

Actividad 6

José Agustín Parada Peralta
Departamento de Física
Universidad de Sonora

20 de Febrero de 2021

1. Introducción de la actividad

En la presente actividad, se analizan las series de tiempo de temperaturas máximas y mínimas de la Ciudad Lerdo, en el estado de Durango, en un periodo de tiempo correspondiente a 20 años. De 1986 a 2006

Con ese fin, se utilizaron distintas bibliotecas que nos permitieron llevar a cabo análisis estadísticos de los datos correspondientes a la estación meteorológica, brindados por la Comisión Nacional del Agua. Tales bibliotecas son: *Matplotlib*, *Statsmodels*, *Pandas*, entre otras, utilizando Python.

2. Información analizada y comentarios generales

Se analizaron, como anteriormente se mencionó, las series de tiempo correspondientes a las temperaturas máximas y mínimas en un cierto periodo de años. Se llevó a cabo una prueba de hipótesis nula en las series de tiempo, que buscaba rechazar la hipótesis de que los datos no son estacionarios.

En ambas series, la prueba de Dickey - Fuller nos arrojó son estacionarias. No obstante, para aplicar los aprendido, se les aplicó escala logarítmica, y se practicó una diferencia con promedio móvil exponencial. Entonces, se obtuvieron los pronósticos con gran grado de precisión, obteniendo errores muy pequeños para ambas series.

Ahora, vamos a profundizar en el proceso de obtención del pronóstico de las series de tiempo:

- i. Los datos por sí mismos resultaron ser estacionarios, pues sus *valores p* son menores que 0,05. Sin embargo, por los motivos mencionados con anterioridad, se aplicaron escalas logarítmicas a ambas series, y respectiva a la de temperaturas mínimas, se diferencié utilizando un promedio móvil exponencial de 30 días. Con esto, obtuvimos valores p del orden de 10^{-27} .

- ii. La prueba Aumentada de Dickey - Fuller nos dicta, mediante los valores críticos del 1 % y 5 % que las series de tiempo no poseen tendencia, lo que implica que sus propiedades estadísticas se mantienen constantes a lo largo del tiempo.
- iii. En general, los datos se mantienen oscilando en un rango constante. En ciertas ocasiones (sobre todo en un rango de tiempo externo al pronosticado) el rango de oscilación de las temperaturas variaba (aumentaba y disminuía).
- iv. Los datos meteorológicos de la Ciudad Lerdo fueron bastante aceptables. En lo que a temperaturas respecta, los huecos en los datos no son muy abundantes. Sin embargo, no inexistentes. Entre estos, los datos que sí pudieron suponer un problema en lo que a huecos se refiere, fue el caso de los datos de la *evaporación*. El rango más amplio sin tropezar con huecos, me parece, fue de unos 4 años. Sin embargo, gracias al bajo nivel de huecos en los datos de temperaturas, los huecos se rellenaron con una interpolación de la biblioteca Pandas.
- v. Podemos concluir, entonces, dados los resultados obtenidos, que el modelo ARI-MA, con gran precisión, nos fue de utilidad en el pronóstico de las series de tiempo. Los pronósticos se anexan a continuación.

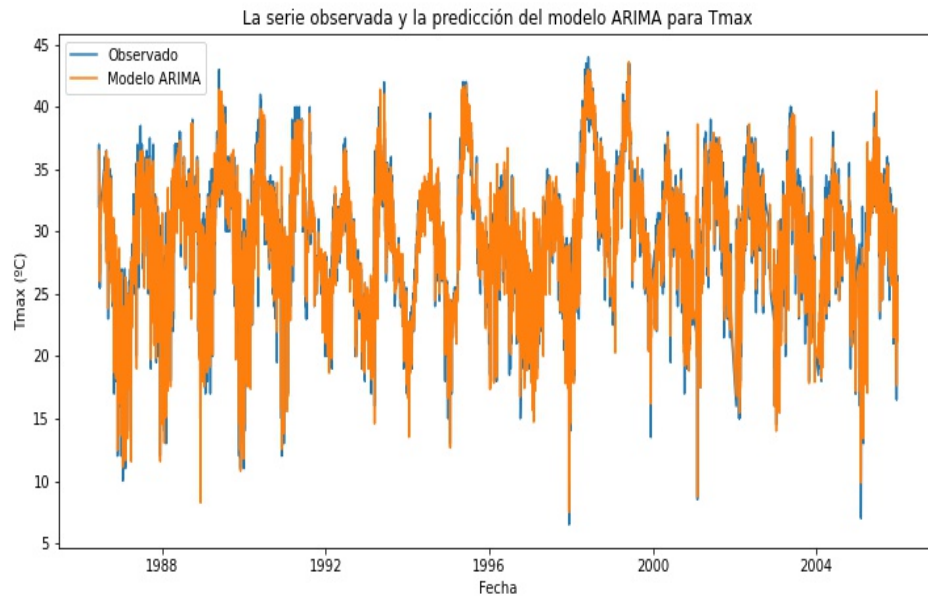


Figura 1: Datos originales contra pronóstico de ARIMA de Temperaturas Máximas

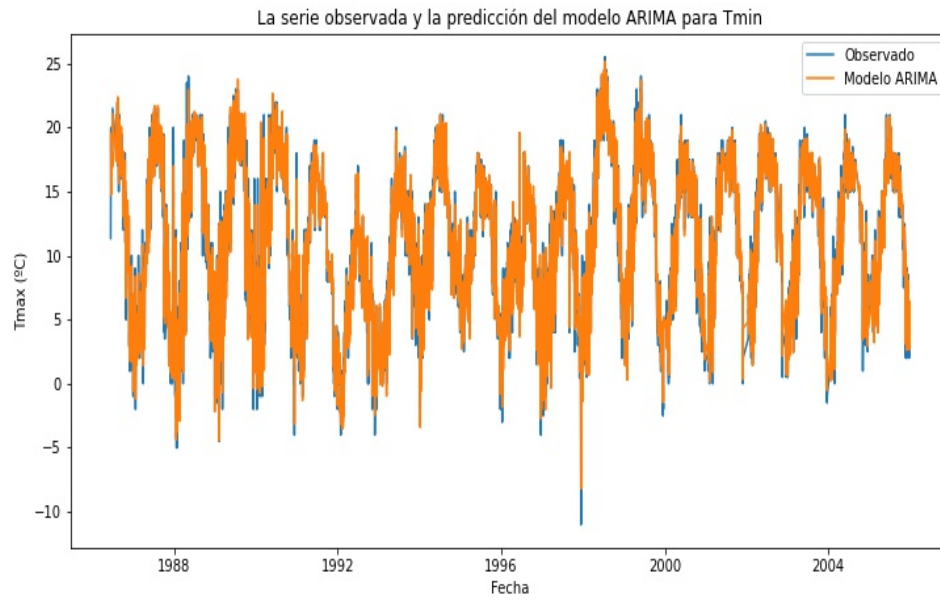


Figura 2: Datos originales contra pronóstico de ARIMA de Temperaturas Mínimas

3. Retroalimentación y comentarios

Me pareció una actividad sumamente interesante, pero confusa de realizar debido a la teoría estadística involucrada, que aún me es desconocida. Me parece que, personalmente, debo continuar investigando las bibliotecas utilizadas, principalmente *Statsmodels*, *Matplotlib*, *Pandas*, *Numpy*. Ahora bien, respecto de las cuestiones de retroalimentación:

1. Me pareció un tema algo complicado, pero interesante, sobre todo pensar en el pronóstico de las temperaturas y el análisis estadístico que conlleva. No obstante, fue este mismo el que hizo a la actividad confusa y compleja, pues se basa en aspectos de estadística avanzados respecto de lo que conocemos.
2. Creo que fue un reto algo complicado y confuso.
3. La parte que más dificultad me presentó fue comprender el sentido de los comandos y la razón de las acciones que se llevaron a cabo para obtener el resultado final. Cabe mencionar que el profesor explicó los temas lo mejor posible, sin embargo, la falta de conocimiento era abismal, pues apenas estoy llevando el curso de estadística.
4. El aburrimiento en esta ocasión, para mí, fue el tema del análisis de los datos: de variables meteorológicas.
5. Para la mejora de la actividad, diría, incluir una explicación más detallada del modelo y la teoría estadística detrás, antes de enfrentarnos al problema del pronóstico de series de tiempo mediante bibliotecas de Python.
6. Avanzado. Creo que lo amerita por las aplicaciones que se pueden obtener.