EMPIRICAL RESEARCH PRESS

International Journal of Engineering and Applied Computer Science

Volume: 04, Issue: 04, May 2022 ISBN: 9780995707545

Outbreaks Predictions of Covid19 in Tabasco, Mexico, using Fast Fourier Transform

Manuel Sandoval Martínez[1], *Janette Moreno Sandoval*[2], *Claudia Morales Barrón*[3], *Luz Elba Castillo Izquierdo*[4]

[1, 2, 3, 4] Computational System Department, Tecnológico Nacional de México, Campus Comalcalco, Tabasco, México.

10.24032/IJEACS/0404/012

Empirical Research Press Ltd.

London, United Kingdom

www.ijeacs.com

Outbreaks Predictions of Covid19 in Tabasco, Mexico, using Fast Fourier Transform

Abstract— **Results of analysis carried out on data of Covid19, in Tabasco State, are presented, from March 2020 to February 2022. The procedure named 7 day´s moving average was applied to smooth the original graph and, in this way, analyze the data more easily. Fast Fourier transform (FFT) was applied to find the frequency that allows us to detect the period (in weeks) of generation of greater Covid19 infections. FFT allows us to determine that for a period of 14n weeks (n=1,3,5,7) there will be a new outbreak. The analysis of data reveals, during the first wave, the week with the highest number of infections was July 7-12 (2751 cases), that is in week 14 (n=1). FFT indicates that the second wave of infection would be in week 42 with n=3, the maximum was obtained at week 40 (2122 cases). The next prediction (variant Delta) was for week 70 (n=5), where there would be a high number of cases (thrid wave); real data indicates that it was reached in week 73 (7,023 cases). The four-wave (Omicron) is predicted for week 98, however, it was reached in week 94 (12,834 cases). It should be noted that due to the high transmissibility of Omicron, the number of infection grew fast during January 2022, so in this case, the difference between the prediction and the reality, present a difference of four week, however, for week 94 number of cases remains very high (1265 cases). FFT has turned out to be an adequate tool to make predictions of four waves that have occurred in Tabasco.**

Keywords- Fast Fourier Transform, 7 days´ moving average, Frequency spectrum, Outbreak prediction.

I. INTRODUCTION

Covid19 was first identified in Wuhan, China in December 2019. By March 2020, the World Health Organization declared it a pandemic since it quickly spread throughout China and then to more than 190 countries in just three months [1], [2]. This virus has multiple characteristics that are different from other infectious deceases, including high infectivity during incubation can be up to 15 days (alpha variant) depending on the patient [3]. During this period, the patient may be asymptomatic, but they can still infect other people [4]. Some studies indicate that about 90% of cases are asymptomatic, and about 45% of infections were from asymptomatic people [5], [6]. This is a big reason why tests must be carried out to detect these cases and thus proceed according to the protocols stablished by the sub-secretary of prevention and health promotion (Mexico). For this reason, it is important to look for mathematic models [7], [8] to predict the possible outbreak of this disease in a given population; until now, these models have been of great relevance to study the behavior of the spread of the deceases.

Another characteristic of transmission is that speaking loudly, shouting, or singing in closed spaces can create a suitable transmission environment for the virus [9], [10] because a large number of infected droplets can be emitted. Lopez & Barrientos [11] point out, that spaces with poor ventilation, and a high density of people such as the busway, buses, markets, and even public toilets could increase the risk of contagion. We should mention, that the last two variants of Covid19 have caused great damage to health services around the world; these are the delta variant and the omicron variant. Studies carried out by Liu & Roclov [12] indicate that the reproduction index of the Delta variant is 5.8 and its transmissibility capacity is 60-120% more transmissible than its alpha predecessor is. On the other hand, the omicron variant reached a reproduction index greater than eight, for which it produced rapid growth in the spread of the virus [13], which causes very rapid growth in the contagion curve.

On the other hand, the technology to perform Fourier spectral analysis was used in many areas of medicine [14], [15] in such a way that Fourier analysis can provide useful information for the behavior of the pandemic. Takefushi [16], proposed to examine the period and spectral density of the pandemic by analyzing the data on deaths from covid19 in the frequency space. Furman [17], conducted research on the detection of covid19 by analyzing the sound of breathing of a sick patient through the computational application of Fourier transform; its results allowed obtain an automatic diagnosis of the deceases. McGobern [18], applied Fourier transformation to the cases of contagion in countries such as England, Brazil, USA, Mexico, and Argentina to analyze the cases of Covid19 and the death by it. The frequency spectrum indicated the periods during which more infections and a greater number of deaths had been generated. Elagan [19] propose remote equipment for the detection of Covid19, they used three types of sensors within which one of which measured the heart rate, and beats were estimated by means of the most prominent peak in the spectral density. The variation in these frequencies was complemented with parameters of temperature and percentage of platelets to determine if a person was infected or not.

II. BASIC DEFINITIONS

A. *Fast Fourier Transform Definition*

Covid19 data can be analyzed, through a powerful mathematical tool named Fourier transform, which transfers time-series data to a space dominated by the frequency of the signal to be analyzed. The Fourier transform definition is [20]:

$$\mathfrak{F}[f(t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-it\omega} dt \qquad (1)$$

It should be noted that FT is applied on stationary functions, when it is applied to a non-stationary function, the windowing method must be used. When collected and graphing the first Covid19 data presented almost oscillatory behavior, although with very pronounced peaks. However, in the context of digital signal processing, the 7 day´s moving average method [21] represents a finite pulse filter, which is defined as:

$$y[n] = \frac{1}{7}\sum_{-3}^{3} x[n-k] \qquad (2)$$

Where x, y and n represent the input data, the output signal, and the time index respectively; this method allows, "signal smoothing" in order to perform a better analysis. We analyze a digital signal, that is, a discrete signal and not a continuous signal, the Discrete Fourier Transform (DFT) must be used, which is defined [22] as:

$$G\left[\frac{n}{NT}\right] = \sum_{k=0}^{N-1} g(kT)e^{\frac{-i2\pi nk}{N}} \qquad (3)$$

With $n = 0,1, \dots N-1$, $\frac{n}{NT}$ represent fundamental frequencies, $g(kT)$ is the value of each sample of the signal, T represents the sampling period and N the number of data collected. In addition, the calculation is carried out computationally, an algorithm is also used, that allows the calculation time to be reduced considerably, and it is the Fast Fourier Transform. In order to analyze Covid19 data, we work with the FFT algorithm.

III. METHODOLOGY

We analyzed data of Covid19 in Tabasco, Mexico, from April 2020 to February 2022, taken from the daily report issued by the Secretary of Health in Tabasco State. In order to perform a more adequate analysis, the 7 day´s moving average method was applied to smooth the data curve since the original (daily cases) will present a diversity of pronounced peak and cause some problems to analyze them adequately with Fourier transform. Then Fourier transform was applied in order to the modified data and perform an analysis in the frequency space and, with it, find the period of the fundamental frequencies of the signal. The periods obtained were carefully analyzed and it was sought which of them could approximate the high concentration of infections by the pandemic in Tabasco. The programming of the algorithm to analyze the covid19 data, the graph, and the application of the algorithm for the FFT, was running in the GNU Octave, which is free to download. The type of research developed is quantitative-explanatory, the population studied was all cases of covid19 infections in Tabasco, and the dependent variable is the number of accumulated infections, frequencies detected by FFT, and the spectral density. The independent variable is the time measured in days. During this analysis, there was no discrimination based on gender, social class, age, or religion. With the analysis in the frequency space, the fundamental frequency, which indicates the cases with the highest number of infections per week, is spotted and compared with the real data in order to determine the relative error in the prediction, of a certain wave of infection.

IV. RESULTS

After collecting Covid19 data in Tabasco (March 2020-February 2022) the graph was made, finding that the first maximum of the contagions wave was obtained, more than 2750 cases in week 14 (July, 7-12). During August and September, the number of infections remained relatively high, with values between 1,500 and 2,000 cases per week.

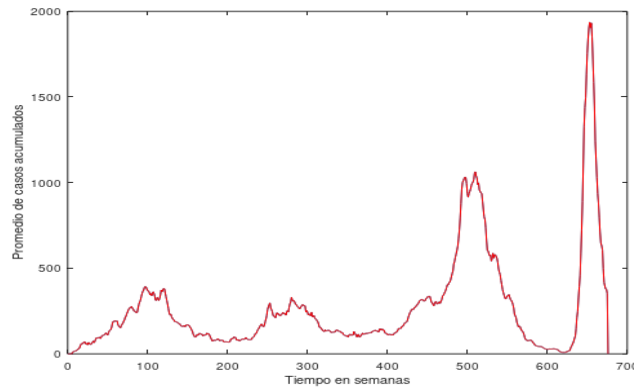

Figure 1. Covid19 data with 7 day´s moving average

During October and November, the infections remain low between 500 and 600 cases per week. However, in mid-December the cases of Covid19 began to increase again; by January 2021 appear the second wave of infection, and more than 2,100 new cases were recorded (week 40). The trend began to go down in mid-February and March of that year (See Figure 1).

To smooth the graph in Figure 1, the 7 day´s moving average method was applied, thus avoiding very high peaks in the original graph, which could cause computation problems in the Fourier analysis. Once this process was carried out, the FFT was applied to analyze its behavior in the frequency space and, thus, locate where the largest number of contagions in the population is concentrated. From this point, the period of that frequency can be calculated and thus predict in which week a very high number of new infections could be had. The application of FFT to Covid19 data produces the frequency spectrum that is shown in Figure 2.

It can be seen, that there are a series of peaks that represent the frequencies of the signal, however, we are interested in analyzing the one that presents a better option to make a prediction about the density of Covid19, that is, what period can be predicted the date of a new wave of infection. In the spectral analysis, we found a period of 14 weeks as a candidate.

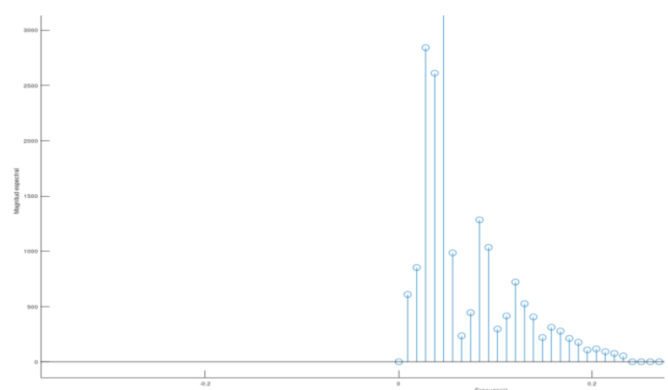

Figure 2. Spectral density of Covid19 data

We did not consider the smallest spectral density since they represent cases of contagions with small values per week and do not constitute an alarm for the State health authorities. Analyzing the spectrum of frequencies, it was found that the one that best predicts official data is $f\omega = 0.069 rad/s$, so the oscillation period is at week 14. With this information, and analyzing Table 1, we can see that in week 14 the number of infections was 2751; in week 40 there were 2122 cases (wave 1 and wave 2, for alpha variant), and at week 73 more than 7,000 cases were obtained (third wave, delta variant) and at week 92, more than 12,800 cases were reported (Omicron variant). These four weeks represent the maximum of the data signal, that is, they correspond to the four waves of high infections in Tabasco.

According to what we found, it is predicted that there will be a maximum in week 14 (n=1), coinciding 100% with official data. For week 42 (n=3) a maximum is predicted, however, real data indicate that it occurred during week 40, in this case, the relative error is 5%. On the other hand, for week 70 (n=5) a new maximum is predicted and this happened in week 73 (4% of error). Note that, for variant alpha and variant delta, the approximation in the predictions presents a very small error. Finally, the prediction for a fourth wave derivate from Omicron was stablished for week 98 (n=7), and reviewing Table 1, the real maximum was reached at week 94 (relative error 5%). Although the error is small, we can see that the net difference is four weeks; and for delicate situations like Covid19, it could be an important difference to consider in public health decision-making. However, a possible explanation for this case is that the Omicron variant presented a very high rate of infections, higher than the rates of the two previous variants, and the cases derived from this variant grew very quickly.

Week	Date	Accumulated cases	Variant
1	April 6-12, 2020	54	
9	June 1-7, 2020	1344	
14	July 7-12, 2020	**2751**	**Alpha variant**
15	July 13-20, 2020	2462	
28	October 12-19, 2020	**502**	
40	January 4-11, 2021	**2122**	
42	January 18-24, 2021	**1945**	**Alpha variant**
56	April 27 to May 04, 2021	967	
70	August 4-10, 2021	6042	
73	August 25-31, 2021	7023	**Delta Variant**
84	November 11-16, 2021	285	
94	January 12-18, 2022	12834	**Omicron Variant**
98	February 24–March 3, 2022	1265	

However, FFT is still a good tool to alert the authorities to a possible new outbreak, since for week 98 the number of infections continued to be high (over a thousand cases). As the final analysis of this section, using FFT we were able to identify the peaks of high contagions which would be reached in odd multiples of the period found that is $n = 1$, $n = 3$, $n = 5$ y $n = 7$. Although the predictions were very close to reality, this tool does not allow measure the number of infections, only allow to predict the periods for which there would be a high number of infections. However, for public health issues, it is more important to know the possible period of maximum contagions in order to prepare for the fight against the new wave.

V. CONCLUSIONS

There are various mathematical models to analyze the data of Covid19, and from them make predictions in the medium or long term. In this work, we found that the use of FFT makes it possible to analyze data of Covid19 infections (Tabasco) in the frequency space and in this way determine the period of the most representative frequencies of the signal. We found that the frequency that best predicts high infections has a period of $14n$ weeks ($n = 1, 3, 5, 7$). In this way, and sorting out the accumulated data by weeks it was possible to verify that the prediction made by FFT is very good since in week 14 the maximum (2751 cases) of the first wave was reached. The prediction indicates that in week 42 there would be a high rate of infection (1660 cases) but the maximum was reached in week 40 (2,122 cases) obtaining a 5% of error. Third-wave, was reached in week 73, and FFT was predicted for week 70, so there is a relative error of 4%. Finally, a fourth wave was predicted for week 98, although, it was reached in week 94 (5% of error). In this case, the error increased a little because the Omicron variant generated a very rapid increase in the number of infections due to its high transmissibility (much higher than the previous variant). Despite this, in week 98 the number of infections was still high, above a thousand cases, so FFT keep indicating that there would be a considerable number of cases. As consequence, FFT is a tool that has made it possible to roughly predict the week in which a wave of infections could occur regardless of the type of variant, although with Omicron the prediction error was slightly higher due to its behavior; the predictions can be considered as correct. This information can be very useful for health authorities so that they can take the appropriate measure to contain the spread of the disease.

ACKNOWLEDGMENT

Our acknowledgment to National Technology of Mexico, campus Comalcalco for all the facilities that they offer to us in order to develop this research, including academic and economical support.

REFERENCES

[1] Hu, B. Guo, H. Zhou, P. Shi, Z. Characteristics of Sars-Cov-2 and Covid19. *Nat. Rev. Microbiol.* 19, 2021,141-154.

[2] Shereen, M. Khan, S. Kasmi, A. Bashir, N. Siddique, R. Covid19 infection: origin, transmition, and characteristics of human coronaviruses. *Journals of advanced research*, 24, 2020, 91-98.

[3] Wang, C. Horby, P. Hayden, F. Gao, G. A novel coronavirus outbrake of global health concern. *The Lancet*, 395, 2020, 470-473.

[4] Zhao, S. Chen, H. Modeling the epidemic dymics and control of covid19 outbrake in chine. *Quantitive Biology*, 2020, 1-9.

[5] Li, R. Pei, S. Chen, B. Song, Y. Zhang, T. Substancial undocumented infection facilities the rapid dissemination of novel coronavirus. *Science*, 368 (6490), 2020, 489-493.

[6] Liu, J. Zhen, X. Tong, Q. Li, W. Wang, B. Sutter, K. Trilling, M. Lu, M. Dittmer, U. Yang, D. Overlapping and discrete aspect of the pathology and phatogenesis of emerging human phatogenic coronaviruses Sars-Cov, Mers-Cov and 2019 nCov. *J Med. Virol*. 92,5, 2020, 491-494.

[7] Lu, H, Stratton. C.W. Tang. Outbreak of pneumonia of unknown etiology un Whunan China. The mystery and the miracle. J. Med. Virol. 92, 2020, 401-402.

[8] Ndairou, F. Area, I. Nieto, J. Torres, D. Mathematical modelling of covid19 transmission dynamics with case study of Wuhan. *Chaos, Solitons and Fractals*, 135, 2020, 1-6.

[9] Aguilar, N. Hernández, A. Ibanes, C. Characteristics of Sars-Cov.2 and its transmissions mechanism. Rev. Latin.Infect. Pediatr, 33, 3, 2020, 143-148.

[10] Características del Sars-Cov-2 y sus mecanismos de transmisión. Rev. Latin. Infect. Pediatr, 33,3, 2020, 143-148.

[11] McCloskey, B. Zumla, A. Ippoloto, G. Blumberg, L. Arbon, P. Cicero, A. Endericks, T. Lim, P. Borodina, M. Mass ghatering events and reducing further global spread of Covid19: A political and public health dilema. *Lancent*, 395, 2020, 1096-1099.

[12] López, N. Barrientos, T. The role of speech in the transmition of Sar-Cov-2: recommendations for confined space. *Salud pública de México*, 62, 5, 2020, 455-456.

[13] Liu, Y. Roclov, J. The reproductive number of the delta variant of Sar-Cov-2 is far higher compared to the ancestral Sar-Cov-2 virus. *J. Travel Med*, 28, 7, 2021.

[14] Mahadi, S.A (2021). South African population immunity and severe covid19 with omicron variant. MedRXviv, DOI 10.1101/2021.12.20.21268096.

[15] Bao, Y (2015). A digitally generated ultrafine optical frequency comb for spectral measurements with 0.01 pm resolution and 0.7ms response time. *Light: Science and Applications*, 4, 2015,300.

[16] Fhaty A. On chip parallel, Fourier transform spectrometer for broadband selective infrared spectral sensing. *Microsystems and nanoengineering*, 6, 10,2020.

[17] Takefuji, Y. Fourier analysis using the number of covid19 daily deaths in the US. *Epidemiology infection*. 149, 2021, E64. doi:10.1017/S0950268821000522.

[18] Furman, E. Charushi, A. Eirikh, E. Furman, G. Sokolovsky, V, Malinin, S. Sheludko, V. The remote analysis of breath sound in covid19 patients: a series of clinical cases. *MedRxiv*, 9, 16, 2020.

[19] McGovern, S. Spectral processing of covid19 time series data. *Electrical engineering and systems science, Cornell University, 2020*.

[20] Elagan, S. Abdelwahab, Zanaty, S. Alkinani, M. Alotaibe, H. Zanaty, M (2021). Remote diagnostic and detection of coronavirus disease (covid19) system bades on intelligent healthcare and internet of things. *Results in Physics*, 2021,903-910.

[21] Severov, V. Fourier series, Fourier transform and their application to Mathematical Physics. *Springer International Publishing*, 2017.

[22] Cecil, W. Covid19: daily fluctuation, a weekly cycle, and a negative trend. *Am. J. Managed Care*. 26, 7, 2020, 248-285.

[23] Oran Brigham. *The Fast Fourier Transform and its Applications*. Prentice Hall, 2018.

AUTHORS PROFILE

M. Sandoval is a Professor in Instituto Tecnológico Superior de Comalcalco since 1997. He pursued a Bachelor of Physics, Master of Science in Applied Math, and PhD in Physics Education from Instituto Politecnico Nacional in the year 2011. His main research work focuses on Physics Education and Mathematics Modeling.

M.I. J. Moreno is a Professor in Instituto Tecnológico Superior de Comalcalco since 2014. She has a degree in Computer Science from Universidad Autónoma de Morelia and a Master of Science from University TecMilenio in the year 2013. Her main research work focuses on Mobile App Development.

M.C.D C. Morales is a Professor in Instituto Tecnológico Superior de Comalcalo since 1997. She has degree in Computer Science from Universidad Cristobal Colon, and Master in Teacher Competences from Instituo Universitario Puebla in year 2014. Her main work focuses in Analysis of Database.

Lic. L. Castillo is a Professor in Instituto Tecnológico Superior de Comalcalco since 1995. She has degree in Computer Science from Universidad Juarez Autónoma de Tabasco in the year 1992. Her main work focuses in Analysis of Database.

ANNEX

Annex 1. Tabasco´s data of Covid19 (a small sample)

Date	6	7	8	12	30	01-may	2	3	4	5	31	01-jun	2	3	30	01-jul	2	31	01-ago
Days	1	2	3	7	25	26	27	28	29	30	56	57	58	59	86	87	88	117	118
Acumulated case	102	108	112	153	984	1066	1077	1196	1247	1277	4142	4323	4609	4807	10695	10921	11165	21301	21747
daily case	6	4	1	3	82	11	119	51	30	42	181	286	198	190	226	244	279	446	417
7´s day M.A	0	0	0	7.71	59.6	52.857	66.86	61.14	55.29	54	155.3	178.4	189	189	243.1	239.857	241.714	344.571	372.57
Average	7.714	12.4	14	23	62.4	66.429	74.43	65.14	73	85.57	190.3	192	167	163	311.9	333.429	351	354.571	334.14
Weeks				54				428	4		1087	8							

Annex 2. Script to compute FFT on Covid19 data

```
%%%%%%%%%%%%%%%%% fast Fourier Transform
w=(1:n)*dw;
w2=(0:n2-1)*dw2;
wc=pi/dt;
wc2=pi/dt2;
g=fft(f);
g2=(fft(f2));
power=(abs(g).^2)/n;
figure(3)
stem(w,power)
[M,I]=max(power);
periodo=2*pi/w(I)
text(w(I)+0.1, power(I), ['Periodo=',num2str(periodo)]);
m=periodo/30;
sem=periodo/7
figure(4)
indice=power2>50;
g2=indice.*g2;
power21=(abs(g2).^2)/1000000;
nf=length(power21);
 i=1;
while i<nf
   if power21(1,i)>300
   dfreq=power21(1,i)
   f3=w2(i)
   periodo=2*pi/f3
   end
i=i+1;
end
stem(w2,power21)
[M2,I2]=max(power2);
periodo=2*pi/w2(I2);
```