

# Pronóstico epidemiológico en México entre 2005 y 2015

Alberto Benavides

*Posgrado de Ingeniería de Sistemas  
Facultad de Ingeniería Mecánica y Eléctrica  
Universidad Autónoma de Nuevo León  
Nuevo León, México*

---

## Abstract

Se clasifican entre no pronosticables y pronosticables focos epidemiológicos a partir de series de tiempo semanales proporcionadas por la Secretaría de Salud de México durante 2005 a 2015 mediante análisis de componentes temporales y preprocesamiento por tipo de estacionalidad presente en tales series de tiempo.

---

## 1. Introducción

Una reciente investigación de maestría realizada por Alberto Benavides [4] toma los registros epidemiológicos semanales publicados por la Secretaría de salud de México durante el periodo 2005–2015 [10] y los agrupa por  $k$ -medias, un algoritmo estadístico no supervisado, a partir de las autocorrelaciones con retraso semanal de hasta un año. Dicho trabajo constituye una primera aproximación a esos datos debido a que antes estaban solamente contenidos en documentos PDF, impidiendo su análisis. Con los datos actualmente disponibles para analizar desde diversas aproximaciones científicas, su interés se vuelve relevante puesto que estos datos contienen potencialmente información valiosa para el gobierno mexicano que los publica [18].

Por otro lado, el pronóstico de aparición de focos epidemiológicos ha sido ampliamente estudiado [1, 2, 5, 6, 12, 24] y los resultados de estos estudios propician la planificación y mejor toma de decisiones respecto a las enfermedades estudiadas, por lo que esta práctica se torna relevante para instituciones de salud, farmacéuticas y gubernamentales.

Este artículo presenta un estudio de los componentes temporales de los datos mencionados además de los resultados de algoritmos de pronóstico aplicados a tales datos distribuido del siguiente modo: La sección 2 contiene una revisión bibliográfica de los artículos más recientes y relevantes sobre el tema de pronóstico epidemiológico; en la sección 3 se definen los conceptos necesarios para comprender la metodología utilizada; dicha metodología se detalla en la sección 4; a continuación, en la sección 5, se muestran los resultados de los experimentos computacionales de este estudio; las conclusiones y discusión de los resultados mencionados aparecen en la sección 6.

## 2. Revisión bibliográfica

Muy recientemente Thompson and Brooks-Pollock [20] recopilaron estudios que involucran el modelado de epidemias que afectan seres humanos, plantas y animales, y proponen la necesidad de unificar estudios que involucren combinaciones de estos modelos por la interrelación que existe entre estos tres tipos de enfermedades.

En este contexto, Chatters et al. [7] analizaron los movimientos de ganado como fuentes de infección con el uso de redes de movimiento del ganado para predecir aparición de enfermedades provenientes del contacto entre distintos grupos de animales.

Además, Tian et al. utilizaron efectivamente un modelo SARIMA para pronosticar aparición de casos epidemiológicos que afectan a manos, pies y boca con base en registros de incidencias reportadas entre 2008 y 2018 de los que estudiaron sus variaciones estacionales y periódicas. Por su parte, Wang et al. [21] predijeron ocurrencias desde datos de reportes de casos de VIH y SIDA en China durante el periodo 2005–2016 con la finalidad de monitorear y controlar estas enfermedades en Guangxi, segunda provincia china con el mayor número de estas incidencias. Pronósticos de esta misma enfermedad fueron estudiados por Xie [23] con datos reportados por la OMS del estimado número de personas que padecen VIH durante 2000 a 2015 en periodos de cinco años.

Por último, la relevancia de este tipo de estudios por modelos lineales, no lineales, estocásticos y determinísticos ha sido ampliamente explorada y demostrada por Dattner and Huppert en un artículo publicado el año pasado [9].

## 3. Marco teórico

Las herramientas utilizadas para la elaboración de este artículo y los experimentos computacionales asociados a éste fueron una computadora portátil Acer Aspire Nitro 15

---

*Email address:* jose.benavidesvz@uanl.edu.mx (Alberto Benavides)

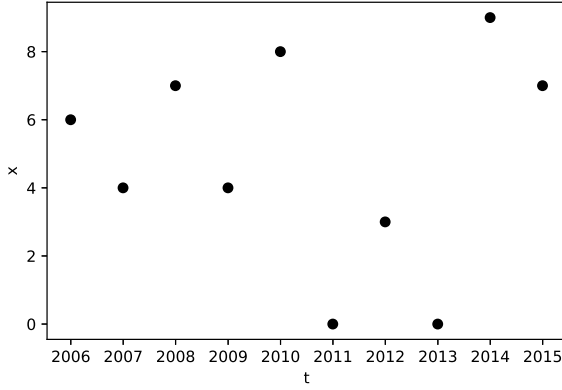


Figura 1: Ejemplo de serie de tiempo de datos generados aleatoriamente con distribución normal y frecuencias anuales.

con sistema operativo Windows 10 Pro de 64 bits, procesador Intel Core i7-8750H a 2.20 GHz, con 8 GB de memoria RAM y un disco de estado sólido Kingston SA40037480G. Los algoritmos se escribieron y ejecutaron en **Python** versión 3 [16] con un extenso uso de librerías entre las que destacan **pandas** [13], **numpy**, **statsmodel**, [19] y el entorno de prototipado **Jupyter** [17]

Los registros de casos de las enfermedades están agrupados por *semana epidemiológica* entendida como una medida temporal utilizada para comparar datos epidemiológicos donde la primera semana epidemiológica del año termina el primer sábado de enero [3]. Además, cada enfermedad tiene asociada una *Clasificación Internacional Estadística de Enfermedades*, abreviada CIE por simplicidad [15], dada por la Organización Mundial de la Salud con la finalidad de entender las causas y efectos que estas enfermedades tienen en la morbilidad y mortalidad de la población [14].

Para pronosticar a partir de datos temporales, se requiere una *serie de tiempo* de los datos. Las series de tiempo se definen como un conjunto de observaciones  $X$  tomadas en un tiempo  $t$  determinado [5], en las que cada observación  $x_t$  expresa los casos registrados de cada enfermedad a lo largo de todas las semanas epidemiológicas reportadas durante el periodo especificado. La figura 1 muestra un ejemplo de una serie de tiempo de números enteros generados al azar registrados en el periodo 2005 a 2015.

Entre los componentes de una serie de tiempo se hallan la *tendencia*, la *estacionalidad* y los *residuales* [5]. La tendencia  $\beta_0$  viene dada por la pendiente que se obtiene de la regresión lineal  $\hat{y}$  de las series de tiempo con observaciones  $Y$ , tal que

$$\hat{y}(t, \beta) = \beta_0 + \beta_1 y_1 + \dots + \beta_t y_t \quad (1)$$

cuya función objetivo consiste en minimizar la suma de errores cuadrados, es decir

$$\sum_t (y_t - \hat{y}_t)^2. \quad (2)$$

La condición de que una serie de tiempo sea pronosticable es que sea *estrictamente estacionaria* [5]. La estacionalidad, por su parte, permite conocer si las propiedades de una serie no cambian con el tiempo [8]. Así, dada una serie de tiempo  $Y = y_t; t \in \mathbb{R}$ , se considera estrictamente estacionaria si

$$F_Y(y_{t_1+\tau}, \dots, y_{t_n+\tau}) = F_Y(y_{t_1}, \dots, y_{t_n}) \quad (3)$$

para  $T \in \mathbb{R}, n \in \mathbb{N}$  y  $\tau \in \mathbb{R}$ .

A su vez, los residuales se conocen por el ruido blanco  $\epsilon_i \in \mathbb{R}$  de la ecuación del modelo autorregresivo

$$y_t = c + \omega_1 y_{t-1} + \dots + \omega_p y_{t-p} + \epsilon_t. \quad (4)$$

En esta ecuación lineal, los  $\omega$  del modelo son los pesos que se ajustan a la serie de tiempo con observaciones  $X$  con un retraso de  $p$ , dado un vector de constantes  $c$  correspondientes a las intercepciones [5].

Las pruebas de hipótesis de Dikey Fuller aumentada y la de Kwiatkowski-Phillips-Schmidt-Shin permiten conocer si una serie de tiempo es estacionaria. La primera propone como hipótesis nula que  $\Omega = 1$  mientras que la hipótesis alternativa es que  $\Omega \neq 1$ , así que si se falla en rechazar la hipótesis nula la serie de tiempo es estacionaria pero se podría convertir en estrictamente estacionaria  $y'$  diferenciándola mediante

$$y'_t = y_t - y_{t-n} \quad (5)$$

donde  $n$  es el retraso estacional que en este caso  $n = 7$  puesto que los datos consisten en series de tiempo tomadas cada siete días. La prueba de Kwiatkowski-Phillips-Schmidt-Shin propone como hipótesis nula que la serie de tiempo posee tendencia estacionaria y como hipótesis alternativa que  $\Omega = 1$  tal que si se rechaza la hipótesis nula, la serie no es estacionaria debido a su tendencia por lo que hay que restar la pendiente  $\beta_0$  de la regresión lineal a la serie de tiempo para obtener una señal  $Y''$  sin tendencia. Ahora bien, en caso de que ambas pruebas den por resultado que determinada serie de tiempo no es estacionaria, tal serie de tiempo no se podría someter a pronóstico por los métodos descritos en el estado del arte ni bibliografía relacionada.

Por último, el modelo de *suavizamiento exponencial* es un modelo de pronóstico definido por Charles Holt en 1957 [11] como sigue

$$s_t = \alpha \cdot x_t + (1 - \alpha) \cdot s_{t-1} = s_{t-1} + \alpha \cdot (x_t - s_{t-1}) \quad (6)$$

donde  $0 < \alpha < 1$  es un factor de suavizado que se aplica a la media ponderada  $s_t$  entre la observación actual  $x_t$  y la anterior  $x_{t-1}$ .

#### 4. Metodología

Tras la extracción, limpieza y conversión en series de tiempo de los datos detallada en [4], en la que se agrupan por enfermedad los casos registrados semanalmente normalizados entre la derechohabencia del año correspondiente, se descompone cada serie de tiempo en su tendencia, componentes estacionarios y residuales además de obtener su

tendencia y desviación estándar acumuladas con la finalidad de comprender el comportamiento de las enfermedades a lo largo del periodo 2005–2015.

Con esta información se realizan las pruebas de Dickey Fuller y la de Kwiatkowski-Phillips-Schmidt-Shin para conocer si las series de tiempo son estacionarias. Si ambas pruebas dan por resultado que la serie de tiempo no es estacionaria, se descarta del pronóstico. Si sólo Dickey Fuller da por resultado que no es estacionaria, se diferencia la serie de tiempo. Por otro lado, si la serie de tiempo resulta no ser estacionaria sólo por la prueba de Kwiatkowski-Phillips-Schmidt-Shin, se resta a la serie de tiempo original la tendencia. Por último, si por ambas pruebas se concluye que la serie es estacionaria, dicha serie no requeriría preprocesamiento para pronosticarla.

Posteriormente se separa cada serie de tiempo en una serie de tiempo de entrenamiento y una de prueba que comprenden el 70 % y 30 % respectivamente del total de observaciones de cada serie de tiempo. A partir de cada conjunto de entrenamiento, se ajusta un modelo de suavizamiento exponencial de Holt-Winter (con o sin estacionalidad, según la serie de tiempo) y se predicen los casos normalizados por serie de tiempo para el periodo comprendido por el conjunto de entrenamiento correspondiente.

Finalmente, se comparan por modelo lineal los logaritmos de los pronósticos contra los logaritmos de los valores reales para conocer el ajuste que tiene el pronóstico propuesto.

## 5. Resultados

Debido a la extensión de las enfermedades, para cada combinación de resultados de pruebas de hipótesis para determinar estacionalidad de series de tiempo, se dará un ejemplo. Cabe señalar que ninguna serie de tiempo fue estacionaria según las dos pruebas.

### 5.1. Serie no estacionaria

El primer caso son las series no estacionarias, definidas como aquellas que no son estacionarias ni por Dickey Fuller ni por Kwiatkowski-Phillips-Schmidt-Shin. Un ejemplo de estas series se presenta en las figuras 2 (p. 4) y 3 (p. 5). En la primera aparece la serie de tiempo original, la tendencia y la desviación estándar acumuladas, y en las figuras 3a, 3b y 3c los componentes de tendencia, estacionalidad y residuales respectivamente. A causa de que esta serie estadísticamente no es estacionaria, no se procede a su pronóstico.

### 5.2. Serie no estacionaria por Dickey Fuller

El segundo caso corresponde a las series de tiempo que no cumplen con ser estacionarias solamente según la prueba de Dickey Fuller. Estas series de tiempo fueron luego diferenciadas mediante la ecuación 5 para luego obtener su predicción por el modelo de suavizamiento exponencial con  $\alpha = 0.1$  y componente estacional de 52 semanas. La CIE A10.0 representa un ejemplo de este grupo presente en las figuras 4 (p. 6), 5 (p. 7) y 6 (p. 8). La primera figura

muestra la serie de tiempo con sus acumulados de la tendencia y desviación estándar. El segundo grupo de figuras está compuesto por las componentes de tendencia (figura 5a), estacionalidad (figura 5b) y residuales (figura 5c). Por último, el pronóstico por suavizamiento exponencial se despliega en la figura 6.

### 5.3. Serie no estacionaria por KPSS

En el caso en que las series de tiempo sólo se consideran no estacionarias por su tendencia se asigna por rechazar la hipótesis nula de la prueba de Kwiatkowski-Phillips-Schmidt-Shin. Puesto que la tendencia no es estacionaria en estas series de tiempo, se procede a eliminar la tendencia de las mismas y, a partir de la señal resultante de esta diferenciación, se pronostica por suavizamiento exponencial. La CIE A03 se comparte como ejemplo de estas series de tiempo en las figuras 7 (p. 8), 8 (p. 9) y 9 (p. 10). La primera figura muestra la serie de tiempo con sus acumulados de la tendencia y desviación estándar. El segundo grupo de figuras está compuesto por las componentes de tendencia (figura 8a), estacionalidad (figura 8b) y residuales (figura 8c). Por último, el pronóstico por suavizamiento exponencial se despliega en la figura 9.

### 5.4. Series de tiempo estacionarias

Las pruebas de hipótesis para comprobar estacionalidad no coincidieron con ninguna de las series de tiempo con las que se realiza este estudio.

### 5.5. Relación real contra pronóstico

Ahora se muestra, en la figura 10, la relación lineal estadísticamente significativa que existe entre el pronóstico del primer valor predicho del conjunto de prueba contra el último valor presente en el conjunto de entrenamiento, esto para todas las CIEs. Se logra un error de 0.0552 y una pendiente de 0.9288.

## 6. Conclusiones y discusión

Este estudio consistió en una metodología robusta de pronóstico para series de tiempo epidemiológicas con resultados satisfactorios tras el estudio, comprensión y preprocesamiento de estas series de tiempo. Destaca como relevante el hecho de que con esta metodología no se logra pronosticar con intervalos de confianza superiores al 95 % las series de tiempo consideradas no estacionarias por las pruebas de hipótesis de Dickey Fuller ni por Kwiatkowski-Phillips-Schmidt-Shin, esto porque sus propiedades tales como la pendiente y la desviación estándar no permanecen en el paso del tiempo.

Sin embargo, se muestran como estrategias indispensables para el pronóstico de series de tiempo el uso de funciones que diferencien las series de tiempo en señales sin tendencia para el caso de las series de tiempo que tengan tendencia no estacionaria según Kwiatkowski-Phillips-Schmidt-Shin, o diferenciándolas por resta de la serie de

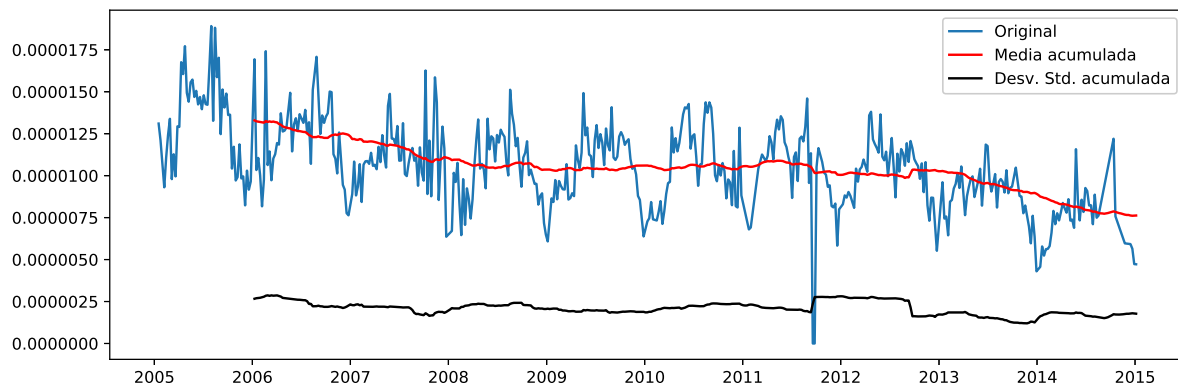


Figura 2: Serie de tiempo (azul) con sus respectivas tendencia acumulada (rojo) y desviación estándar acumulada (negro).

tiempo original menos la serie de tiempo con retraso periódico cuando no son estacionarias por Dickey Fuller.

Es importante remarcar que un pronóstico más preciso se logra al establecer las frecuencias estacionales correctas, que en este caso fueron de 52 semanas dado que la periodicidad de los ciclos mostrados por los componentes estacionarios coincidía con ciclos anuales de 52 semanas cada uno. Este ajuste permite incluir en el modelo la frecuencia y amplitud del conjunto de entrenamiento previamente seleccionado y plasmarlo en los resultados pronosticados.

Pese a que el análisis intuitivo de los componentes estacionarios junto a los acumulados de la tendencia y desviación estándar de las series del tiempo pueden dar una idea aproximada de la estacionalidad de la misma, las pruebas de hipótesis se presentan como una herramienta definitiva para el preprocesamiento de este tipo de datos con la finalidad de conocer su pronosticabilidad.

A partir de estos resultados, es posible considerar otras variables usualmente empleadas en modelos de pronóstico como las autocorrelaciones y autocorrelaciones parciales a la par que los coeficientes de Fourier representados en espectrogramas, todas ellas para mejorar la precisión de los pronósticos en modelos ARIMA (que incluyen modelos ARMA, AR y AM) para series de tiempo no estacionales y SARIMA para series estacionales.

Así también, se puede aproximar el pronóstico de series de tiempo no estacionarias según criterios de Dickey Fuller ni Kwiatkowski-Phillips-Schmidt-Shin por suavizamientos exponenciales simples cuando hay pocos e irregulares datos sin tendencia, modelos lineales de Holt para datos con tendencia moderada o modelos de tendencia amortiguada de Holt, estos últimos para series de tiempo con tendencias fuertemente marcadas.

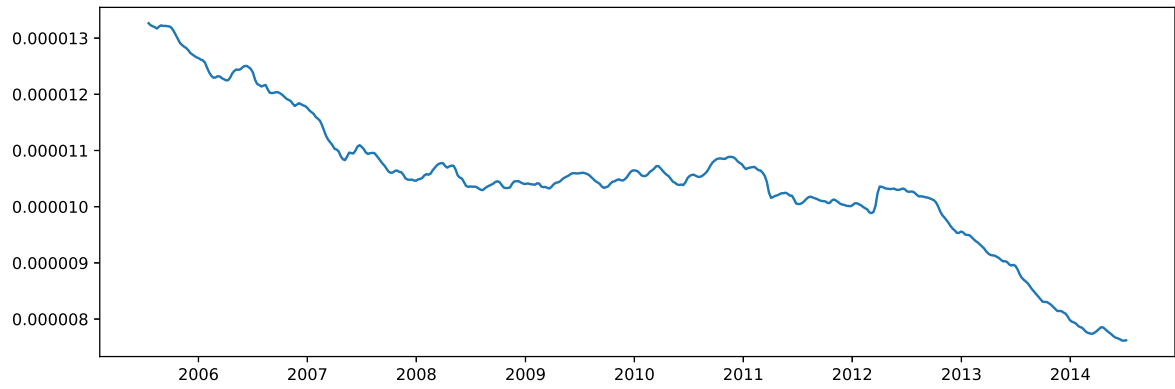
## 7. Agradecimientos

Agradezco primeramente al CONACyT por otorgarme una beca que me ha permitido dedicarme a la investigación y compartir mis conocimientos adquiridos en el proceso,

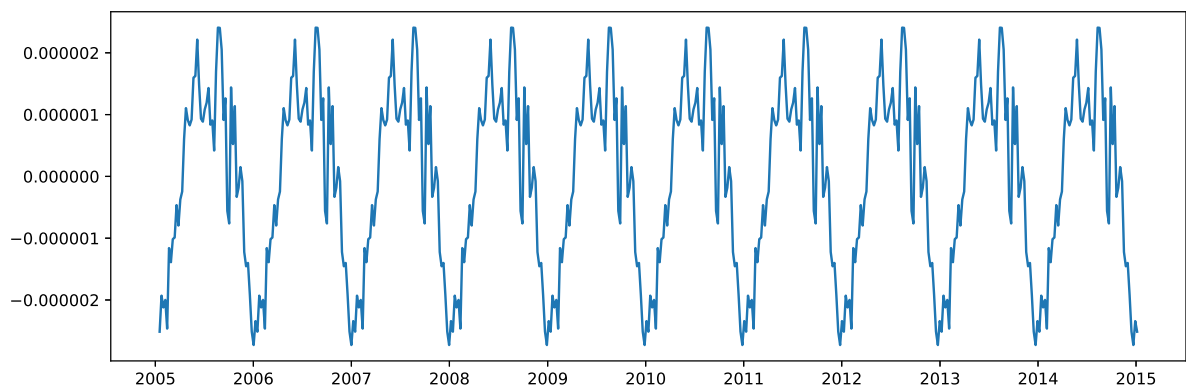
a la Secretaría de Salud de México que hace públicos los datos con la finalidad de que puedan ser de utilidad para investigadores y público en general y a mis asesores Elisa Schaeffer y Arturo Berrones del Posgrado de Ingeniería de Sistemas de quienes he aprendido (y sigo aprendiendo) sobre ciencia de datos, metodología indispensable para el tratamiento de esta información.

## Referencias

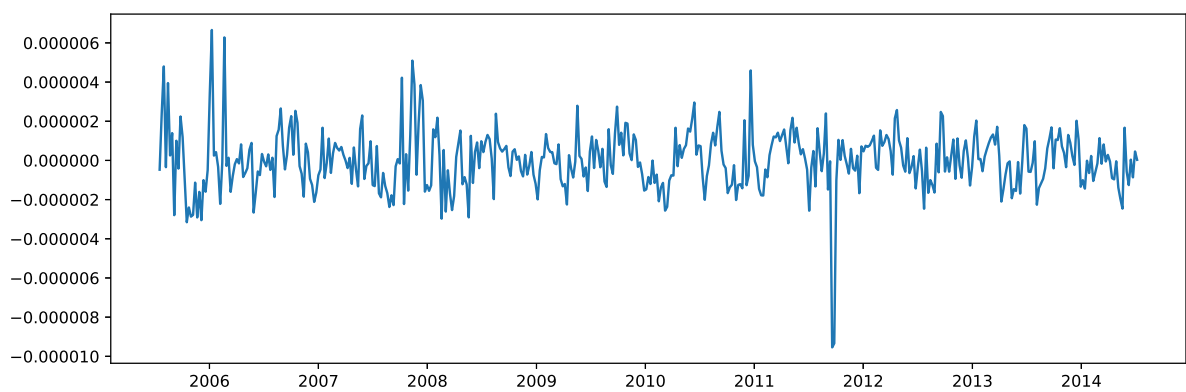
- [1] S. Abdullah, N. Sapii, S. Dir, and T. M. T. Jalal. Application of univariate forecasting models of tuberculosis cases in kelantan. In *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, pages 1–7, Sep. 2012. doi: 10.1109/ICSSBE.2012.6396582.
- [2] Ozgur M. Araz, Dan Bentley, and Robert L. Muellemann. Using google flu trends data in forecasting influenza-like-illness related ed visits in omaha, nebraska. *The American Journal of Emergency Medicine*, 32(9):1016 – 1023, 2014. ISSN 0735-6757. doi: <https://doi.org/10.1016/j.ajem.2014.05.052>. URL <http://www.sciencedirect.com/science/article/pii/S0735675714004215>.
- [3] Jorge R. Arias. What is an epidemiological week and why do we use them? *The Seeker*, 6(1):7, apr 2006. URL [https://www.cmmcp.org/sites/cmmcp/files/uploads/spring\\_skeeter\\_06.pdf](https://www.cmmcp.org/sites/cmmcp/files/uploads/spring_skeeter_06.pdf).
- [4] Alberto Benavides. *Agrupamiento no supervisado de series de tiempo epidemiológicas de México entre 2005 y 2015*. PhD thesis, FIME, UANL, 2019. Tesis no publicada.
- [5] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, Switzerland, 2016.
- [6] Howard S. Burkom, Sean Patrick Murphy, and Galit Shmueli. Automated time series forecasting for biosurveillance. *Statistics in Medicine*, 26(22):4202–4218, 2007. doi: 10.1002/sim.2835.
- [7] G. L. Chatters, P. C. D. Johnson, S. Cleaveland, J. Crispell, W. A. de Glanville, T. Doherty, L. Matthews, S. Mohr, O. M. Nyasebwa, G. Rossi, L. C. M. Salvador, E. Swai, and R. R. Kao. Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1776):264, 2019. doi: 10.1098/rstb.2018.0264. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2018.0264>.
- [8] David Cox and Hilton David Miller. *The Theory of Stochastic Processes*. Methuen & Co, Reino Unido, 1965.
- [9] Itai Dattner and Amit Huppert. Modern statistical tools for inference and prediction of infectious diseases using mathematical



(a) Tendencia.



(b) Componentes estacionarios.



(c) Residuales.

Figura 3: Serie de tiempo y descomposición de la CIE A05 durante el periodo 2005–2015

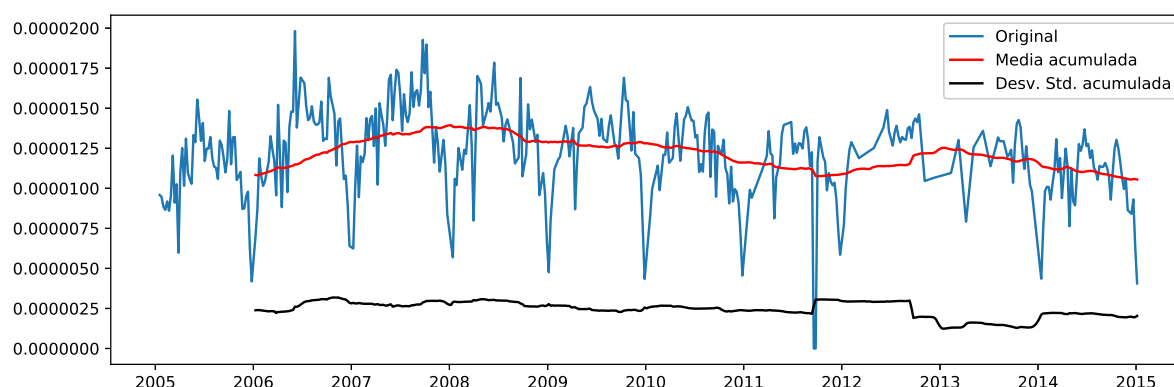
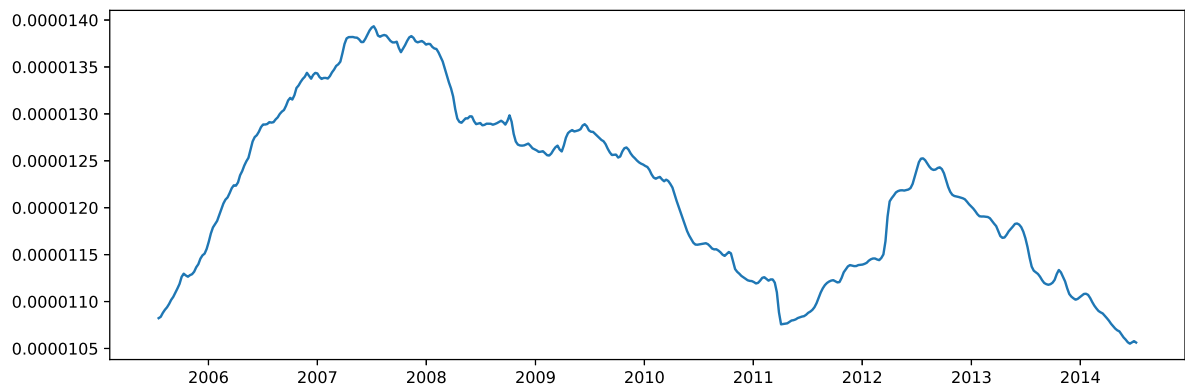
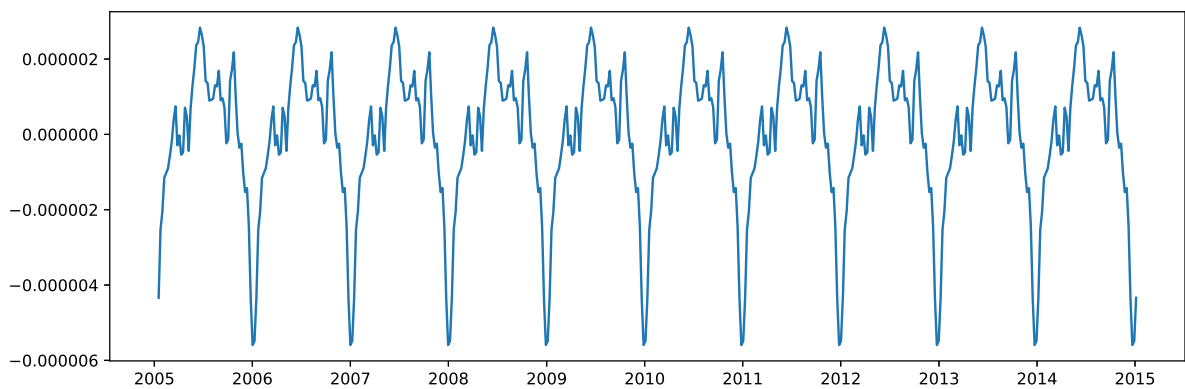


Figura 4: Serie de tiempo (azul) con sus respectivas tendencia acumulada (rojo) y desviación estándar acumulada (negro).

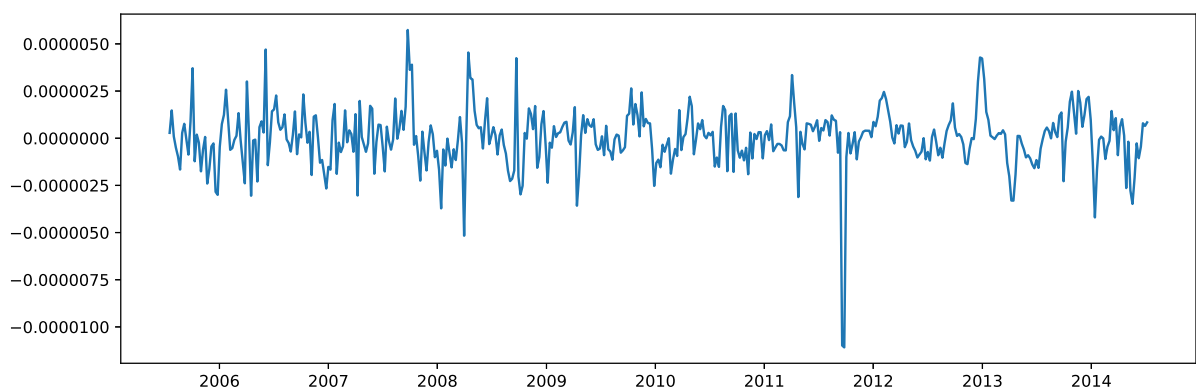
- models. *Statistical Methods in Medical Research*, 27(7):1927–1929, 2018. doi: 10.1177/0962280217746456.
- [10] gob.mx. Boletín epidemiológico sistema nacional de vigilancia epidemiológica sistema Único de información, oct 2016. URL <https://www.gob.mx/salud/acciones-y-programas/direccion-general-de-epidemiologia-boletin-epidemiologico>. Accedido: 2019-05-23.
- [11] Charles C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, 2004. ISSN 0169-2070. doi: <https://doi.org/10.1016/j.ijforecast.2003.09.015>. URL <http://www.sciencedirect.com/science/article/pii/S0169207003001134>.
- [12] Ariel Linden, John L. Adams, and Nancy Roberts. Evaluating disease management program effectiveness: An introduction to time-series analysis. *Disease Management*, 6(4):243–255, 2003. doi: 10.1089/109350703322682559.
- [13] NumFOCUS. Python data analysis library, 2019. URL <https://pandas.pydata.org/>. Accedido: 2019-04-07.
- [14] Organización Mundial de la Salud. La organización mundial de la salud (oms) publica hoy su nueva clasificación internacional de enfermedades (cie-11), sep 2018. URL [https://www.who.int/es/news-room/detail/17-06-2018-who-releases-new-international-classification-of-diseases-\(icd-11\)](https://www.who.int/es/news-room/detail/17-06-2018-who-releases-new-international-classification-of-diseases-(icd-11)). Accedido: 2019-03-20.
- [15] World Health Organization. *International statistical classification of diseases and related health problems—10th revision*. WHO Library Cataloguing, France, 2016.
- [16] Python Software Foundation. Python 3.7.0, 2018. URL <https://www.python.org/downloads/release/python-370/>. Accedido: 2018-08-13.
- [17] Project Jupyter. Project Jupyter, 2019. URL <https://jupyter.org/>. Accedido: 2019-03-01.
- [18] Secretaría de Salud. ¿qué son los datos abiertos?, jul 2018. URL <https://datos.gob.mx/blog/que-son-los-datos-abiertos?category=casos-de-uso&tag=educacion>. Accedido: 2018-08-13.
- [19] The SciPy community. scipy.stats.linregress – scipy v1.2.1 reference guide, 2019. URL <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html>. Accedido: 04-07-2019.
- [20] Robin N. Thompson and Ellen Brooks-Pollock. Detection, forecasting and control of infectious disease epidemics: modelling outbreaks in humans, animals and plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1775):20190038, 2019. doi: 10.1098/rstb.2019.0038. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2019.0038>.
- [21] C. W. Tian, H. Wang, and X. M. Luo. Time-series modelling and forecasting of hand, foot and mouth disease cases in china from 2008 to 2018. *Epidemiology and Infection*, 147:e82, 2019. doi: 10.1017/S095026881800362X.
- [22] G. Wang, W. Wei, J. Jiang, C. Ning, H. Chen, J. Huang, B. Liang, N. Zang, Y. Liao, R. Chen, and et al. Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting hiv incidence in guangxi, china. *Epidemiology and Infection*, 147:e194, 2019. doi: 10.1017/S095026881900075X.
- [23] Liming Xie. Analyzing and Forecasting HIV Data Using Hybrid Time Series Models. *Asian Journal of Probability and Statistics*, 2(3):1–12, 2019. doi: 10.9734/AJPAS/2018/46566.
- [24] Kate Zinszer, Aman D Verma, Katia Charland, Timothy F Brewer, John S Brownstein, Zhuoyu Sun, and David L Buckneridge. A scoping review of malaria forecasting: past work and future directions. *BMJ Open*, 2(6), 2012. ISSN 2044-6055. doi: 10.1136/bmjopen-2012-001992. URL <https://bmjopen.bmj.com/content/2/6/e001992>.



(a) Tendencia.



(b) Componentes estacionales.



(c) Residuales.

Figura 5: Serie de tiempo y descomposición de la CIE A01.0 durante el periodo 2005–2015

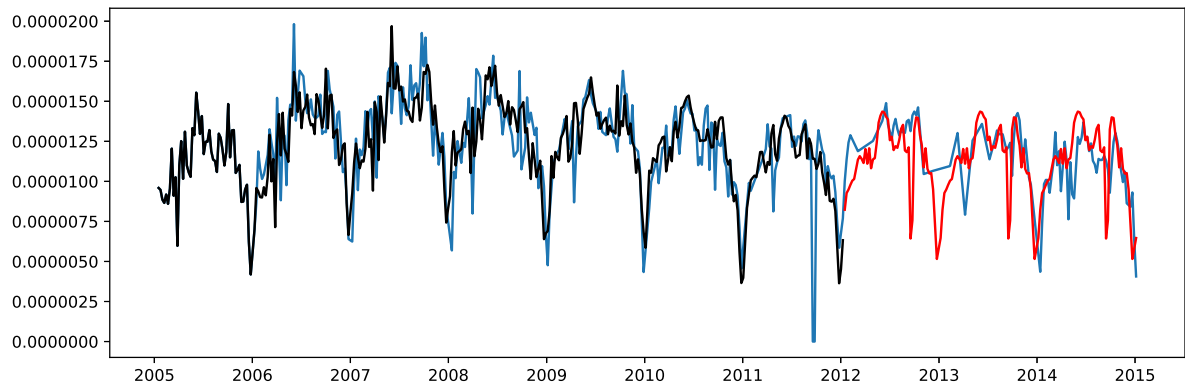


Figura 6: Serie de tiempo de la CIE A01.0 (azul) separada en 70 % inicial de los datos como conjunto de entrenamiento cuyo ajuste al modelo se muestra en color negro y los valores pronosticados a partir del modelo aparecen en rojo, correspondientes al último 30 % de los datos como conjunto de prueba.

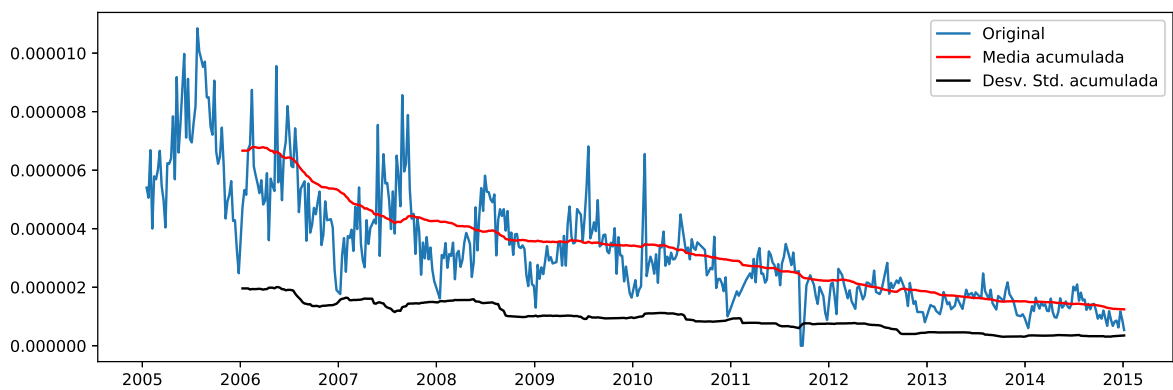
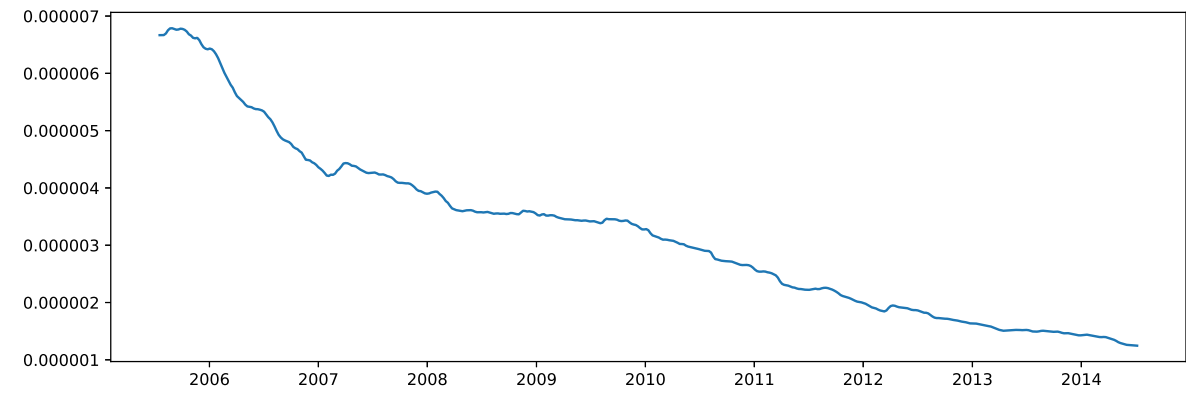
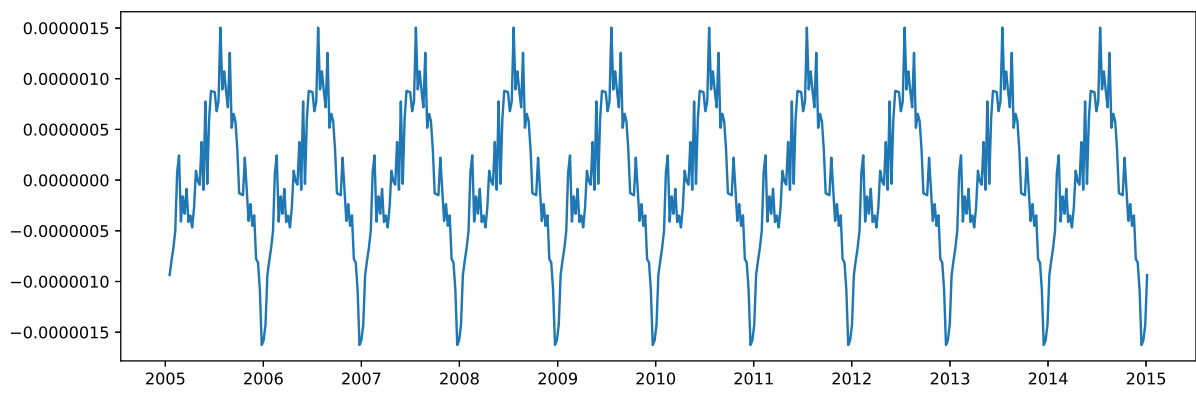


Figura 7: Serie de tiempo (azul) con sus respectivas tendencia acumulada (rojo) y desviación estándar acumulada (negro).

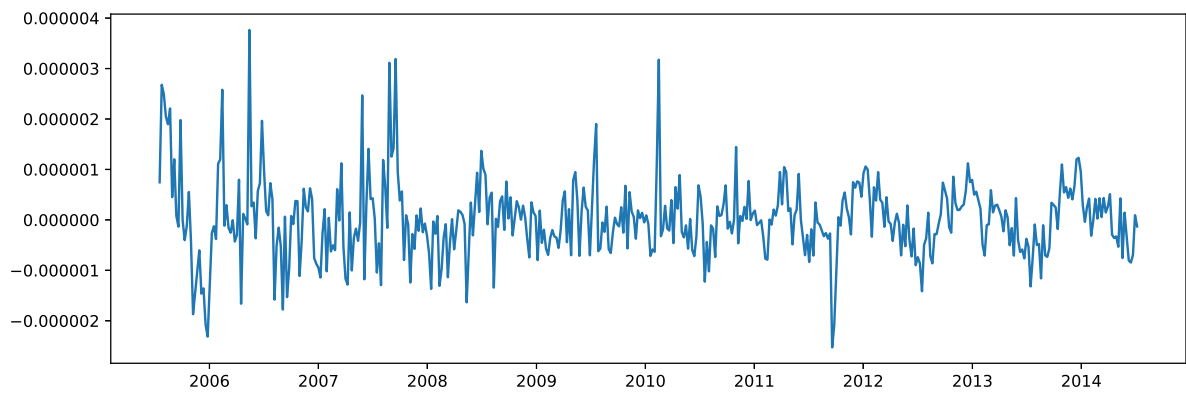




(a) Tendencia.



(b) Componentes estacionarios.



(c) Residuales.

Figura 8: Serie de tiempo y descomposición de la CIE A03 durante el periodo 2005–2015

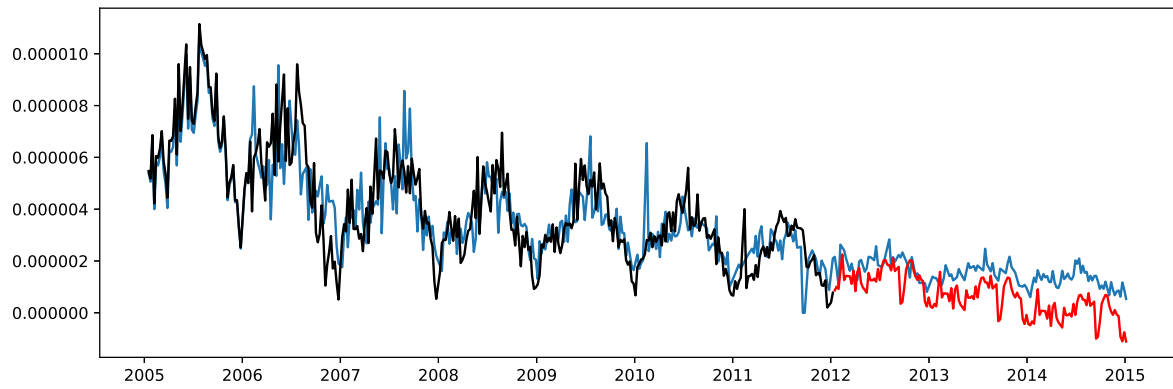


Figura 9: Serie de tiempo de la CIE A03 (azul) separada en 70 % inicial de los datos como conjunto de entrenamiento cuyo ajuste al modelo se muestra en color negro y los valores pronosticados a partir del modelo aparecen en rojo, correspondientes al último 30 % de los datos como conjunto de prueba.

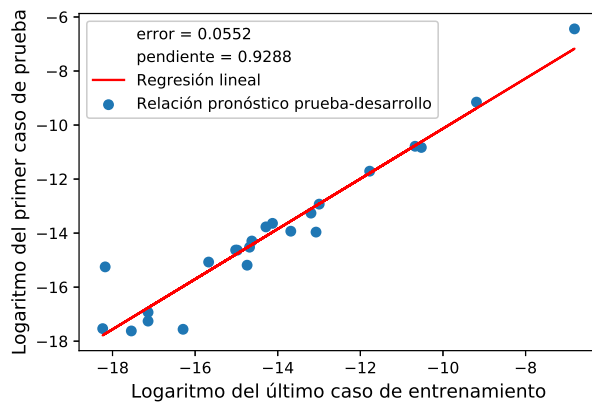


Figura 10: Relación lineal entre los logaritmos del primer dato pronosticado contra el último presente en los conjuntos de prueba y entrenamiento, respectivamente, para todas las CIEs presentes en el estudio.