

## Análisis de la serie de tiempo de Tuberculosis en el Estado de Veracruz de 2000 a 2018 Time series analysis of tuberculosis in the state of Veracruz from 2000 to 2018

S. I. Vázquez-Chena <sup>a</sup>, B. Tapia-Santos <sup>a</sup>, R. Ávila-Pozos <sup>b,\*</sup>

<sup>a</sup>Departamento de Matemáticas, Universidad Veracruzana, 91097, Xalapa, Veracruz, México.

<sup>b</sup>Área Académica de Matemáticas y Física, Universidad Autónoma del Estado de Hidalgo, 42184, Mineral de la Reforma, Hidalgo, México.

### Resumen

En este trabajo se revisa la información de la Dirección General de Epidemiología, particularmente los casos registrados de Tuberculosis pulmonar en el Estado de Veracruz, desde el año 2000 y hasta el año 2018. Esta información, vista como una serie de tiempo, se analiza usando uno de los enfoques más empleados para el estudio de series temporales. Se revisa el comportamiento de la serie para encontrar si esta tiene algún tipo de tendencia constante o estacional, se realizan pruebas de estacionalidad y finalmente se obtienen modelos integrados autoregresivos de medias móviles, a fin de realizar pronósticos. Como parte de los resultados, se muestran los parámetros de los mejores modelos ARIMA obtenidos, así como los valores de los criterios de información de Akaike y Bayesiano, para las series reportadas por mes y por semana epidemiológica. Para el conjunto de datos registrados por mes, el AIC fue 2111.43 y el BIC fue de 2131.98, mientras que para los datos por semana, el AIC fue de 6899.45 y el BIC fue de 6938.61. Además, al realizar el pronóstico de casos se observó que, para el modelo correspondiente a los datos por semana epidemiológica, se obtiene una diferencia menor entre los casos de tuberculosis pulmonar que se pueden presentar.

**Palabras Clave:** Análisis de series de tiempo, Estacionaridad, Tuberculosis pulmonar, Modelos ARIMA, Pronósticos.

### Abstract

In this work, the information of the General Directorate of Epidemiology is reviewed, particularly the registered cases of pulmonary tuberculosis in the State of Veracruz, from the year 2000 to the year 2018. This information, seen as a time series, is analyzed using one of the most widely used approaches for the study of time series. The behavior of the series is reviewed to find out if it has some type of constant or seasonal trend, seasonality tests are carried out and finally integrated autoregressive models of mobile means are obtained, in order to make forecasts. As part of the results, the parameters of the best ARIMA models obtained are shown, as well as the values of Akaike and Bayesian information criteria, for the series reported by month and by epidemiological week. For the data set recorded by month, the AIC was 2111.43 and the BIC was 2131.98, while for the data by week, the AIC was 6899.45 and the BIC was 6938.61. In addition, forecast of cases shown that, for the model corresponding to the data per epidemiological week, a smaller difference is obtained between the cases of Pulmonary Tuberculosis that may occur.

**Keywords:** Time Series Analysis, Stationarity, Pulmonary Tuberculosis Disease, ARIMA Models, Forecast.

### 1. Introducción

La tuberculosis (TB) es una infección causada por la bacteria *Mycobacterium tuberculosis*, esta bacteria puede afectar cualquier órgano del cuerpo humano, tal como: sistema nervioso, huesos, piel, intestinos, genitales, ganglios linfáticos, pulmones, etc., siendo este último donde ataca con mayor frecuencia originando así tuberculosis pulmonar (TBP). De todas las

formas en las que se presenta la tuberculosis únicamente la pulmonar es contagiosa, el contagio se produce cuando una persona infectada de TBP tose, platica, o estornuda y al hacerlo elimina bacterias al aire que pueden entrar a los pulmones de otras personas (INSP, 2019).

Aunque hoy en día la tuberculosis es una enfermedad prevenible y curable, se sigue considerando un problema importante

\* Autor para correspondencia: ravila@uaeh.edu.mx.

**Correo electrónico:** sobeidachena@gmail.com (Sobeida Itzel Vázquez-Chena), btapia@uv.mx (Brenda Tapia-Santos), ravila@uaeh.edu.mx (Roberto Ávila-Pozos).

de salud pública en todo el mundo. En 2019 un total de 1.4 millones de personas murieron por esta causa y, por su parte, la Organización Mundial de la Salud (OMS) informa que a nivel mundial un tercio de la población está infectada por *Mycobacterium Tuberculosis* y que entre 5% y 10% de estas personas desarrollarán la enfermedad en algún momento de su vida. Los enfermos de TB suelen manifestar algunos síntomas tales como: tos, fiebre, sudores nocturnos y pérdida de peso principalmente. Estas manifestaciones pueden ser leves por muchos meses por lo que los pacientes tardan en buscar atención médica y, en ese lapso, pueden transmitir la enfermedad a otras personas. Se sabe que a lo largo de un año un enfermo con tuberculosis es capaz de infectar entre 10 y 15 personas y, si no reciben algún tratamiento, pueden morir en el transcurso de 5 años (SSA, 2014).

En la región de América, en 2010 se identificaron 10 países que concentraban más del 80% de la carga de tuberculosis, México ocupó el cuarto lugar por el número de casos después de Brasil, Perú y Haití. La incidencia de la TB en México para el año 2012 fue de 16.8 por cada 100 mil habitantes, de 19,697 casos, 15,858 fueron de localización pulmonar afectando mayormente a la población entre los 15 y los 49 años de edad, y 2,253 personas murieron por esta causa, con una tasa de 1.9 por cada 100 mil habitantes. En ese mismo año las entidades federativas con mayor número de mortalidad por TB fueron: Baja California, Sonora, Nayarit, Chiapas, Nuevo León, Guerrero y Veracruz (SSA, 2014). En 2019 se identificaron más de 19 mil casos nuevos de TB respiratoria y poco más de 4 mil casos de otras formas de tuberculosis, siendo más frecuente en el grupo de 25 a 44 años. Además, se registraron 1,982 defunciones por esta patología; el 89.2% correspondió a la forma pulmonar, 5.4% del sistema nervioso y 5.4% por secuelas de tuberculosis (INSP, 2021).

La tuberculosis se puede manifestar por signos y síntomas pulmonares o extrapulmonares, los individuos cuya sintomatología, signos físicos, elementos auxiliares de diagnóstico y respuesta terapéutica sugieren la existencia de tuberculosis pulmonar, se les denomina diagnosticados y de estos, quienes resultan positivos en estudios de Baciloscopia, Cultivo o Histología para TBP se denominan casos confirmados, en caso contrario se les denominan no confirmados. México ha asumido una serie de compromisos en salud compatibles con la realidad y en concordancia con organismos internacionales encaminados a abatir enfermedades que aún siguen considerándose flagelos de la modernidad: paludismo, infección por VIH y tuberculosis, entre otras (SSA, 2008).

La globalización ha permitido grandes avances en el diagnóstico, tratamiento y prevención de algunas enfermedades que causan alta morbilidad y mortalidad, un ejemplo palpable es el avance en el conocimiento de la interrelación de *Mycobacterium tuberculosis* y el hospedero mediante la biología molecular, la bioingeniería en vacunas, la farmacodinámica y el aporte de la microbiología que, en conjunto, han impactado en la evolución de los enfermos de tuberculosis. La Secretaría de Salud en México se ha dado a la tarea de fortalecer el Programa Nacional de Tuberculosis (PNT) mediante el diseño de herramientas que permitan o faciliten la prevención, el diagnóstico y el tratamiento, además de medir el grado de cumplimiento y

el impacto de las acciones suficientes del programa en los diferentes escenarios que se le presenten al servidor de la salud.

La vacuna BCG (Bacilo Calmette-Guérin) es una herramienta de prevención contra la TB, fue utilizada por primera vez para inmunizar a seres humanos en 1921, dando inicio a la vacunación en forma masiva a partir de 1930. En 1951, México adopta los lineamientos para la implementación del control de infecciones por tuberculosis en la Américas desarrollado por la OMS (Organización Mundial de la Salud). Para el año 1988, la OMS incluye la vacuna BCG en el programa salud para todos; esta vacuna ha demostrado ser efectiva para prevenir la tuberculosis pulmonar en un 50%.

Otra herramienta contra la TB para su control es el tratamiento antituberculoso. Los objetivos del tratamiento son interrumpir la cadena de transmisibilidad de *M. tuberculosis*, lograr la curación, prevenir las complicaciones y evitar la muerte. Un tratamiento bien guiado tiene alta probabilidad de éxito. Los tratamientos pueden variar según sea el estado del individuo, sin embargo, para aquellos que presentan la enfermedad por primera vez se aplica el tratamiento primario que consta de una fase intensiva con duración promedio de 2 meses o hasta completar 60 dosis de los fármacos: Isoniazida (INH), Rifampina (RIF), Etambutol (EMB), Pirazinamida (PZA). Una vez concluida esta fase se continúa con la fase de sostén que tiene una duración de 4 meses o hasta completar 45 dosis de Isoniazida (INH) y Rifampina (RIF) (SSA, 2008).

## 2. Antecedentes

Una forma de estudiar el comportamiento pasado e intentar entender cómo podrá ser en el futuro de la dinámica de alguna enfermedad, es analizando la información que se ha ido recopilando en el transcurso de los años. Esto se puede hacer mediante análisis de series de tiempo.

Los métodos comúnmente utilizados para el análisis de series de tiempo, incluyen análisis espectral, métodos de suavizado exponencial y el modelo autorregresivo integrado de medias móviles (ARIMA).

En Chen (Chen et al., 2013) identificaron un modelo autorregresivo integrado de medias móviles (ARIMA) para predecir la incidencia de tuberculosis pulmonar. El modelo de serie de tiempo que desarrollaron es univariante, en él emplearon las funciones de autocorrelación muestral (ACF) y autocorrelación parcial (PACF) para identificar el modelo ARIMA (SARIMA) del orden apropiado. Las estimaciones de los parámetros del modelo se obtuvieron mediante el método de máxima verosimilitud y, para comparar la bondad de ajuste entre los modelos ARIMA, utilizaron controles de diagnóstico que incluían análisis de residuales, los criterios de información de Akaike (AIC) y el criterio de información Bayesiana (BIC).

En Kohei (Kohei et al., 2016) se analizaron series de tiempo por grupos de edad de 0 – 9, 10 – 19, . . . , 70 – 79 y mayores o iguales de 80 años. Realizaron un análisis espectral con el método de máxima entropía (MEM) y el método de mínimos cuadrados (LSM).

En la búsqueda de un modelo adecuado para representar una serie de tiempo de tuberculosis, en Kumar (Kumar et al., 2014) realizaron ajustes usando el software 21 donde, para verificar la estacionariedad de los datos, emplearon la función de

autocorrelación (ACF) y la función de autocorrelación parcial (PACF). Mientras que, para determinar el pico de variación estacional, usaron el factor de ajuste estacional (SAF). Y la prueba de Ljung-Box (Box-Pierce modificado) para determinar si el modelo fue especificado correctamente.

Otra forma de analizar una serie tiempo para tuberculosis la presentan en Li (Li et al., 2020), quienes utilizaron el análisis de varianza para comparar la amplitud estacional de la TB entre subgrupos, a fin de identificar los factores responsables de la variación estacional. De esta manera, encontraron que los factores relacionados con la variación estacional fueron la edad y la ocupación. Emplearon el modelo X12-ARIMA que engloba dos procedimientos: X11 y RegARIMA. El procedimiento X11 lo usaron para descomponer la serie de tiempo original y RegARIMA se usó para construir el modelo ARIMA. Además, emplearon el análisis de varianza y el método de Student-Newman-Keuls para comparar la amplitud estacional en los diferentes subgrupos.

A fin de estudiar los patrones temporales y estacionales de la incidencia de TB en Pakistán, Khaliq y colaboradores (Khaliq et al., 2015) realizaron un estudio retrospectivo de los datos obtenidos del programa de tratamiento corto de observación directa, agrupándolos por sexo, en siete grupos de edad. Para la relación temporal de TB, realizaron un análisis de correlación y regresión entre el número trimestral de casos de TB notificados y la temperatura media trimestral. Usaron un modelador experto del software SPSS (versión 20) para ajustar el modelo más adecuado para los datos de la serie de tiempo. La estacionariedad de los datos la verificaron mediante la función de autocorrelación y la función de autocorrelación parcial, para determinar si el modelo se especificó correctamente, emplearon la prueba Ljung-Box (Box-Pierce modificada).

En Wang (Wang et al., 2018) realizaron un análisis de la serie de tiempo de la tasa de notificaciones de TB mensual, entre 2005 y 2017, en China. El análisis se basó en un modelo SARIMA, y en un modelo híbrido de modelo de red neuronal de regresión generalizada SARIMA (GRNN), que considera información lineal y no lineal para evaluar la estacionalidad de la TB a nivel nacional. Hicieron el suavizado de la serie de tiempo para discernir patrones subyacentes debido a las variaciones de alta frecuencia de la serie original. Utilizaron el método Dickey-Fuller aumentado (ADF) para determinar si la secuencia era estacionaria y se podría adoptar una transformación y/o diferenciación logarítmica, si la secuencia no fuera estacionaria. Emplearon el coeficiente de autocorrelación (ACF) y el coeficiente de autocorrelación parcial (PACF) de la secuencia estacionaria, para identificar los parámetros opcionales del modelo, a fin de establecer uno o más modelos alternativos. Con el criterio de información de Akaike (AIC) y el criterio de información bayesiana (BIC) se seleccionó el mejor modelo SARIMA. El error cuadrático medio y el coeficiente de determinación ( $R^2$ ) fueron utilizados para evaluar la precisión del modelo elegido.

Para estudiar la naturaleza de los datos, en Moosazadeh (Moosazadeh et al., 2015) examinaron el tipo de serie en relación con el diagnóstico estacionario y no estacionario de la media, la varianza y la tendencia estacional. Utilizaron la prueba de Bartlett para estudiar la igualdad de varianza y la prueba de Kolmogorov-Smirnov para analizar la normalidad de los datos. Dado que la serie obtenida no tenía condiciones de esta-

cionariedad en su promedio, utilizaron la diferenciación de una etapa para eliminar la tendencia y crear series estacionarias. Los parámetros desconocidos fueron estimados a partir de la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF), antes y después de diferenciar, sin efectos estacionales y con efectos estacionales.

En otro estudio, cuyo propósito fue obtener un modelo de series de tiempo para estimar la incidencia de casos de TB notificados en el Hospital de Korle-Bu, entre 2008 y 2017, Aryee y colaboradores (Aryee et al., 2018) establecieron y compararon varios modelos para encontrar el mejor, basado en el criterio de información de Akaike y el criterio de información bayesiano. Emplearon el enfoque de series de tiempo de Box-Jenkins presentado como modelo autorregresivo integrado de medias móviles (ARIMA). La metodología de Box-Jenkins comprende la identificación del modelo, la estimación de parámetros, el diagnóstico del modelo y la previsión del mismo. La estacionariedad fue evaluada y confirmada por la prueba de Dickey-Fuller aumentada (ADF) en los datos transformados. Graficaron una función de autocorrelación (ACF) y una función de autocorrelación parcial (PACF) para obtener los órdenes  $p$  y  $q$  del modelo autorregresivo y de medias móviles, respectivamente. El modelo que obtuvieron se comparó con otros modelos ARIMA, resultando el mejor el de menor valor de AIC y de BIC. Realizaron pruebas de diagnóstico en el mejor modelo, mediante un análisis de residuales, evaluando la normalidad e independencia de estos, mediante la gráfica cuantil-cuantil y se confirmó mediante la prueba de Shapiro-Wilk. Usaron el mejor modelo para pronosticar el número estimado de casos de TB mensuales.

En Willis (Willis et al., 2012) realizaron un análisis de la descomposición de series de tiempo de los casos de tuberculosis notificados a los Centros para el Control y la Prevención de Enfermedades desde 1993 hasta 2008, en Estados Unidos. Los recuentos de casos mensuales fueron analizados con el método X-11. El componente estacional fue aislado de la tendencia general y los factores irregulares (ruido) mediante un modelo multiplicativo. Obtuvieron una estimación preliminar de la tendencia promediando los valores logarítmicos en una ventana móvil de 13 meses, dicha tendencia preliminar fue eliminada de la serie original, dejando los componentes estacionales e irregulares combinados. La presencia de estacionalidad fue evaluada mediante la prueba combinada X-11, que combina los estadísticos F estable y móvil con la prueba de Kruskal-Wallis. Calcularon el mes de menos casos y, la amplitud estacional anual, con intervalos de confianza (IC) del 95 % para las poblaciones con estacionalidad identificable. Los intervalos de confianza fueron calculados utilizando el método de Wald basado en la varianza alrededor de la media, con las 16 mediciones de amplitud anuales.

En un estudio para analizar la estacionalidad de los datos sobre los casos notificados de tuberculosis, reportados en el anuario de vigilancia de enfermedades infecciosas 2017, publicado por los Centros para el Control y Prevención de Enfermedades en Corea, Kim y Bae (H. and Bae, 2018) estudiaron la estabilidad de los datos estadísticos y obtuvieron el número de notificaciones por mes durante 10 años. Para evaluar la estacionalidad, la variable que representaba la estación fue construída al dividir el año en períodos de 3 meses. Consideraron que para los modelos de series de tiempo de descomposición estacional, existen

enfoques multiplicativos y enfoques aditivos. Debido a que los datos de su estudio tenían una variación estacional consistente a lo largo del tiempo, emplearon el enfoque aditivo. Utilizaron el método de media móvil simple, para que el ancho de los datos para cada período fuera igual al número de semanas y se agregaran de manera idéntica en todos los puntos de tiempo para ajustar la variación estacional.

En este trabajo se analizaron las series de tiempo correspondientes a los casos de tuberculosis pulmonar durante el período de 2000 a 2018, en el estado de Veracruz, registrados por años, meses y semanas epidemiológicas. Se analizaron las componentes de la series, tales como: estacionaridad, tendencia y aleatoriedad. Se determinaron modelos autorregresivos integrados de medias móviles (ARIMA) bajo la metodología Box-Jenkins, para pronosticar la incidencia de tuberculosis pulmonar en el estado. Para comprobar el modelo que se determina del estudio de las funciones de autocorrelación simple y parcial, se utiliza la prueba de Ljung-Box. Presentamos los parámetros de los modelos, así como el valor de los criterios de información de Akaike y Bayesiano.

### 3. Datos

A continuación se muestran los datos de los casos de Tuberculosis pulmonar registrados en el estado de Veracruz durante el período de 2000 a 2018. Los datos se obtuvieron de la Dirección General de Epidemiología (antes Sistema Nacional de Vigilancia Epidemiológica). En esa base de datos, los registros se pueden consultar por año, mes o semana epidemiológica, como se puede apreciar en las figuras 1, 2 y 3. Esta base de datos también permite consultar los registros por grupo de edad, por sexo, por jurisdicción sanitaria, entre otras opciones. El análisis de la serie de tiempo entonces puede considerar todos estos casos. En este trabajo, sólo consideramos el total de casos registrados por mes y por semana epidemiológica.

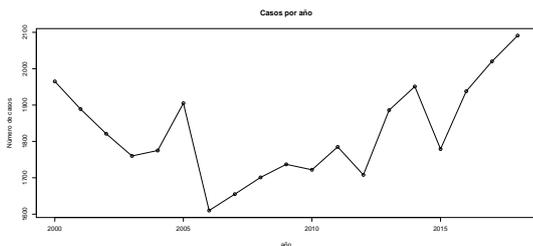


Figura 1: Número de casos de tuberculosis pulmonar por año en el estado de Veracruz.

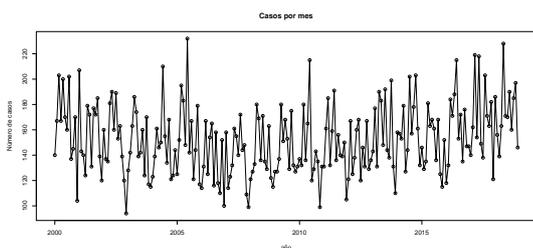


Figura 2: Número de casos de tuberculosis pulmonar por mes en el estado de Veracruz.

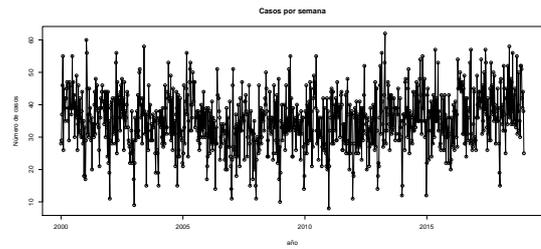


Figura 3: Número de casos de tuberculosis pulmonar por semana epidemiológica en el estado de Veracruz.

### 4. Análisis

Los análisis estadísticos de esta sección se realizaron en R versión 4.1.1 (R Core Team, 2021a) y se utilizaron las siguientes librerías: *astsa* versión 1.13 (Stoffer, 2021), *foreign* versión 0.8-81 (R Core Team, 2020), *quantmod* versión 0.4.18 (Ryan and Ulrich, 2020), *tseries* versión 0.10-48 (Trapletti and Hornik, 2020), *forecast* (RJ and Y., 2008), *stats* (R Core Team, 2021b), *lubridate* (Grolemund and Wickham, 2011), *tidyverse* (Wickham et al., 2019).

Una secuencia de variables aleatorias  $\{Y_t : t = 0, \pm 1, \pm 2, \dots\}$  se conoce como proceso estocástico, y sirve como un modelo para una serie de tiempo. Se sabe que la estructura probabilística completa de tal proceso está determinada por el conjunto de distribuciones de toda la colección finita de  $Y_t$ 's. Afortunadamente, no se requiere tratar explícitamente con esas distribuciones multivariadas. Gran parte de la información en esas distribuciones conjuntas pueden describirse en términos de medias, varianzas y covarianzas.

Consideremos la situación más sencilla, donde asumimos una función media  $\mu_t$ . El modelo puede escribirse como

$$Y_t = \mu_t + X_t \tag{1}$$

donde  $E[X_t] = 0$ .

Considere la tendencia determinista expresada como

$$\mu_t = \beta_0 + \beta_1 t. \tag{2}$$

La suposición más general si los datos tienen un comportamiento estacional es que se propongan parámetros constantes  $(\beta_1, \beta_2, \dots, \beta_{12})$ . En este caso, la tendencia estacional se puede escribir como

$$\mu_t = \begin{cases} \beta_1 & \text{para } t = 1, 13, 25, \dots \\ \beta_2 & \text{para } t = 2, 14, 26, \dots \\ \vdots & \vdots \\ \beta_{12} & \text{para } t = 12, 24, 36, \dots \end{cases} \tag{3}$$

En la figura 4 se muestra la descomposición de la serie de tiempo, suponiendo que es de la forma mencionada en la ecuación 1, por lo que se puede ver la tendencia, la componente estacional y la parte estocástica, llamada aquí residuales.

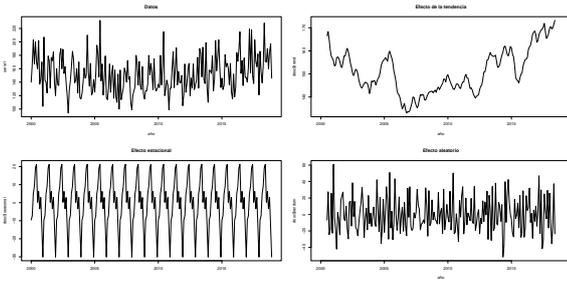


Figura 4: Descomposición de la serie de tiempo, por meses, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

La componente estocástica  $X_t$  se puede estimar mediante el residual  $\hat{X}_t = Y_t - \hat{\mu}_t$ . Si la tendencia del modelo es razonablemente correcta, los residuales deberían tener el comportamiento del verdadero componente estocástico. La figura 4 muestra parte del estudio descriptivo de la serie de tiempo bajo sus componentes de variación básicas: tendencia, estacionalidad y aleatoriedad, mientras que el diagrama de residuales mostrado en la figura 5 corresponde a los llamados residuales estandarizados, que realizan la mayoría de los programas.

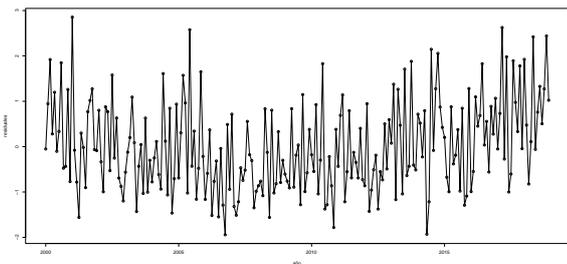


Figura 5: Diagrama de los residuales de la serie de tiempo, por meses.

La normalidad se puede verificar al graficar los cuantiles de los datos contra los cuantiles teóricos de una distribución normal, como se muestra en la figura 6. Si los datos tienen una distribución normal, el diagrama Q-Q representará aproximadamente una línea recta. El valor de la Kurtosis para la serie de tiempo por semana es 0.2166654, y en la prueba de Jarque Bera se obtuvo un p-valor de 0.312 Para la serie de tiempo por mes es de -0.356772 y en la prueba Jarque Bera el p-valor fue de 0.01952.

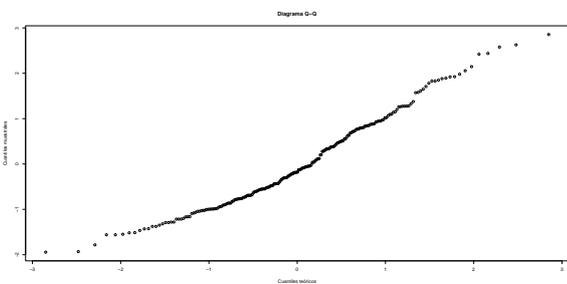


Figura 6: Diagrama cuantil-cuantil de los residuales de la serie de tiempo, por meses, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

Una herramienta de diagnóstico muy importante para examinar la dependencia es la función de autocorrelación. Si supo-

nemos que la serie de tiempo es estacionaria, es decir, se tiene una media y varianza común para la serie, se puede definir la función de autocorrelación muestral como

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \tag{4}$$

La gráfica presentada en la figura 7 presenta el valor de  $r_k$  contra el retardo  $k$ . A este diagrama se le conoce como correlograma.

Dado que se requiere determinar el orden de los modelos autoregresivos, se introduce una función definida como la correlación entre  $Y_t$  y  $Y_{t-k}$ , después de remover el efecto de las variables  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ . Este coeficiente se conoce como autocorrelación parcial al retardo  $k$ , y se denota por  $\phi_{kk}$ . La función de autocorrelación parcial, definida como

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_j}, \tag{5}$$

donde  $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j}$ , para  $j = 1, 2, \dots, k - 1$ .

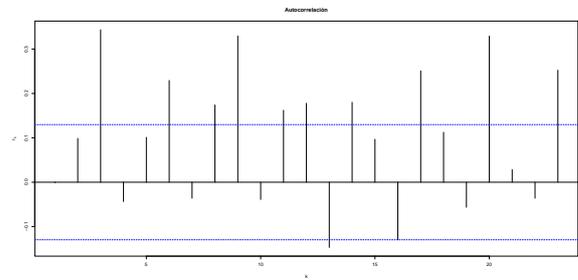


Figura 7: Función de autocorrelación de la serie de tiempo, por meses, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

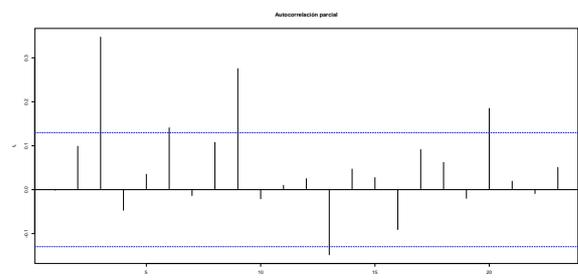


Figura 8: Función de autocorrelación parcial de la serie de tiempo, por meses, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

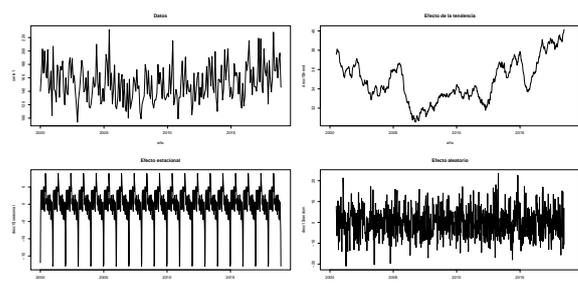


Figura 9: Descomposición de la serie de tiempo, por semana, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

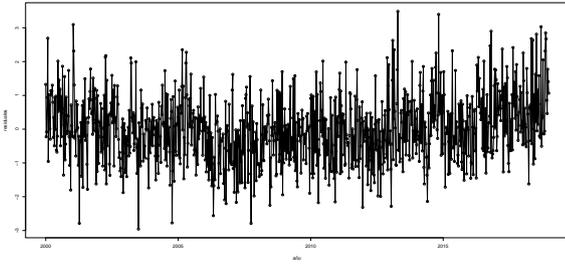


Figura 10: Diagrama de los residuales de la serie de tiempo, por semana.

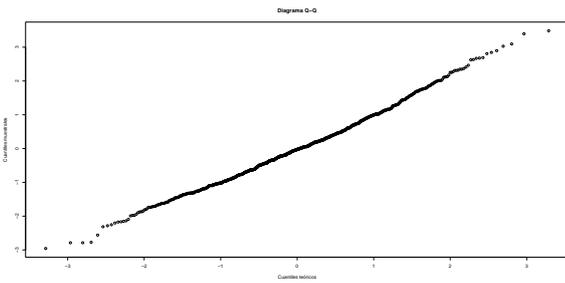


Figura 11: Diagrama cuantil-cuantil de los residuales de la serie de tiempo, por semana, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

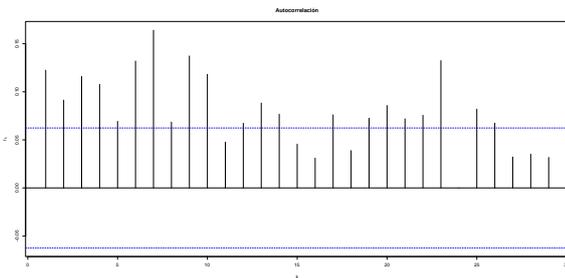


Figura 12: Función de autocorrelación de la serie de tiempo, por semana, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

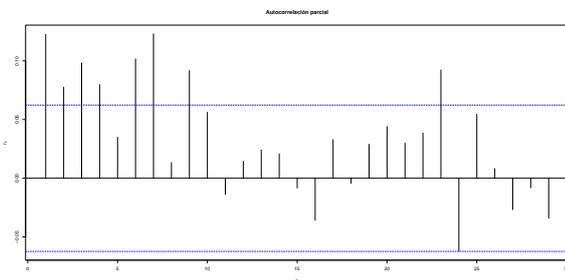


Figura 13: Función de autocorrelación parcial de la serie de tiempo, por semana, del número de casos de tuberculosis pulmonar en el estado de Veracruz.

Decimos que una serie de tiempo  $\{Y_t\}$  se puede representar mediante una modelo integrado autoregresivo de medias móviles (*ARIMA*), si la  $d$ -ésima diferencia  $W_t = \nabla^d Y_t$  es un proceso estacionario autoregresivo de medias móviles (*ARMA*). Si  $\{W_t\}$  se representa con un modelo *ARMA*( $p, q$ ), se dice que el proceso  $\{Y_t\}$  es un proceso *ARIMA*( $p, d, q$ ).

Los modelos *ARIMA*( $p, q, d$ ) pueden operar en series de tiempo tanto estacionarias como no estacionarias, ya que incluyen el proceso de estabilización de la serie por medio del parámetro  $d$ , que representa el grado de diferenciación aplicado a la serie. Los órdenes ( $p, q$ ), corresponden al modelo autorregresivo *AR* y al modelo de medias móviles *MA*, respectivamente. En la figura 7 la función de autocorrelación simple muestra una espiga significativa en el retardo  $k = 3$ , lo que indica una media móvil, es decir,  $q = 1$ . Mientras que, la función de autocorrelación parcial en la figura 8 muestra también una espiga significativa en  $k = 3$ , originando un autoregresivo,  $p = 1$ . Lo anterior daría lugar al modelo *ARMA*(1, 1) siempre que la serie de tiempo fuera estacionaria o el modelo *ARIMA*(1, 1, 1), en relación a la serie de tiempo en meses.

Por otro lado, considerando la serie de tiempo por semana epidemiológica, la función de autocorrelación simple en la figura 12, muestra una espiga significativa en  $k = 7$ , indicando una media móvil,  $q = 1$ . Para la función de autocorrelación parcial en la figura 13, se tiene una espiga significativa para  $k = 1$  por lo que, se tendría un autoregresivo,  $p = 1$ , originando el modelo *ARMA*(1, 1) o el modelo *ARIMA*(1, 1, 1).

Para la serie de tiempo que se reporta por mes, utilizando la función `auto.arima` de **R**, obtuvimos un modelo *ARIMA*(1,1,2) con los siguientes parámetros:

AR1	MA1	MA2	AIC	BIC
-0.9259	0.0214	-0.9389	2111.43	2131.98

La figura 14 ilustra el pronóstico realizado con este modelo *ARIMA*.

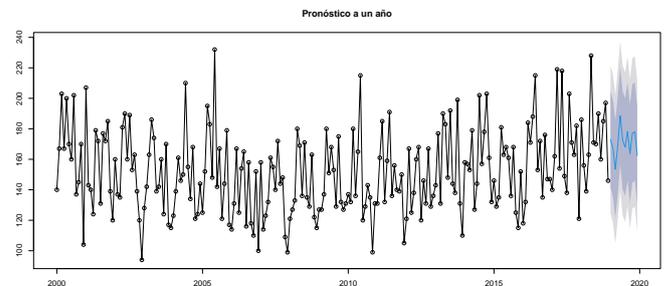


Figura 14: Parámetros estimados para la tendencia estacional de la serie de tiempo de datos mensuales.

Para la serie de tiempo que se reporta por semana epidemiológica, obtuvimos un modelo *ARIMA*(2,1,2) con los siguientes parámetros:

AR1	AR2	MA1	MA2	AIC	BIC
0.9289	-0.0706	-1.7562	0.7586	6899.45	6938.61

La figura 15 ilustra el pronóstico realizado con este modelo *ARIMA*. Los pronósticos muestran una diferencia menor entre los casos de tuberculosis que se pueden presentar por mes, con picos de incidencia menos pronunciados en ambas gráficas.

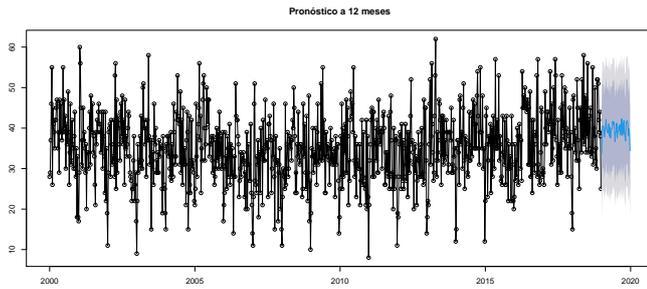


Figura 15: Parámetros estimados para la tendencia estacional de la serie de tiempo de datos mensuales.

El criterio de información de Akaike es una herramienta ampliamente utilizada en modelos estadísticos. Se emplea para la selección de un modelo entre varias propuestas que emplean ajustes con máxima verosimilitud. Por otra parte, el modelo favorecido por el criterio de información Bayesiano corresponde al modelo que se vuelve más plausible por los datos disponibles (Cavanaugh and Neath, 2019).

## 5. Conclusiones

A partir de los casos registrados de Tuberculosis pulmonar en el Estado de Veracruz entre los años 2000 y 2008, se obtuvieron los parámetros de dos modelos ARIMA, uno de ellos para la serie de tiempo con reporte mensual y el otro para la serie de tiempo con reporte por semana epidemiológica. Para cada uno de estos modelos se realizó el pronóstico de casos observándose que, para el modelo correspondiente a los datos por semana epidemiológica, se obtienen picos de incidencia menos pronunciados, es decir, una diferencia menor entre los casos de TBP que se pueden presentar.

Los modelos ARMA y ARIMA proporcionan importantes ventajas en el análisis de series de tiempo. Los pronósticos pueden ser muy útiles para determinar un posible comportamiento en la incidencia ocasionada por la enfermedad. Esto permite tomar acciones anticipadas para el manejo y control de la misma. Particularmente se obtienen predicciones sobre los casos de tuberculosis pulmonar en el Estado de Veracruz partiendo de los registros del periodo 2000 - 2018 por mes y semana epidemiológica.

## Agradecimientos

Agradecemos a los revisores sus comentarios y observaciones, que dieron forma a la versión final de este trabajo.

Este trabajo ha sido realizado gracias al apoyo del Consejo Nacional de Ciencia y Tecnología, mediante una beca a SIVC.

## Referencias

- Aryee, G., E., K., Essuman, R., Agyei, A. N., Kudzawu, S., Djagbletey, R., Darkwa, E. O., and Forson, A. (2018). Estimating the incidence of tuberculosis cases reported at a tertiary hospital in Ghana: a time series model approach. *BMC Public Health*, 18:1292.
- Cavanaugh, J. E. and Neath, A. A. (2019). The akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *WIREs Computational Statistics*, 11:e1460.
- Chen, Y., Wu, A., Wang, C., Zhou, H., and Feng, S. (2013). Time series analysis of pulmonary tuberculosis incidence: Forecasting by applying the time series model. *Advanced Materials Research*, 709:819–822.
- Grolemund, G. and Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25.
- H., K. E. and Bae, J.-M. (2018). Seasonality of tuberculosis in the republic of Korea, 2006-2016. *Epidemiology and Health*, 40:e2018051.
- INSP (2019). Día mundial de la tuberculosis 2019. último acceso 22 de febrero de 2022.
- INSP (2021). Tuberculosis: una epidemia que debería quedar en el pasado. último acceso 22 de febrero de 2022.
- Khaliq, A., Batool, S. A., and Chaudhry, M. N. (2015). Seasonality and trend analysis of tuberculosis in Lahore, Pakistan from 2006 to 2013. *Journal of Epidemiology and Global Health*, 5(4):397–403.
- Kohci, Y., Sumi, A., and Kobayashi, N. (2016). Time-series analysis of monthly age-specific numbers of newly registered cases of active tuberculosis in Japan from 1998 to 2013. *Epidemiol. Infect.*, 144:2401–2414.
- Kumar, V., Singh, A., Adhikary, M., Daral, S., Khokhar, A., and Singh, S. (2014). Seasonality of tuberculosis in Delhi, India: A time series analysis. *Tuberculosis Research and Treatment*, 2014:1–5.
- Li, Y., Zhu, L., Lu, W., Chen, C., and Yang, H. (2020). Seasonal variation in notified tuberculosis cases from 2014 to 2018 in eastern China. *Journal of International Medical Research*, 48(8):1–11.
- Moosazadeh, M., Khanjani, N., Nasehi, M., and Bahrapour, A. (2015). Predicting the incidence of smear positive tuberculosis cases in Iran using time series analysis. *Irani Journal of Public Health*, 44(11):1526–1534.
- R Core Team (2020). *foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...* R package version 0.8-81.
- R Core Team (2021a). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- R Core Team (2021b). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RJ, H. and Y., K. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.
- Ryan, J. A. and Ulrich, J. M. (2020). *quantmod: Quantitative Financial Modeling Framework*. R package version 0.4.18.
- SSA (2008). Estándares para la atención de la tuberculosis en México. último acceso 22 de febrero de 2022.
- SSA (2014). Prevención y control de la tuberculosis 2013-2018. último acceso 22 de febrero de 2022.
- Stoffer, D. (2021). *astsa: Applied Statistical Time Series Analysis*. R package version 1.13.
- Trapletti, A. and Hornik, K. (2020). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-48.
- Wang, H., Tian, C. W., Wang, W. M., and Luo, X. M. (2018). Time-series analysis of tuberculosis from 2005 to 2017 in China. *Epidemiology and Infection*, 146(8):935–939.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
- Willis, M. D., Winston, C. A., Heilig, C. M., Cain, K. P., Walter, N. D., and Mac Kenzie, W. R. (2012). Seasonality of tuberculosis in the United States, 1993-2008. *Clinical Infectious Diseases*, 54(11):1553–1560.