



"El saber de mis hijos  
hará mi grandeza"

UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Licenciatura En Física

Física Computacional I

---

### Evaluación 1

*"Análisis de series de tiempo con los datos del  
observatorio Tutuila, Samoa Americana"*

---

Ismael Espinoza Arias

Profesor Carlos Lizárraga Celaya

Hermosillo, Sonora

25 de febrero de 2021

# 1 Introducción

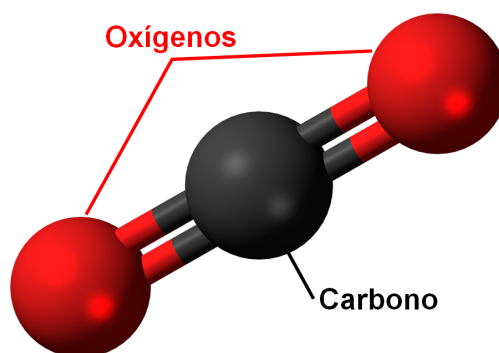
El siguiente reporte forma parte de la primera evaluación del semestre, de la materia de Física Computacional 1, por lo que, a lo largo de él, hablaremos sobre el procedimiento usado para llegar a lograr la práctica propuesta y sobre los resultados que obtuvimos en el transcurso de la actividad.

Para la realización de esta evaluación, fue necesario el uso de los datos del observatorio Tutuila, Samoa Americana, donde las variables que observamos su comportamiento, fueron las concentraciones de CO<sub>2</sub> con respecto a nuestra serie de tiempo. Esta evaluación fue realizada con el propósito de lograr demostrar las habilidades aprendidas en el transcurso del curso de Física Computacional 1, donde aplicaremos los conocimientos que tenemos hasta el momento gracias a las 6 actividades ya realizadas. A lo largo de este reporte veremos e interpretaremos los datos obtenidos de esta actividad de evaluación.

## 2 Antecedentes

Antes de iniciar la actividad, entraremos un poco en contexto acerca del tema, como mencionamos anteriormente nos interesa explorar el comportamiento de CO<sub>2</sub> y su correlación con la temperatura, donde estudiaremos las concentraciones de CO<sub>2</sub> se han venido monitoreando diariamente por los Laboratorios de Investigación del sistema tierra de la agencia NOAA (ESRL NOAA), desde hace un poco más de 60 años. El CO<sub>2</sub> es un gas tipo invernadero que atrapa la radiación solar y por lo tanto calienta la atmósfera. Las emisiones de CO<sub>2</sub> se han incrementado producto de las actividades humanas, sobre todo por la quema de combustibles fósiles, fabricación de cemento y otros.

El dióxido de carbono (CO<sub>2</sub>) es un gas incoloro e inodoro compuesto por un átomo de carbono y dos de oxígeno en enlaces covalentes. El CO<sub>2</sub> produce el desplazamiento del oxígeno y en concentraciones altas, de más de 30.000 ppm, puede producir asfixia. En aire interior el CO<sub>2</sub> es un gran indicador de la calidad de aire ya que actúa como chivato de la renovación de aire. Es sabido que a partir de concentraciones de más de 800 ppm en entornos laborales, se comienzan a producir quejas debidas a olores.



*Molécula de dióxido de carbono*

### 3 Observatorio de referencia de Samoa Americana

El Observatorio de Samoa Americana (SMO) está ubicado en el medio del Pacífico Sur, a medio camino entre Hawai y Nueva Zelanda. Se caracteriza por el calor y la humedad durante todo el año, exuberantes montañas verdes y una fuerte cultura samoana. El observatorio está situado en el extremo noreste de la isla de Tutuila, Samoa Americana, en el cabo Matatula.

El observatorio se estableció en 1974 en un sitio de 26,7 acres. Desde su construcción, ha sobrevivido a dos grandes huracanes, un terremoto y un tsunami con daños menores. Un personal de 3 personas opera la instalación durante todo el año. Este observatorio tiene la particularidad de obtener el 30% de su energía diurna de paneles solares.

El territorio de Samoa Americana consta de 7 islas que se encuentran a 4700 km (2600 millas) al suroeste de Hawai, en el centro del Océano Pacífico y son las más antiguas de las islas de Samoa. La masa terrestre total de las 7 islas es de 197 kilómetros cuadrados (76,1 millas cuadradas) y el 74% pertenece a la isla de Tutuila. Tutuila tiene aproximadamente 30 km de largo y 6 km en su punto más ancho con Pago Pago Harbour en el centro de la isla. Este puerto son los restos de un cráter volcánico y una ubicación profunda y protegida, ideal como un puerto de barcos oceánicos. La mayoría de la población reside en esta isla. El Observatorio NOAA GML se encuentra en el extremo noreste de la isla, en el cabo Matatula, pasando el pueblo de Tula. El viaje desde el aeropuerto hasta el observatorio es de aproximadamente 24 millas, pasando por las principales zonas comerciales, el puerto.

## 4 Desarrollo de la actividad

### 4.1 Archivo de Datos

El archivo necesario para la elaboración de esta actividad fue dado por el profesor, los cuales se incluyeron en el repertorio github para la complementación de la actividad. Analizando los datos podemos llegar a la conclusión de que nuestros datos existen desde el año 1976 hasta el año 2019, es decir alrededor de 43 años de datos, también las columnas las tenemos separadas por años, meses, días, horas, minutos, segundos, longitud, latitud, altitud entre otras cosas más.

### 4.2 Análisis de Datos

Para empezar con el código de esta actividad, en la primera celda vamos a cargar las bibliotecas que vamos a manejar, bibliotecas que vamos hemos estado utilizando anteriormente y nos ayudaran a iniciar con nuestro análisis y lectura de nuestros datos en el archivo correspondiente antes establecido.

```
#importamos bibliotecas que nos puedan ser de ayuda
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Bibliotecas para el análisis de Series de Tiempo y Modelos de Pronóstico
import statsmodels.api as sm
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.arima_model import ARIMAResults
from statsmodels.tsa.stattools import acf, pacf

import pandas.util.testing as tm

# Suprimir mensajes de alertas de cambios futuros
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

# Matplotlib mostrará los resultados en la Notebook de Jupyter
%matplotlib inline
```

Posteriormente empezamos con la creación de nuestra serie de tiempo, para ello se leyó el archivo y se designó sus data frames. Notamos también que los datos fueron agrupados por sus respectivas columnas. Ponemos nuestro respectivo link de donde sacamos la base de datos a utilizar y damos forma a las columnas.

```
url='https://raw.githubusercontent.com/carloslizarzagac/FisicaComputacional1/master/Eval1/data/co2_spo_DailyData.txt'
my_cols_orig = ['0','year','month','day','hour','minute','second','1','CO2','2','3','4','5','6','7','8','9']
df_orig= pd.read_csv(url,names=my_cols_orig,encoding='cp1251',sep='\s+',header=None,skiprows=151,engine='python')
#print(df_orig)
df_2=df_orig.copy()
```

Es importante notar que los nombres de las columnas ahora se encuentran "abreviados". Observamos una muestra de datos para verificar si se estan leyendo.

Comenzamos con la segunda parte de nuestra actividad, ya que corrimos los datos y los administramos en las columnas anteriores, procedemos a realizar nuestro análisis de la serie de tiempo.

```
#Creamos una funcion para probar la estacionaridad de la serie de tiempo
def test_stationarity(df_ts):
    #Determina la estadistica de promedios moviles
    rolmean = df_ts.rolling(window=365,center=True).mean()
    rolstd= df_ts.rolling(window=365,center=True).std()
```

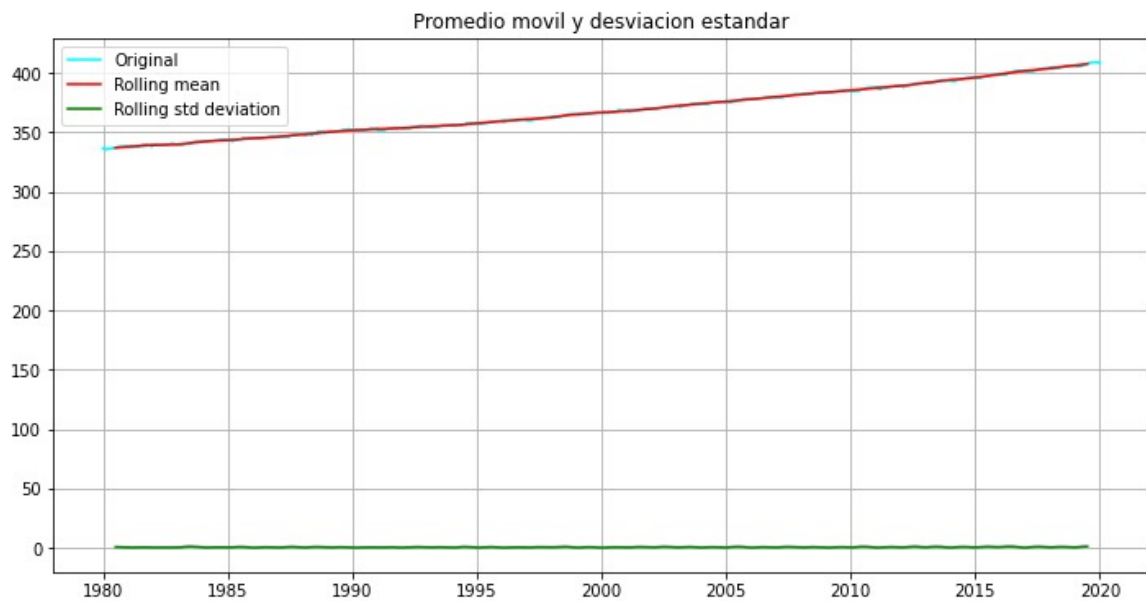
Después empezamos a definir el tamaño de la figura que vamos a utilizar, graficamos las estadísticas de los promedios móviles, realizamos pruebas y obtenemos nuestros resultados.

```
#PROBAMOS ESTACIONARIDAD
#ts_test.head()

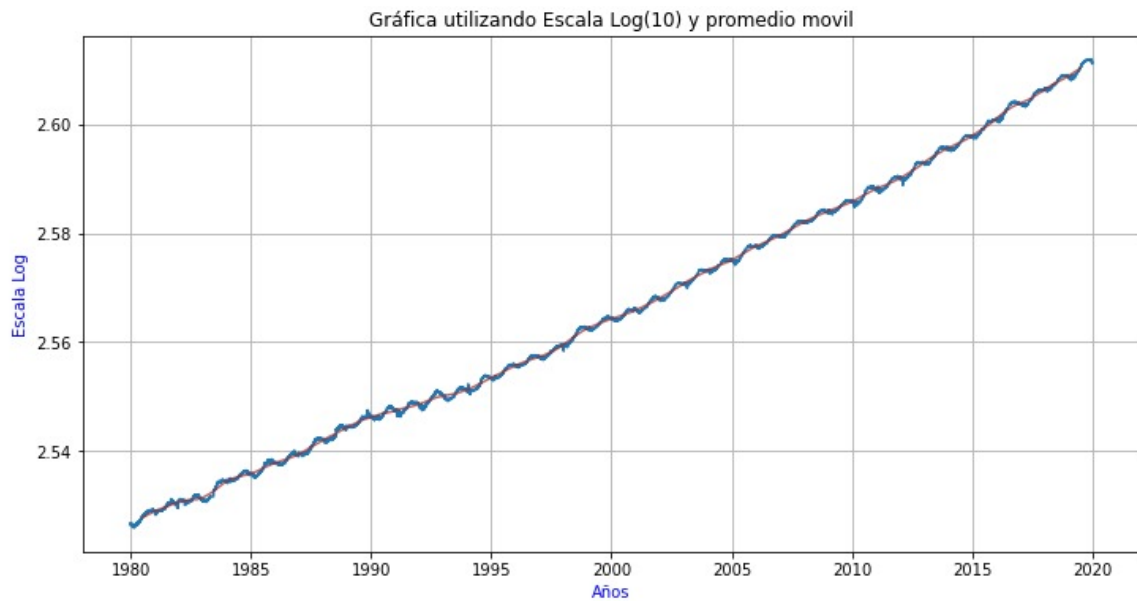
test_stationarity(ts_test['CO2'])
```

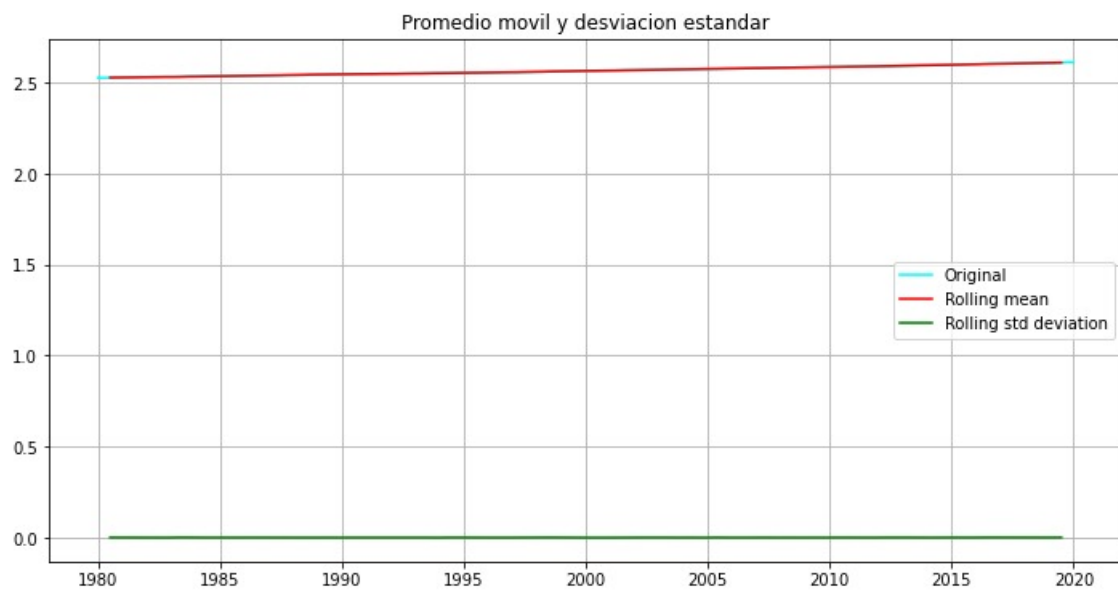
Creamos nuestras funciones:

```
#Creamos una funcion para probar la estacionaridad de la serie de tiempo
def test_stationarity(df_ts):
    #Determina la estadistica de promedios moviles
    rolmean = df_ts.rolling(window=365,center=True).mean()
    rolstd= df_ts.rolling(window=365,center=True).std()
```

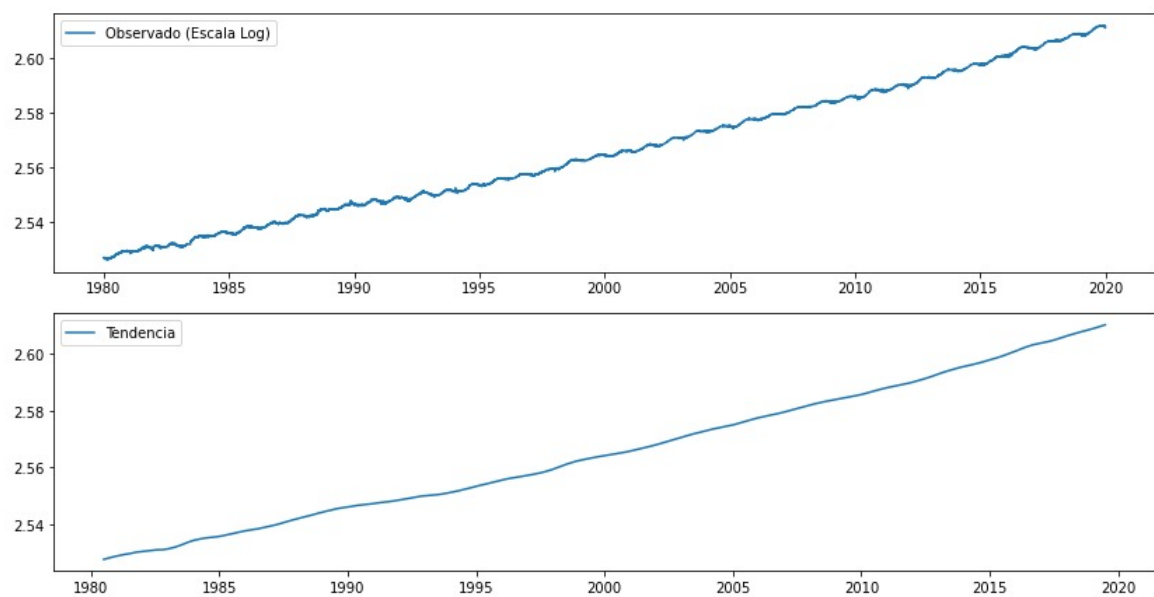


Obtenemos nuestros resultados obtenidos de la prueba de Dickey-Fuller. Volvemos a graficar y nos encontramos con los siguientes datos. Cambiando la escala como hemos hecho anteriormente, cambiamos la escala a logaritmo de base 10 y podremos preciar unos datos distintos a nuestra gráfica.

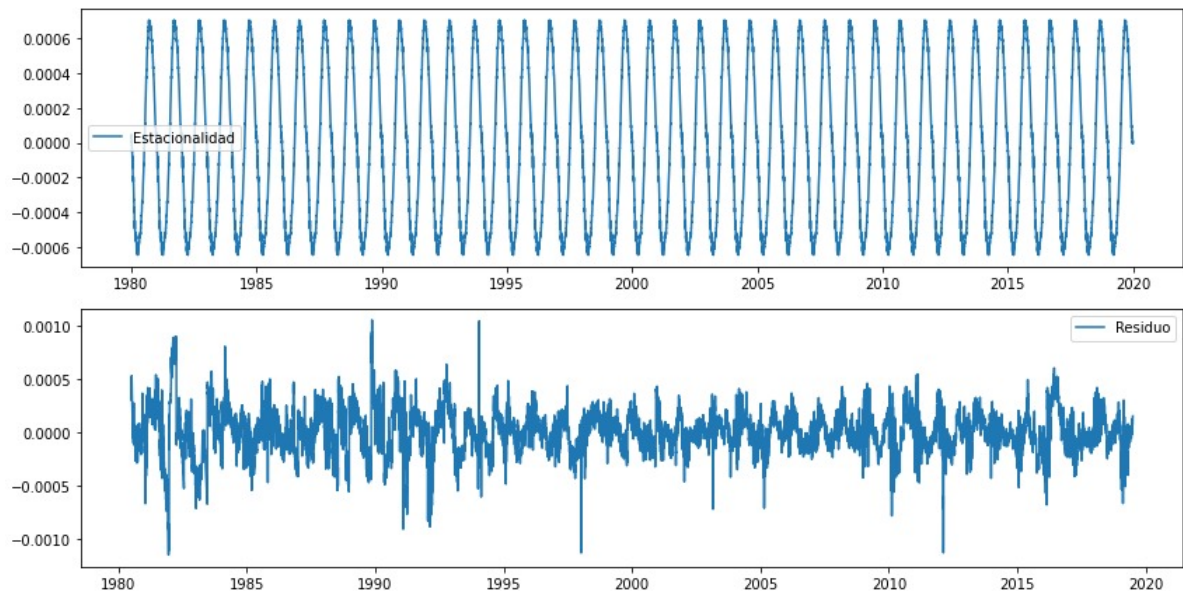




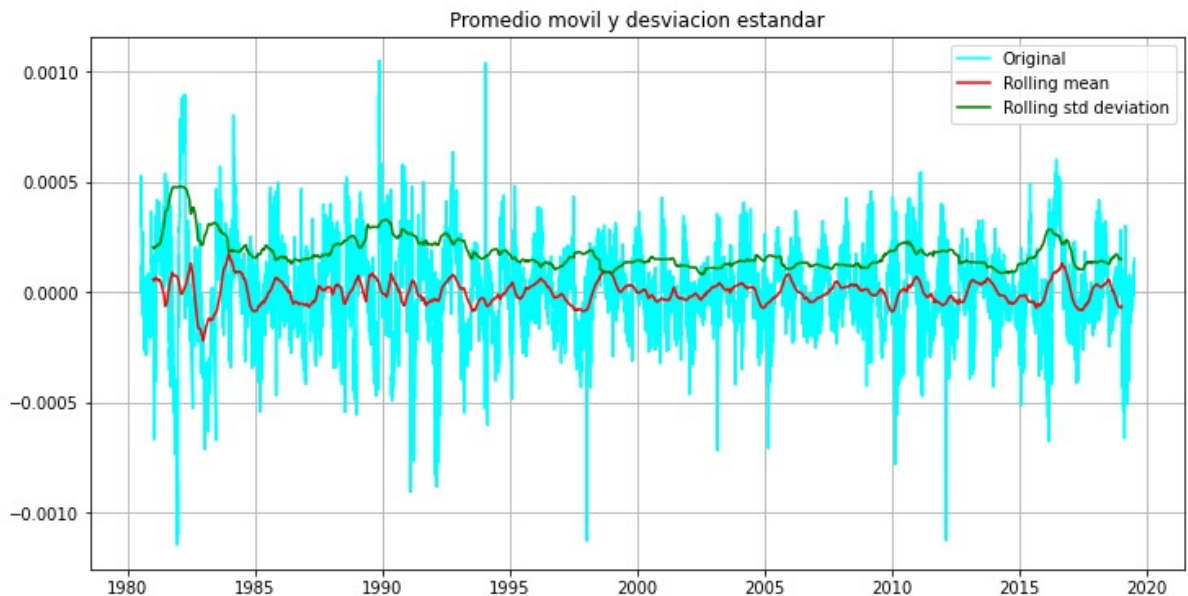
Como podemos ver en nuestros datos, no es estacionaria, por lo tanto, utilizaremos otra biblioteca importante, que es el método de descomposición usando la biblioteca statsmodels. En este caso obtenemos los siguientes resultados:





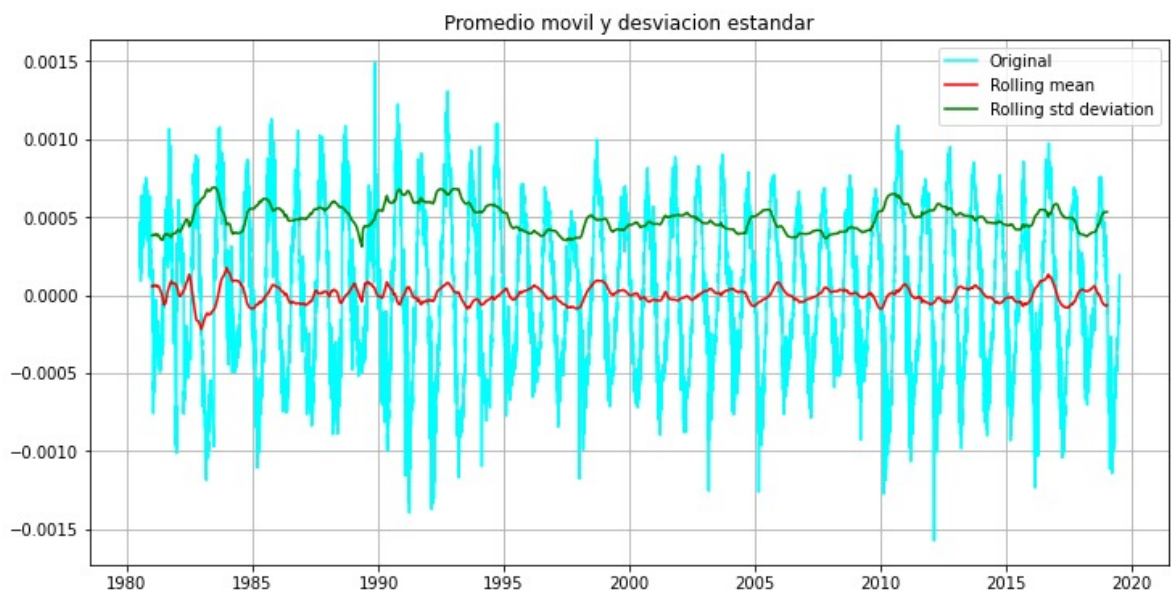


Podemos ver que la tendencia del CO2 va en un aumento constante y de forma relativamente rápida, la dependencia con las estaciones no es muy clara, vemos que si tiene un ciclo por ejemplo baja en 5 ocasiones cada año y 5 veces aumenta también, las dos igualmente espaciadas. En cuanto al residuo me parece que cada vez disminuye la variabilidad de este pero un mínimo.

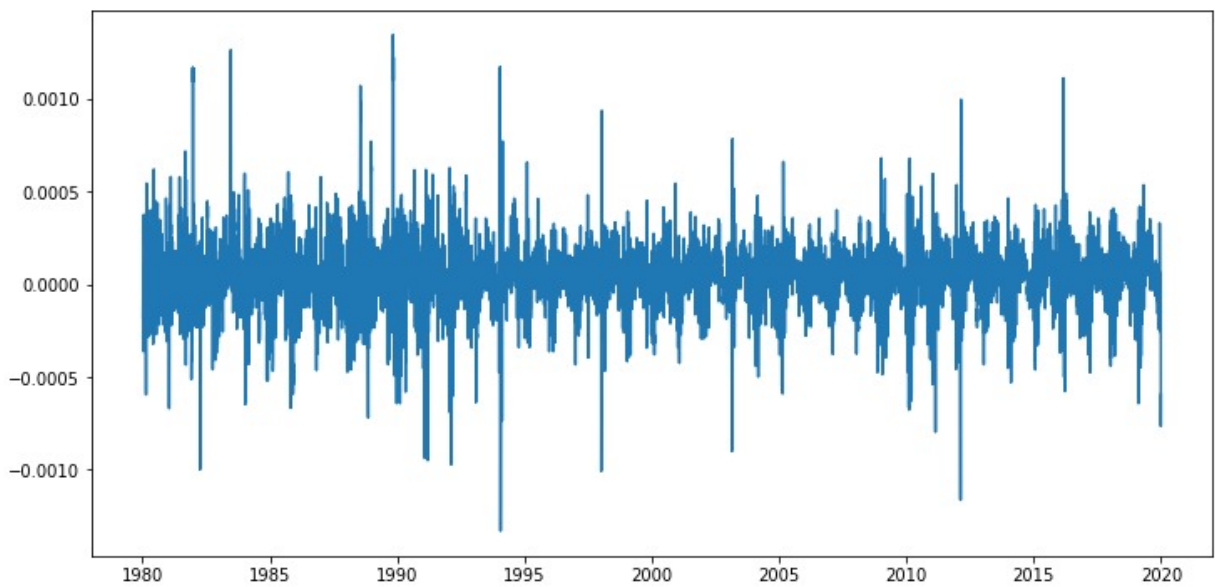


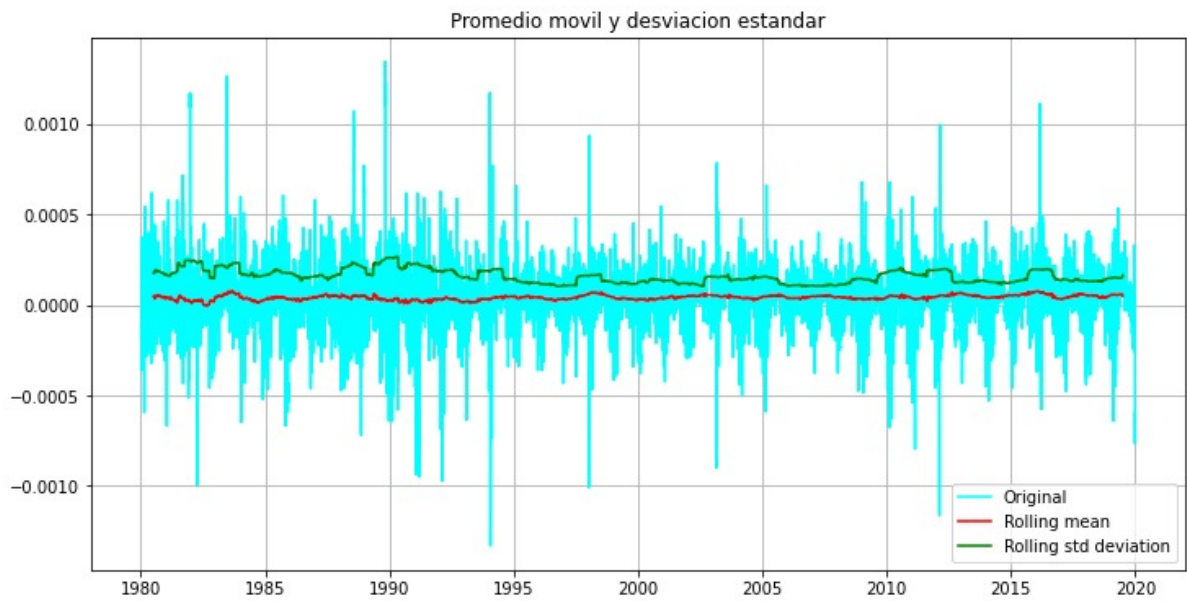
Define un nuevo dataframe con la diferencia entre la serie logaritmo original y el promedio móvil, y prueba después su estacionaridad.



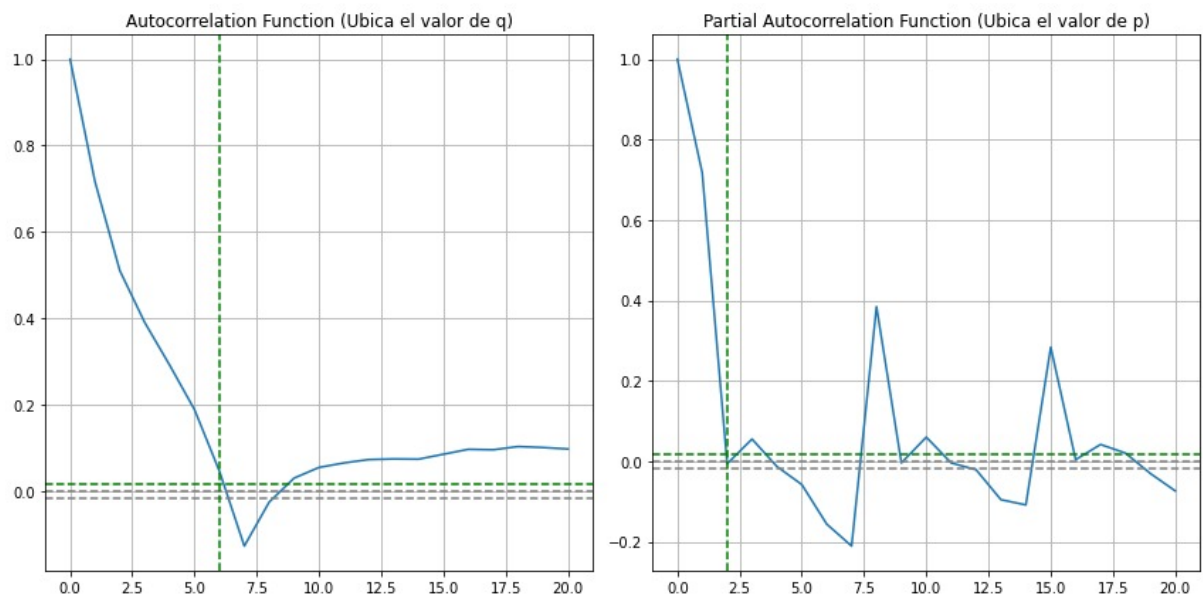


Diferenciamos usando la función:





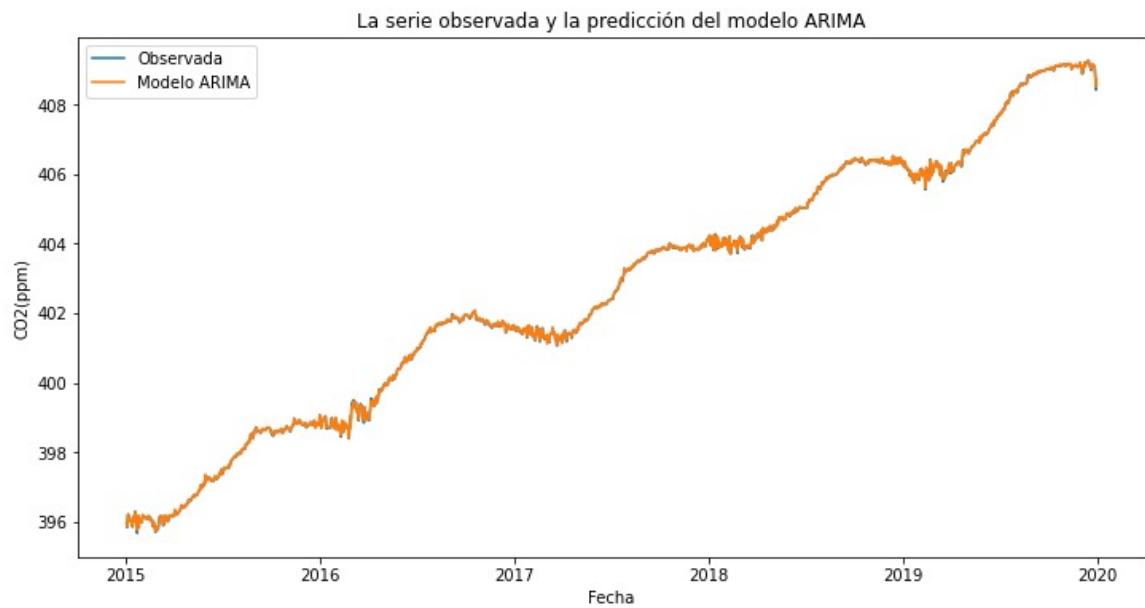
Ahora aplicamos el modelo ARIMA, cambiamos los tamaños de las gráficas. Graficamos unas líneas verticales como referencias:

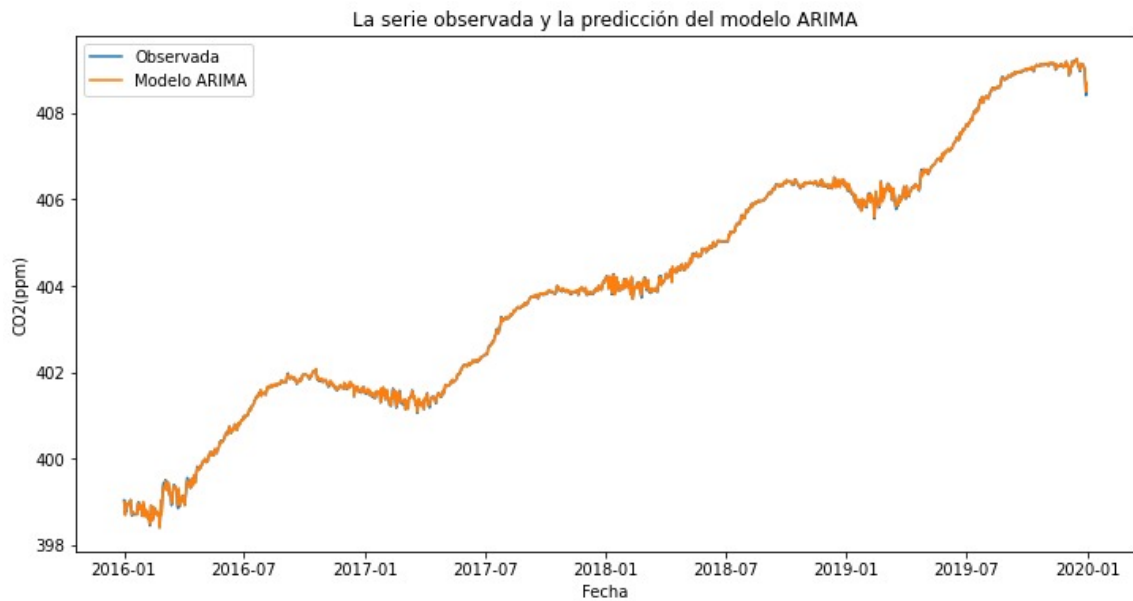


Contrasta la serie observada de concentraciones de CO2 contra la construida por el modelo ARIMA. Definimos una serie de tiempo con las predicciones obtenidas por el método ARIMA.



Graficamos solo los ultimos 5 años de la Serie de CO2 y el modelo ARIMA para ver mas a detalles el modelo.





## Conclusión

Al concluir esta actividad, podemos ver la gran funcionalidad que tienen estas bibliotecas para la hora de trabajar con estos datos, ya que con cada una de estas nos ayudó en una función específica. También podemos ver como es que las distribuciones de CO2 tienen una gran correlación con respecto a las temperaturas manejadas en la actividad. Solo queda pensar que esta actividad fue de las mejores si de retroalimentar lo aprendido se trata.

## 5 Bibliografía

1. *Dióxido de carbono CO2*. (2020, 24 abril). Instituto para la Salud Geoambiental. <https://www.saludgeoambiental.org/dioxido-carbono-co2>.
2. *Global monitoring laboratory*. (s. f.). ESRL Global Monitoring Division - American Samoa Observatory. Earth system research laboratories. Recuperado 24 de febrero de 2021, de <https://www.esrl.noaa.gov/gmd/obop/smo/>.
3. *Series de Tiempo*. (2018, octubre). Series de tiempo. <http://www.estadistica.mat.uson.mx/Material/seriesdetiempo.pdf>
4. Vermorel, J. (2105, 15 enero). Qué es el pronóstico de series de tiempo: Una introducción no técnica. LOKAD. <https://www.lokad.com/es/que-es-el-pronostico-de-series-de-tiempo:%7E:text=El%20pron%C3%B3stpersonal>.