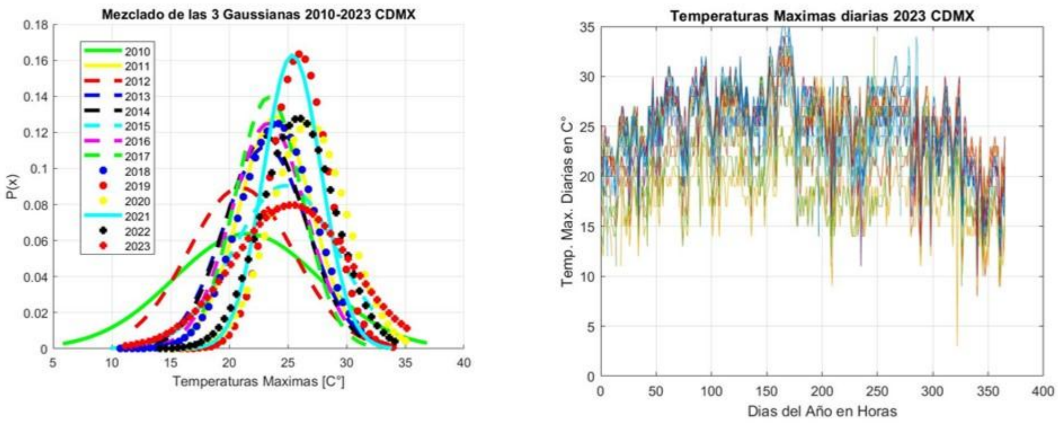


Figure (15)

The average Maximum Temperature until June 2023 is 26.12 C° of the 1D Gaussian Mixture and the CONAGUA of México is 25.8 C° and now the Maximum average Temperature until November 12, 2023 is 25.34 C° and until October the CONAGUA is 25.5 C° on average, now that of the full year of 2023 is 25.3 C° of the 1D Gaussian Mixing and that of CONAGUA until November 30 is 25.4 C°

Conclusions

It can be concluded with this more complete analysis that the methodology gives us the behavior or trend of the maximum ozone concentrations of each year taken from the database of the Air Secretary of Mexico City, as well as the temperatures. Daily maximums When starting with the cluster analysis, when making the groups, a certain symmetry is observed, then when taking the clusters by pairs of years, a similar group pattern is observed in the daily ozone maximums except in the 2021-22 clusters and 2022-23 are totally united without distinction and in the clusters by pairs of years of the daily maximum temperatures in the years 2021-22 and 2022-23 totally separated except for a slight union, now in the Gaussian mixture you can see the patterns of daily maximum temperatures, in the 2021-22 and 2022-23 mixture of maximum daily ozone concentrations is a single Gaussian with a slight pattern to the right of the x-axis, having a slight rebound beyond 85 ppb, but In general I feel the downward trend and showing average values approximately coinciding with the official trend graph.

Now, by further combining the mixture of several Gaussians per year for ozone, a slight trend is shown, shown with the arrow on the image, but it is very slight. For daily maximum temperatures, this effect is seen a little more and with the mixture of With the 3 Gaussians you can see the trend of the last one in 2023 a little better, Mexico City is also a heat island and more like the effect of the child that hovers over the entire globe, this effect of high temperatures increases a little more plus the heat waves that have been observed.

We can say that there is already a background concentration of ozone in Mexico City which perhaps the effect could be observed more or those maximum daily temperatures could be seen with a more geostatistical treatment.

There are also the days of environmental contingencies due to ozone for Mexico City in 2023 and so far in 2024, with a very similar count so far.

Índice de calidad del aire mayores a 150					
Fecha	NOO3	NEO3	CEO3	SOO3	SEO3
2023-02-23	124	123	151	167	142
2023-03-25	157	131	137	132	125
2023-03-26	136	130	143	153	122
2023-11-20	123	154	104	138	49

Índice de calidad del aire mayores a 150					
Fecha	NOO3	NEO3	CEO3	SOO3	SEO3
2024-02-22	109	127	161	172	172
2024-02-23	131	183	134	138	129
2024-02-24	120	133	162	140	136
2024-03-06	175	164	121	119	90
2024-03-23	158	151	136	105	69

References

[1]. Jyh-Shing Roger Jang, "Machine Learning Toolbox", available at "http://mirlab.org/jang/matlab/toolbox/machineLearning", accessed on [date]. http://mirlab.org/jang/books/dcpr/ Data Clustering and Pattern Recognition

[2]. Li, Z. Applications of Gaussian Mixture Model to Weather Observations. IEEE Geoscience and Remote Sensing Letters (Volume: 8, Issue: 6, Nov. 2011) McLachlan, G. & Peel, D. Finite Mixture Models.

[3]. 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.

[4]. Robert, C. P., The Bayesian Choice: A Decision-Theoretic Motivation, Springer Ser. Stat., Springer Verlag, New York, 1994.

[5]. Otten, A., and M. A. J. Van Montfort, Maximum-likelihood estimation of the general extreme-value distribution parameters, J. Hydrol., 47, 187–192, 1980.

[6]. Zenteno Jimenez Jose Roberto International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031 www.ijlret.com || Volume 07 - Issue 03 || March 2021 || PP. 01-17 http://www.ijlret.com/Papers/Vol-07-issue-03/1.B2021164.pdf An Analysis of the Relationship between Maximum Daily Temperatures and Maximum Ozone Concentrations in México City 2010 - 2020. A Bivariate Approach

[7]. Zenteno Jimenez Jose Roberto International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031 www.ijlret.com || Volume 05 - Issue 06 || June 2019 || PP. 36-54 Analysis of the Trend Annual Maximum Temperatures in Mexico City 2005 - 2018 and with the WRF Program Case Two: Daily Average Maximum Temperature Data, with Gaussian behavior

[8]. Zenteno Jimenez Jose Roberto International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031 www.ijlret.com || Volume 08 - Issue 08 || August 2022 || PP. 01-20 Multivariate and Mixed Gaussian Analysis of the Maxima of Ozone, Temperatures, Nitrogen Dioxides and Total Suspended Particulate from 2019-2022 in México City http://www.ijlret.com/Papers/Vol-08-issue-08/1.B2022274.pdf