

Statistical analysis for the characterization of the wastewater in the influent of a treatment plant (Case of study)

Facundo Cortés-Martínez, Alejandro Treviño-Cansino, Agustín Sáenz-López, Rajeswari Narayanasamy

Abstract—A problem facing wastewater treatment systems is to identify the discharges of wastewater from the industry and classify the concentration of pollutants in it that are received in the treatment plant, weak, medium or strong. So a statistical analysis of the concentration of pollutants in the influent, can guide the operators of these systems to control decisions. Statistical analysis was carried in a treatment plant and the concentration data was considered in the influent from an external source. The following parameters were analyzed: biochemical oxygen demand and chemical oxygen demand. According to the results, Wastewater was classified as middle class, and some industrial wastewater discharges could be identified. The criteria to identify the possible infringing users has been included.

Index Terms— Statistical analysis, discharge control, sewage, industrial discharges, biochemical oxygen demand.

I. INTRODUCTION

The information that is generated by the characterization of wastewater in the influent of a municipal treatment plant provides essential information like the concentration tendency of some or one specific pollutants. According to The Environmental Protection Agency (EPA) of United States of America. The discharging regulation to commercial, industrial, and service sewers systems is a total challenge for authorities in charge, because of its scattering, and the needed hard work to control them. (EPA, 2002; 2003; Wills *et al.*, 2010). Therefore, the quoted agency recommends to identify, track, and control the non-domestic wastewater discharges; that is, to prevent possible interference in biological treating at the municipal treatment plant (EPA, 1987; 1991, UNAM, 2000). A consequence of high pollutants content, like biochemical oxygen demand (BOD), is that treated wastewater does not meet the quality regulation for pouring receptive bodies NOM-001-ECOL-1996 (DOF, 1997). This represents an economical problem for the plant, the town

council, and the water operator corporation.

Some possible relevant problems that can affect the municipal pipes, due to the fact that pouring pollutants higher than the indicated by the norm, are: a temperature increase over 40 degrees in wastewater moving through pipes systems, fat and solid blocking pipes, and the risk of explosions in sewers systems, among others. (DOF, 1998; EPA, 1987; CNA e IMTA, 2000; 2007).

Two parameters are being analyzed in this project: Biochemical oxygen demand (BOD) and chemical oxygen demand (COD). BOD is the among of dissolver oxygen that microorganisms require to oxidize organic substance. The process lasts 5 days, for that reason the parameter is indicated as BOD₅ (Crites and Tchobanoglous, 2000; CNA e IMTA, 2000). COD measures organic substance in wastewater as an indirect way; due to dichromatic potassium is used for oxidation. The bonds of this parameter are generally higher than BOD₅, since COD oxidizes any kind of substance; while BOD₅ oxidizes only those ones which can be biologically degraded (Metcalf & Eddy, 1991; CNA e IMTA, 2000).

The main aim of this article was to apply the descriptive (non parametrical) statistics to concentration data BOD₅, COD in the influent of the treatment plant, all this with the purpose of classifying the state of the wastewater and identifying the possible discharges of it within industrial process.

The contribution of this document was the analysis criteria and the results interpretation of the statistical analysis, applied to the characterization of wastewater in the treatment plant.

II. METHODOLOGY

A. Classification of wastewater.

According to Metcalf & Eddy (1991) the typical concentration of wastewater is composed by three states: weak, medium, and strong. On table 1, a segment of the parameters and concentrations is being shown.

Table 1. Composition of gross domestic wastewater. Metcalf & Eddy (1991).

Pollutants	Units	Concentration		
		Weak	Medium	Strong
Biochemical oxygen Demand 5 days, 20 °C (BOD ₅)	mg/L	110	220	400
Chemical Oxygen Demand (COD)	mg/L	250	500	1000
Settleable solids	mL/L	5	10	20

Facundo Cortés-Martínez, Faculty of Engineering, Science and Architecture Juárez University of Durango State, Mexico. (FICA-UJED)
Alejandro Treviño-Cansino, FICA-UJED, México
Agustín Sáenz López, FICA-UJED, México.
Rajeswari Narayanasamy, FICA-UJED, México

Next, the basic nomenclature of descriptive statistics is being described. The mathematical expressions and nomenclature were taken from Pérez (2002) and Guarín (n.d.).

Statistical analysis for the characterization of the wastewater in the influent of a treatment plant (Case of Study)

B. Nomenclature

\bar{X}	=	Arithmetic mean or average
x_i	=	Values of the variable X
n	=	Number of observations
Σ	=	Sign sum Σ
f_n	=	Frequency
Me	=	Median
LI	=	Lower limit of the range where the median is placed
$fa_{(i-1)}$	=	Cumulative frequency preceding the median interval
f_i	=	Median frequency range
A	=	Crest factor
k	=	Order quartile = k = 1, 2, 3
$L1$	=	Lower limit of the range that contains the Quartile
R	=	Range
X_{max}	=	Maximum value of the variable
X_{min}	=	Minimum value of the variable
g_1	=	Fisher asymmetry coefficient
g_2	=	Fisher kurtosis coefficient
L_i	=	Lower limit of the modal interval
f_m	=	Frequency of the modal class
$f(m-1)$	=	Frequency of premodal class
$f(m+1)$	=	Frequency of posmodal class
$fa(i-1)$	=	Cumulative frequency to the pre containing quartile range
N_i	=	Absolute cumulative frequency.

F_i	=	Cumulative relative frequency.
f_R	=	Relative frequency.
n_i	=	Absolute frequency..
m_3	=	Third order moment with regard to the average
S_x^3	=	Cubed Standard Deviation
m_4	=	Fourth order moment with regard to the average

C. Frequency distributions.

In statistics, frequency is often referred to as the number of times you repeat a variable, also called absolute frequency (n_i). After being divided by the total of the observations, it is called relative frequency (f_R).

$$f_R = \frac{n_i}{N} \quad (1)$$

The absolute cumulative frequency allows to know the number of cases that are located below a certain value (N_i).

$$N_i = \sum n_i \text{ with } J = 1 \dots i. \quad (2)$$

The relative cumulative frequency refers to the cumulative absolute frequency divided by the total number of values of the variable under study.

$$F_i = \frac{N_i}{N} \quad (3)$$

Table 2 shows the way the frequency distribution is usually presented.

Table 2. Table of frequencies.

I_i	x_i	n_i	f_R	N_i	F_i
$[L_0 L_1]$	x_1	n_1	$f_1 = \frac{n_1}{N}$	$N_1 = n_1$	$F_1 = \frac{N_1}{N}$
$[L_1 L_2]$	x_2	n_2	$f_2 = \frac{n_2}{N}$	$N_2 = n_1 + n_2$	$F_2 = \frac{N_2}{N}$
$[L_2 L_3]$	x_3	n_3	$f_3 = \frac{n_3}{N}$	$N_3 = n_1 + n_2 + n_3$	$F_3 = \frac{N_3}{N}$
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
$[L_{k-1} L_k]$	x_k	n_k	$f_k = \frac{n_k}{N}$	$N_k = n_1 + \dots + n_k$	$F_k = \frac{N_k}{N}$
		$\sum n_i = N$	$\sum f_R = 1$		

Source: Pérez (2002).

A histogram includes the variable frequencies in intervals where the area of the rectangle is proportional to the frequency interval in question. According to Pérez (2002), a frequency histogram is the representation of data which can be defined in three important properties of the distribution:

shape, central tendency and dispersion.

$$R = X_{\max} - X_{\min} \quad (8)$$

D. Measures of central tendency or position

The main measures of central tendency are: the arithmetic mean or average, median and mode.

The arithmetic mean is the sum of all divided by the number of data values. Some of its properties are: all variables are involved; the value is unique and is interpreted as a balance. In equation (4) the expression for grouped variables is shown.

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_i f_i + \dots + x_m f_m}{n} = \frac{\sum_{i=1}^m x_i f_i}{n} \quad (4)$$

Median is the value of the variable considered below half of the data, the other half is located above, this measure is used in nonparametric statistics. Some properties are: it is less sensitive to outliers than the average and dispersion does not affect the value: it is considered more real than the arithmetic mean. The information is grouped in equal intervals, then the median is calculated using the following expression:

$$Me = LI + \frac{\frac{n}{2} - fa(i-1)}{fi} A \quad (5)$$

Trimmed mean: is a more robust measure, as it is less sensitive to outliers. It deletes a number of observations, both above and below the variable under study.

When data is grouped into intervals of equal size, mode is calculated using the following expression.

$$Mo = LI + \frac{f_m - f_{(m-1)}}{2f_m - f_{(m-1)} - f_{(m+1)}} A \quad (6)$$

Quantiles, percentiles, quartiles and deciles. It is the dataset but formed into groups with the same number of elements: In the case of quartiles variable is divided into four groups with the same number of data.

$$Q = L_{\inf} + \frac{\frac{kn}{4} fa(i-1)}{f_i} A \quad (7)$$

E. Measures of dispersion

They indicate the concentration of data with regard to measures of centralization. They are divided into: variance, standard deviation, coefficient of variation and range. In the present paper only the range will be used, as this study is only an exploratory analysis of the original data without transforming it (non-parametric statistics). The range refers to the difference between the highest and lowest number of distribution.

The interquartile range is the difference between the third and first quartiles: in this range, 50 percent of the data is included.

$$Range_{IQ} = Q_3 - Q_1 \quad (9)$$

The box plots show median, interquartile range, and outliers of the variable under study. The lower edge of the box corresponding to the first quartile (25 percent); while the higher corresponds to 75 percent. To set much lower limits outliers as top of the box are determined. To accomplish the foregoing is considered the breadth of the box, that is, the interquartile range. The first limit is obtained as 1.5 times the IQ Range, while the second limit is set 3 times the breadth of the quoted range.

F. Skewness and kurtosis

Fisher coefficient is a measure of asymmetry, which analyzes the proximity of the data average (\bar{X}). Therefore if the coefficient of Fisher ($g_1 = 0$), the distribution will be symmetrical if $g_1 < 0$ asymmetric negative (left), and if $g_1 > 0$ asymmetric distribution is positive (right). To analyze the asymmetry coefficient is first necessary to calculate a statistic known as time of order three with respect to the mean (m_3). asymmetry coefficient is determined by expression (10).

$$g_1 = \frac{m_3}{S_x^3} \quad (10)$$

Kurtosis evaluates the distribution frequency in the central region with regard to the normal curve. It is said that a distribution is mesokurtic (equal to the normal curve) when the kurtosis, $g_2 = 0$; be leptokurtic (pointing higher than the normal curve) where $g_2 > 0$; and platykurtic (pointing lower than the normal curve), where $g_2 < 0$. The kurtosis is determined using equation (11).

$$g_2 = \frac{m_4}{S_x^4} \quad (11)$$

The measurement data of wastewater was taken from an external source: Lichman (2013), which were also used in the following studies: Belanche et al (1992); Garcia (1993) and Bejar et al (1993). The measurements were only taken from the influent in a treatment system: 1046 measures. Due to the volume of information, it is not included in this document; however it can be verified in the already quoted resource

III. RESULTS AND DISCUSSION

A. Biochemical oxygen demand (BOD)

In figure 1 the frequency histogram is presented.

Statistical analysis for the characterization of the wastewater in the influent of a treatment plant (Case of Study)

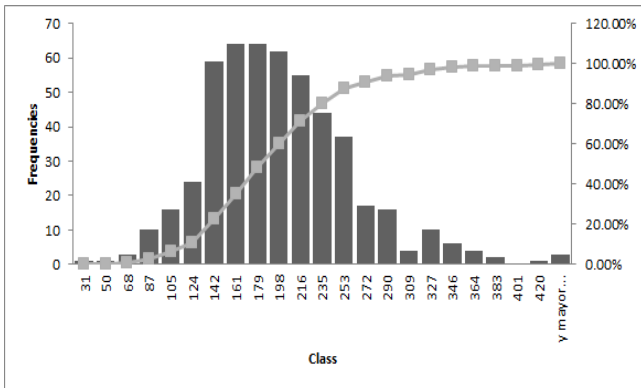


Figure 1. Histogram to BOD₅ in the influent of the treatment plant.

Class mark represents the concentration of organic matter in the range of 19 mg / L.

Shape measures: the asymmetry of the distribution was positive: 0.78 versus normal zero. According to Figure 1 the left side plummets; while right side goes down gently. The tail of the distribution is larger with values above average. 1.36 kurtosis proved less than 3 so the distribution is platykurtic. The five classes in the peak are the concentrations of organic matter appearing more frequently.

Position measurements: the statistics calculated by the SPSS program were: with a confidence of 95 percent is estimated that the values are between 183 and 194 upper and lower limits respectively, average is 189, trimmed 5 percent mean equals to 186, Medium is 183, and fashion turned out to be 133. The relationship between mean and median for positive asymmetry is that the average is greater than the median. It is noted that the trimmed mean is the closest to the arithmetic median.

Scattering measurements: In Figure 2 the dispersion of data BOD₅ is shown with a minimum value of 31 mg / L and a maximum of 438 mg / L. The rank is very wide: 407 mg / L.

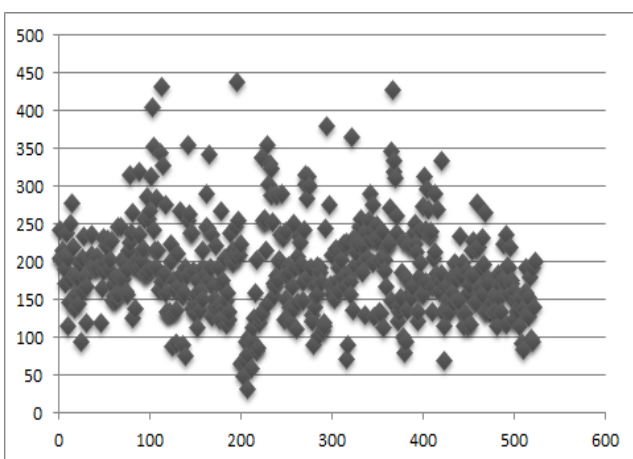


Figure 2. Dispersion of BOD₅ in the influent of the treatment plant.

According to Figure 2 some measurements that at first glance are low and high values are observed. In order to identify outliers and extremes, in Figure 3 the box and whisker plot is presented for BOD₅.

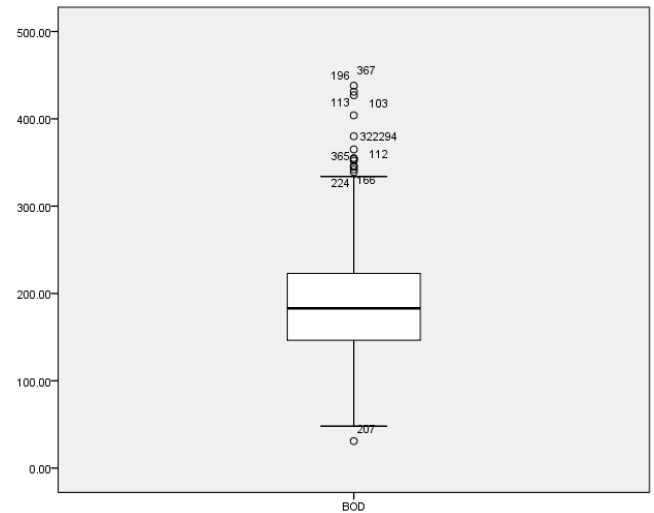


Figure 3. Box diagram of BOD₅ in the influent of the treatment plant.

Results quartiles: Q₁ = 146 corresponds to the bottom of the box 183, Q₂ = 183 or medium and Q₃ = 223 top of the box. The interquartile range = 77, lower limit = 30.5 whisker is rounded to 31 and higher = 338.5 is rounded to 339. According Quevedo and Perez (2008) the interquartile range represents the 50 percent dispersion of the core data. The upper outliers seen in Figure 3 are far more than 1.5 lengths of the third quartile. These outliers suggest a positive skewness of the distribution. A displaced median center of the box as shown in Figure 3 shows also positive skewness. The lowest and highest values that resulted from the analysis are shown in Table 3.

Table 3. Extreme values for BOD₅

		Case Number	Value	
BOD	Higher	1	196	438.00
		2	113	431.00
		3	367	427.00
		4	103	404.00
		5	294	380.00
	lower	1	207	31.00
		2	203	48.00
		3	212	58.00
		4	201	64.00
		5	202	66.00

All major values given in Table 3 were above the upper limit determined in quartile analysis: 339. Therefore they are considered as outliers; while lower values are located below the weak concentration shown in Table 1; that is less than 110 mg / L. Only the case 207 was similar to the lower limit, although the box and whisker plot indicates that it is out of the limit. All this because the value was 30.5 but it was rounded up to 31. The last situation is not a problem for a biological treatment system. In the same table 1, top concentration values are observed.

The statistical literature mentions that it is wise to conduct a study before removing outliers. Since they are not eliminated, the conclusions may be wrong, or the results could also be deformed. In the present study we observed atypical

measurements, then it was applied a connection to BOD₅ and COD.

According to Table 1, the value of organic matter concentration average for wastewater is 220 mg / L. The average BOD₅ was 189, this value is located between the weak and mean value, namely 110 and 220. But the value of the average is closer to 220 to 110; therefore it is classified as medium concentration for organic matter, although Figure 2 shows values from below weak concentration to above strong concentration.

B. Chemical Oxygen Demand

Figure 4 shows the histogram.

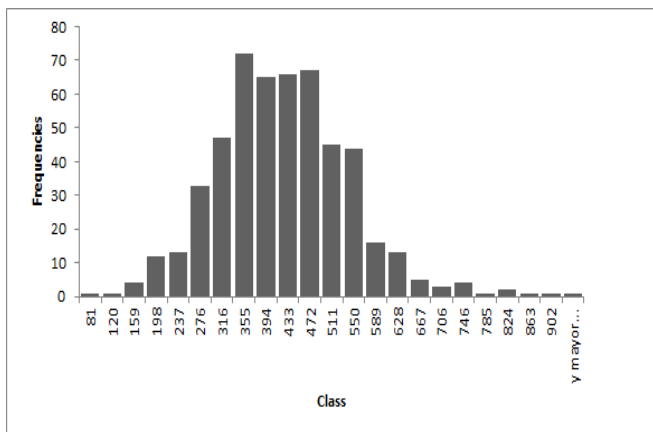


Figure 4. Histogram for COD in the influent of the treatment plant.

Shape measures: The asymmetry of the distribution was positive: 0.58. According to Figure 4 the distributions of the tail is larger with values above average. 4 peaks can be observed and 355 resulted the more frequently class, followed by 472, 433 and 394. The kurtosis was platykurtic.

Position measures: defined statistical were the following; with a confidence interval of 95 percent is estimated that the values are located between 396 lower limit and 417 upper limit. 407 average, Trimmed Mean at 5 percent 403; for outliers. Median 398, mode 380. As the BOD₅ the Trimmed Mean is closer to the arithmetic median.

In order to calculate in an approximate way the measurements number for each classification according to Table 1. The cumulative frequency percentage is displayed.

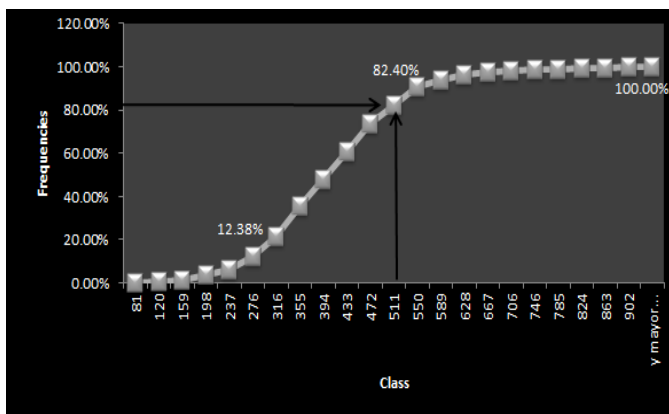


Figure 5. Accumulated Frequencies Graph in percentage of the COD in the influent of the treatment plant.

According to Figure 5 approximately 12.38 percent of the measurements are less than the weak concentration: 65; while 82.40 percent are less than the average 431. 100 percent of the values were lower than the maximum concentration shown in Table 1.

Another observation is that the percentage difference between the weak and average concentration resulted on 70.02 percent: 366 values. This confirms that most measurements are of the average concentration type (Table 1). While in 17.6 percent: 92 were between average and high concentration.

Dispersion Measures: in Figure 6 the dispersion of the data of COD with a minimum value of 81 mg / L, and maximum value of 941 is shown. As well as the BOD₅ variability is also very wide: 860 mg/L.

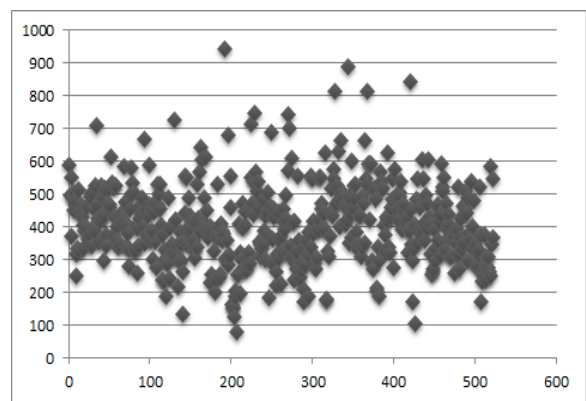


Figure 6. Dispersion of COD on the treatment plant influent.

In order to identify atypical measurements and outliers, the box and whisker plot is presented in Figure 7.

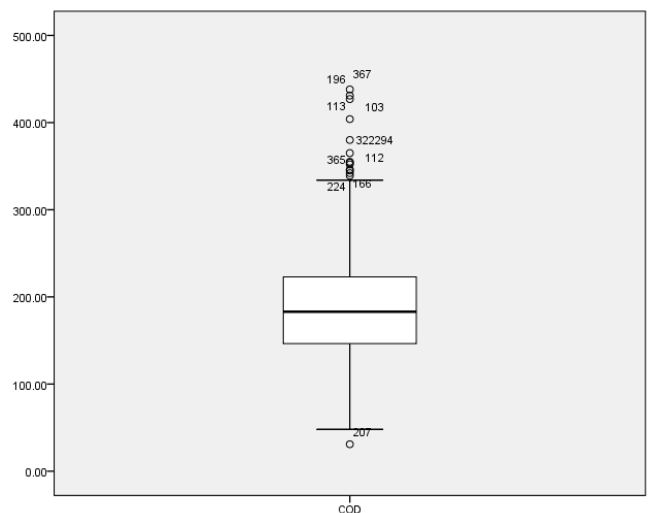


Figure 7. Box and whisker diagram for COD in the treatment plant influent.

Quartiles results were: $Q_1 = 325$, $Q_2 = 398$ y $Q_3 = 478$. The interquartile range = 153, lower limit whisker = 96 and upper limit whisker = 708. In Figure 7 can be observed that only the 207 case was below the lower limit. The lowest and

Statistical analysis for the characterization of the wastewater in the influent of a treatment plant (Case of Study)

highest values that resulted from the analysis are shown in Table 4.

Table 4. Outliers values for COD

		Case Number	Value
COD	higher	1	192
		2	344
		3	420
		4	367
		5	328
	lower	1	207
		2	426
		3	204
		4	140
		5	202

The higher values shown in table 4 were above the upper limit defined in the quartiles: 708 so they are considered atypical and outlier measures.

The lower values shown in table 4 were below the weak concentration: 250 mg/L. which is indicated in Table 1.

Due to atypical and outlier values were observed, the arithmetic mean does not represent an actual value. Therefore, the Trimmed Mean at 5 percent was considerate: 403 mg/L. The classification resulted average according to Table 1

C. Relationship between BOD₅ / COD

According to Metcalf & Eddy (1991) and CNA and IMTA (2000) there is a relationship between the BOD₅/ COD which varies between 0.4 and 0.8. The range of these values indicates that it is wastewater of the domestic type; while for industrial waters this ratio is greater.

Applying to relationship to the data the following results were obtained: average 0.48, median 0.45 and 0.50 mode. The condition for a positive asymmetric distribution, as previously indicated is media > median. The minimum value was 0.17, maximum value 1.27 and 1.10 range. Figure 8 shows the results.

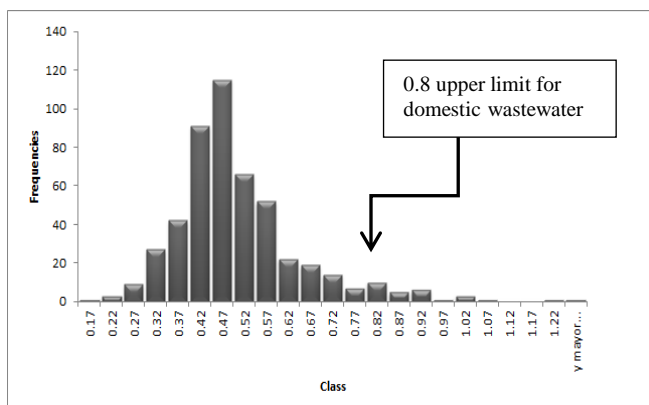


Figure 8. Histogram ratio of BOD₅/COD in the treatment plant influent

As observed on the class mark ("X" axis) Most measurements indicate that wastewater are of domestic type, since the ratio ranges BOD₅/COD resulted from 0.17 Although values exceeding 0.8 are included in the right side of the histogram . The Results match with the atypical observations shown in Figures 3 and 6. This suggests that atypical measurements are rather isolated wastewater discharges from industrial processes into the drainage system, and therefore, it is prudent to establish a discharge control of commercial, industrial and services wastewater. The mentioned above, in order to protect the municipal pipelines system and the operation of the treatment plant (EPA, 1987; CNA and IMTA, 2000; UNAM, 2000).

To identify the number of measurements of industrial wastewater, in figure 9 cumulative frequency percentage is shown.

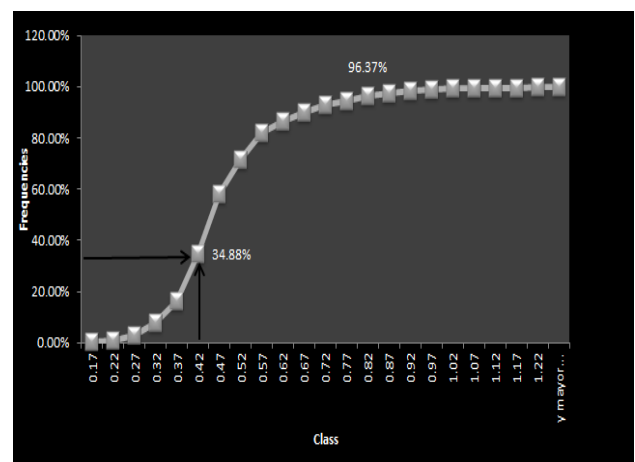


Figure 9 cumulative frequency graph in BOD₅ / COD ratio percentage of the treatment plant influent.

According to Figure 9 a little less than 34.88 percent of the measurements are less than 0.4: 182. While the 96.37 percent: about 504 are less than the upper limit for domestic wastewater. 3.63 percent resulted measurements above 0.8. As mentioned before, this last percentage suggests that industrial process wastewater is downloaded to the drainage system. This coincides with the analysis made between BOD₅ and COD parameters.

C. Identification of possible infringing users

A criterion to identify inferring users is determined by the extinct Secretariat of Commerce and Industrial Development. This branch carried out a categorization of users regarding the types of pollutants generated. That aforementioned document was called: Mexican Classification of Activities and Products. This was published by the National Water Commission (CNA and IMTA, 2000). Table 6 shows a segment of the mentioned table.

Table 6. Industries that generate high levels of contaminants. Adapted UAM (1997) and EPA (1987) cited (CNA e IMTA, 2000).

Process	Descripción	EC	HP	F and O	SS	TSS	BOD	COD
3111	Meat industry			X		X	X	X
3112	Manufacture of dairy products			X		X	X	X
3113	Canned food processing	O		X	X		X	X
3114	Grain milling			X	X	X		
3115	Bread making			X	X			X
3116	Making tortillas	O	O		X	X	X	X
3117	Edible oils and fats			O				X
3118	Sugar industry	O	O			X	X	X
3119	chocolat manufacturing			X	X	X	X	X
3120	Food products			X	X	X	X	X
3121	Animal feed			X		X	X	X
3122	Beverage industry	O		X		X	X	X

Table 6 industries that generate pollutants, among those in which BOD₅ and COD are observed. This information can be used as an initial assessment of the potential users that discharge pollutants above the maximum limits permitted by the standard. Consequently, the water operator corporation of the city can start with the inspection of these industries.

Then propose a treatment system where the generation of pollutants is located, so that it meets the concentration indicating the standard of wastewater discharge to municipal sewer systems: NOM -002- ECOL - 1996 published in the Official Journal of the Federation (DOF).

This indicates 200 mg / L., for BOD₅ daily average for the discharge to the sewer system (DOF, 1988) Table 7 shows the summary of results of the analysis of raw wastewater in the treatment plant influent.

Table 7. Summary of wastewater classification results.

Concept	BOD ₅	COD
Classification of typical wastewater	Average with industrial discharges	Average with industrial discharges

ACKNOWLEDGE

The author would like to thank the strengthening program for quality in educational institutions (PROFOCIE) 2014 for its support in the realization of this research.

IV. CONCLUSION

Descriptive statistics were applied to the data; analyzed pollutants were BOD₅, COD in the influent to the treatment plant. Wastewater state was then determined according to the classification shown in Table 1.

Possible discharges of industrial wastewater were identified as well as processes with high organic matter content.

The application of basic statistical to the characterization of wastewater yields important results, so that those responsible

for the operation of the treatment system may recommend to the water utility, the types of industries and processes to be monitored. In order to establish a control of pollutants prior to discharge into the municipal sewage. Performing a wide database of the characterizations in the influent wastewater treatment system, as well as analytical results provide a basis for future comparisons. In order to identify significant deviations in the concentration of pollutants. With this, it is possible to detect potential offenders.

REFERENCES

- [1] Béjar, J., Cortés, U., & Poch, M. (1993). Linneo+: a classification methodology for illstructured domains. Facultat d'Informtica de Barcelona, Tech. Rep.
- [2] Belanche, L., Sánchez, M., Cortés, U., & Serra, P. (1992). A knowledge-based system for the diagnosis of waste-water treatment plants. In *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems* (pp. 324-336). Springer Berlin Heidelberg.
- [3] Comisión Nacional del Agua. (2000). Guía para el control de descargas a los sistemas de alcantarillado urbano o municipal. México.
- [4] Comisión Nacional del Agua. (2007). Guía para el control de descargas a los sistemas de alcantarillado urbano o municipal. México.
- [5] De Garcia Blanco, J. (1993). *Avaluació de tècniques de classificació per a la gestió de bioprocessos: aplicació a un reactor de fangs activats*.
- [6] Diario Oficial de la Federación. (1997). NOM-001-ECOL-1996 que establece los límites máximos permisibles de contaminantes en las descargas de aguas residuales a los sistemas en aguas y bienes nacionales. SEMARNAT. México.
- [7] Diario Oficial de la Federación. (1998). NOM-002-ECOL-1996 que establece los límites máximos permisibles de contaminantes en las descargas de aguas residuales a los sistemas de alcantarillado urbano o municipal. SEMARNAT. México.
- [8] Environmental Protection Agency. (1987). *Guidance Manual for Preventing Interference at POTWs*. USA.
- [9] Environmental Protection Agency. (1991). *Control de descargas*

Statistical analysis for the characterization of the wastewater in the influent of a treatment plant (Case of Study)

irregulares hacia las POTWs. USA.

- [10] Environmental Protection Agency (2002). Onsite wastewater treatment systems manual. United States.
- [11] Environmental Protection Agency. (2003). Voluntary national guidelines for management of onsite and clustered (decentralized) wastewater treatment systems. United States.
- [12] Guarín, S. N. (Sin fecha). Estadística aplicada. Universidad Nacional de Colombia. Medellín. Recuperado el 27 de diciembre de 2015 desde <https://estadisticaunicaes.files.wordpress.com/2012/05/curso-de-estadistica-aplicada-de-n-guarin-s.pdf>
- [13] Metcalf & Eddy, Inc. (1991). Wastewater Engineering: Treatment, Disposal, Reuse. In wastewater characteristics (pp. 53-135) New York: McGraw-Hill.
- [14] Crites, R., Tchobanoglous, G., Camargo, M., & Pardo, L. P. (2000). Sistemas de manejo de aguas residuales: para núcleos pequeños y descentralizados (Vol. 1). McGraw-Hill.
- [15] Pérez L.C. (2002). Estadística aplicada a través de Excel. Universidad Complutense de Madrid. España.
- [16] Quevedo, U. H. & Pérez, S. B. R. (2008). Estadística para ingeniería y ciencias. En estadística descriptiva (pp. 1-37). México.
- [17] Lichman, M. (2013). UCI Machine Learning Repository [<https://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant>]. Irvine, CA: University of California, School of Information and Computer Science.
- [18] Universidad Nacional Autónoma de México. (2000), Curso uso eficiente del agua y control de calidad de las descargas de aguas residuales en la industria. México.
- [19] Wills, B. A., Vélez, S., Arboleda, A. F., & Garcés, J. P. (2010). Propuesta metodológica para la evaluación de sistemas de tratamiento de aguas residuales domésticas en el sitio de origen. *Revista EIA*, (13), 93-105.