

PRONÓSTICO MEDIANTE SERIES DE TIEMPO DE TEMPERATURAS MÁXIMAS Y MÍNIMAS PROMEDIO MENSUALES, ESTACIÓN 09045, CDMX

**Rocio Cervantes-Osornio^{1*}; Ramón Arteaga Ramírez²; Mario Alberto Vázquez Peña²;
Waldo Ojeda Bustamante⁴; Lucia Victoria Carmona Mota**

¹Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias. Carretera México-Lechería, km 18.5 Chapingo, Estado de México C.P. 56230 A.P. 10, México.

rcervanteso@hotmail.com – 01 800 088 2222 85565 (*Autor de correspondencia)

²Departamento de Irrigación. Universidad Autónoma Chapingo, km 38.5 Carretera México-Texcoco, Chapingo, Estado de México. México.

⁴Riego y Drenaje. Instituto Mexicano de Tecnología del Agua. Paseo Cuauhnáhuac 8532, Progreso, Jiutepec, Morelos, C.P. 62550. México.

Resumen

La variable temperatura resulta importante en diversas áreas, recientemente ha tomado importancia conocer con cierto grado de anticipación su valor debido a los incrementos de esta, y su relación estrecha con el cambio climático. En el presente trabajo se desarrollaron dos modelos de series de tiempo tipo ARMA para el pronóstico de temperaturas promedio mensuales máximas y mínimas. Se analizaron datos de un periodo de aproximadamente 13 años de la estación 009045, en el Municipio de Milpa Alta en CDMX, y se realizó el pronóstico a 24 meses. De ambos modelos encontrados, el modelo de la temperatura mínima realizó un ajuste de los datos con una RMSE de 0.666.

Palabras claves: Box-Jenkins, series de tiempo, promedios mensuales temperaturas

Introducción

Los pronósticos en series de tiempo tienen la tarea de conocer con anticipación algunos valores en el futuro, dada una secuencia de datos históricos de la misma señal o variable, y también se puede utilizar datos históricos de señales correlacionadas (Romeu *et al.*, 2013). Los datos de temperaturas mensuales siempre resultaran necesarios, tanto por los agricultores, como por el personal que toma decisiones: medioambientalistas, gobierno. De igual manera la temperatura es importante porque influye en la tasa de respiración del suelo, en los cambios de fisiología de los cultivos y es una variable necesaria en el cálculo de la evapotranspiración de referencia (Ye *et al.*, 2013). Existen múltiples métodos de estimación y/o pronóstico de las temperaturas, desde los análisis de regresión múltiple (Flocas *et al.*, 1983), pasando por las series de Fourier y los modelos de series de tiempo, entre otros. En series de tiempo una variable es explicada por los valores anteriores de esa misma variable, siendo esta una particularidad de este tipo de modelación. Una característica de lo pronosticado, es que estos valores, van convergiendo a la media del proceso o de la serie.

En el presente trabajo se ajustaron dos modelos autorregresivos de media móvil (ARMA del inglés Autorregresive moving average) de las respectivas variables de temperatura promedio mensuales máximas (T_{max}) y mínimas (T_{min}), con la finalidad de obtener un pronóstico de estas. Uno de los resultados obtenidos fue que el modelo encontrado con Stata para T_{min} ajustó mejor a la serie, con un RMSE de 0.666, a diferencia del ajuste para la T_{max} con un RMSE de 0.963.

Materiales y Métodos

Los datos analizados fueron obtenidos del Servicio Meteorológico Nacional de la estación 09045 Santa Ana Tlacotenco, de la Ciudad de Mexico (CDMX) con una latitud de 19° 10' 44" N, longitud 099° 00' 10" W y altitud de 2595 msnm, perteneciente a la alcaldía de Milpa Alta, del organismo de la CONAGUA-DGE y Cuenca del Rio Moctezuma.

Se utilizó un enfoque empírico para identificar el proceso estocástico que genera la serie de tiempo, a través del correlograma se identificó a cual proceso corresponde dicha serie, ya sea: AR(1), AR(2), ARMA(1,1), etc. Todos los procesos ARMA o ARIMA (Autorregresive Integrated Moving Average) tienen un correlograma, y se utiliza este para identificar qué proceso estocástico tuvieron las series, en este caso, de la serie de datos de los promedios mensuales de la temperaturas máximas y mínimas. Las series de datos presentaron datos faltantes, para solventarlo, dado que las series presentaban un comportamiento estacional cada doce meses, cíclico, se utilizaron modelos de regresión lineal con variables dummy, con estos, se estimaron los datos faltantes de acuerdo al Cuadro 1.

Cuadro 1. Datos mensuales faltantes y sustituidos con modelos de regresión lineal con variables dummy, tanto para la serie de T_{max} como T_{min}

Datos faltantes sustituidos para cada variable	1er modelo (3 datos sustituidos)	2º modelo (21 datos sustituidos)	3er modelo (12 datos sustituidos)	4to. Modelo (4 datos sustituidos)
T_{min}	Marzo-2006, abril-2006, mayo-2006	Enero de 2009 hasta julio del 2010, noviembre-2010, diciembre-2010	Enero-2013 hasta diciembre de 2013	Octubre-2015, Noviembre-2015, Diciembre-2015, Abril-2016
T_{max}	Marzo-2006, abril-2006, mayo-2006	Enero de 2009 hasta julio del 2010, noviembre-2010, diciembre-2010	Enero-2013 hasta diciembre de 2013	Octubre-2015, Noviembre-2015, Diciembre-2015, Abril-2016

Fue necesario que las series, una vez rellenos los datos faltantes, cumplieran con el supuesto de estacionariedad, esto ayuda para que el número de parámetros del modelo de pronóstico sea el menor posible, es decir, se simplifique. Si la serie no es estacionaria, entonces hay que transformar ésta: se aplica logaritmo, primeras diferencias y diferencia estacional. En este estudio se aplicaron tanto a T_{min} como a T_{max} estas transformaciones. Para probar estacionariedad se utilizaron las pruebas de raíces unitarias de Dickey-Fuller y Phillips-Perron (Becketti, 2013).

Una vez las series rellenas, se analizaron los datos de enero de 2005 a octubre de 2017 para ambas variables (temperaturas máximas y mínimas), se realizó el pronóstico para 24 meses a partir de noviembre de 2017, dado que la estación ya no presenta datos desde esta fecha en adelante. Se utilizó Stata 16.0 para realizar el pronóstico mediante series de tiempo y también para ejecutar los modelos de regresión lineal con variables dummy. Cabe resaltar que entre más datos se pronostiquen hacia el futuro el modelo de series de tiempo va perdiendo confiabilidad (Esparza, 2007).

El proceso seguido para obtener el modelo de serie de tiempo fue con la metodología de Box-Jenkins, con las siguientes etapas: 1ª) Identificación. 2ª) Estimación, 3ª) Verificación o diagnóstico, 4ª) Pronóstico. En la tercera etapa se realizó la predicción y se generaron los residuales, a estos se les realizaron las pruebas estadísticas de normalidad Shapiro-Wilk, Jarque-Bera y Liung Box. Si los residuales generados por la predicción del modelo no presentan una distribución normal (campana de Gauss) hay que regresar a la segunda etapa o reconstruir el modelo.

Para los resultados obtenidos con los modelos de series de tiempo se obtuvo la raíz cuadrada del cuadrado medio del error (RMSE), con la ecuación 1, estos valores de RMSE se calcularon de febrero de 2006 a octubre del 2017, porque de acuerdo a los resultados arrojados por Stata, en estas fechas son en las cuales se cuenta tanto con datos estimados como con datos observados para T_{max} y T_{min} .

$$RMSE = \left(\frac{\sum_{i=1}^n (a_i - t_i)^2}{n} \right)^{1/2} \quad (1)$$

donde a_i es el dato estimado por el modelo, t_i , es el dato observado (medido), n es el número de observaciones o estimaciones.

Resultados y Discusión

Se obtuvieron dos modelos ARMA: para T_{min} presenta ocho términos para la parte autorregresiva (AR) y dos términos de media móvil (MA), se aceptó este modelo con un término que no es significativo, esto debido a que el criterio de Akaike indicó que se debe dejar el termino como parte del modelo, además de que el pronóstico resultó aceptable y para T_{max} el modelo se aceptó con 95% de significancia el cual presentó seis términos AR y uno MA (Figura 1).

d12dltmax		OPG							
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]				
ARMA									
ar									
L1.	-.5922216	.0755932	-7.83	0.000	-.7403816	-.4440617			
L2.	-.6034746	.096486	-6.25	0.000	-.7925836	-.4143656			
L3.	-.3240484	.0906788	-3.57	0.000	-.5017755	-.1463212			
L4.	-.2762325	.0782289	-3.53	0.000	-.4295583	-.1229068			
L6.	-.1649957	.0655301	-2.52	0.012	-.2934324	-.0365591			
L23.	.23432	.0620259	3.78	0.000	.1127516	.3558885			
ma									
L12.	-1.000001			
/sigma	.0380329	.0022757	16.71	0.000	.0335727	.0424931			

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

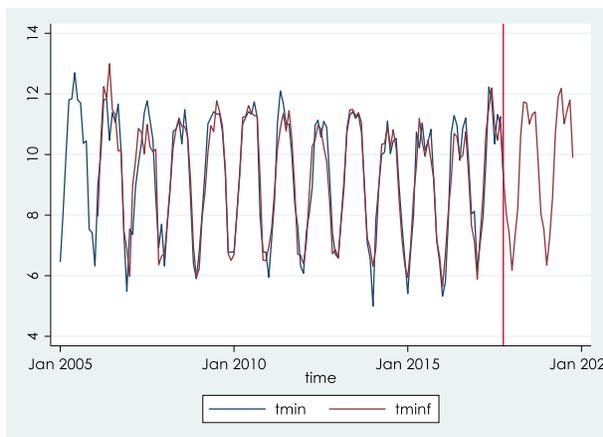
a)

d12dltmin		OPG							
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]				
ARMA									
ar									
L1.	-.8507819	.1125824	-7.56	0.000	-1.071439	-.6301245			
L2.	-.6791791	.1209386	-5.62	0.000	-.9162143	-.4421439			
L3.	-.6476913	.1225253	-5.29	0.000	-.8878364	-.4075462			
L4.	-.3968534	.1129128	-3.51	0.000	-.6181585	-.1755484			
L5.	-.2487758	.0890389	-2.79	0.005	-.4232887	-.0742628			
L10.	-.2035757	.0518756	-3.92	0.000	-.3052501	-.1019013			
L12.	-.1668938	.0916666	-1.82	0.069	-.3465571	.0127694			
L13.	-.1634943	.1121384	-1.46	0.145	-.3832816	.0562929			
ma									
L1.	.2860908	.1109759	2.58	0.010	.068582	.5035996			
L12.	-.6645166	.107993	-6.15	0.000	-.876179	-.4528543			
/sigma	.0703537	.0043957	16.01	0.000	.0617382	.0789691			

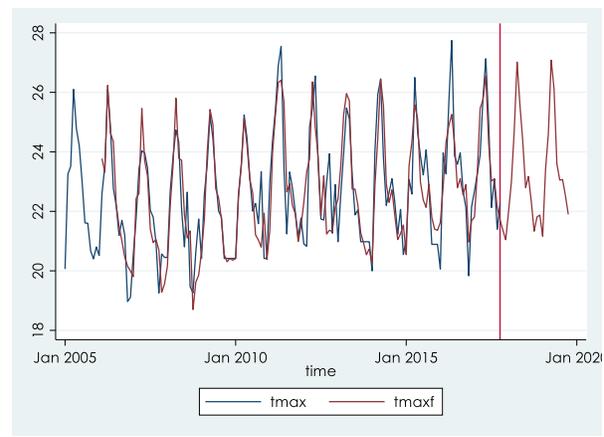
Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

b)

Figura 1. Modelos ARMA con sus coeficientes respectivos para a) T_{max} y b) T_{min} .



a)



b)

Figura 2. Pronóstico de 24 meses de las variables a) temperatura mínima y b) máxima, para la estación 09045 de la CDMX.

Se observa en la Figura 2 el pronóstico de ambos modelos ARMA obtenidos, para T_{min} (Figura 2a) el modelo obtuvo un mejor ajuste, en comparación con T_{max} (Figura 2b), lo corroboran los valores de RMSE obtenidos, el cual fue para T_{min} de 0.666 y para T_{max} de 0.963.

Conclusiones

De ambos modelos encontrados, el que mejor aproximó a los datos observados fue el de T_{min} . Los modelos de series de tiempo muestran ser una herramienta robusta en la estimación y/o pronóstico de las temperaturas promedio mensuales.

Referencias Bibliográficas

- Beckett S. 2013. Introduction to time series using Stata. A Stata Press Publication, StataCorp LP, Texas, United States of America, 443 pp.
- Esparza C.C. 2007. Series temporales, 53 pp.
- Flocas A. A.; Giles B. D.; and Angouridakis V.E. 1983. On the estimation of annual and monthly means values of air temperature over Greece using stepwise multiple regression analysis. Arch. Met. Geoph. Biocl., Ser. B. 32:287-295.
- Meyer S. J.; Hubbard K. G. 1992. Nonfederal automated weather stations and networks in the United States and Canada: a preliminary survey. Bulletin of the American Meteorological Society. 73(4):449-457.
- Romeu P.; Zamora-Martinez F.; Botella-Rocamora P.; and Pardo J. 2013. V. Mladenov et al (Eds.): ICANN 2013, LNCS 8131, pp 451-458.
- Ye L.; Yang.; Van Ranst E.; and Tang H. 2013. Time-series modeling and prediction of global monthly absolute temperature for environmental decision-making. Advances in atmospheric sciences 30(2):382-396.