

Modelado y análisis de datos climáticos de Puerto Ayora a través de series temporales

Emily Alcívar, Josué Alvear, Ambar Borbor

emiaalci@espol.edu.ec , jealvear@espol.edu.ec , aeborbor@espol.edu.ec

Resumen

En este proyecto se realizó el estudio de la serie que representa el cambio de temperatura ($^{\circ}\text{C}$) promedio diario en Puerto Ayora; a través de modelos estacionarios como ARMA, modelos no estacionarios como ARIMA y estacionales como SARIMA. Usando el Software R para la estimación de los coeficientes, indicadores y métricas de los errores para los modelos mencionados. También el software nos permite analizar los supuestos para los errores (normalidad, linealidad, homogeneidad), teniendo como finalidad comparar los modelos y determinar el más óptimo para la representación de la serie original. Así estimar pronósticos apropiados y representarlos con gráficos conforme al comportamiento de la serie, concluyendo por medio del criterio de información de Akaike (AIC) y criterio Bayesiano (BIC) que el mejor es el modelo SARIMA porque representa mejor los puntos antes mencionados.

Abstract

In this project, the study of the series representing the daily average temperature change ($^{\circ}\text{C}$) in Puerto Ayora was carried out; through stationary models such as ARMA, nonstationary models such as ARIMA and seasonal models such as SARIMA. Using Software R for the estimation of error coefficients, indicators and metrics for the above models. The software also allows us to analyze the assumptions for errors (normality, linearity, homogeneity), with the purpose of comparing the models and determining the most optimal for the representation of the original series. Thus estimate appropriate forecasts and represent them with graphs according to the behavior of the series, concluding by means of the Akaike information criterion (AIC) and Bayesian criterion (BIC) that the best is the SARIMA model because it better represents the points mentioned above.

1. Introducción

El pronóstico de datos a través series temporales es un tema importante en economía, negocios, finanzas, etc. Así mismo, la evaluación de la precisión de los pronósticos es necesaria cuando se emplean varias formas de métodos de pronóstico y, más específicamente, pronosticar mediante el análisis de regresión, ya que tienen varias limitaciones debido a los supuestos a evaluar. El objetivo del análisis de series temporales es estudiar las observaciones de trayectoria de las series temporales y construir un modelo para describir con la mayor precisión posible la

estructura de los datos y luego predecir los valores futuros de las series temporales. (Siami-Namini, et al., 2018)

El principal objetivo de este artículo es investigar cuales modelos de series temporales ofrecen mejores predicciones con respecto a criterios previamente establecidos para las mediciones de temperatura media del aire de una localidad en Ecuador. Con ese fin, existen variedades de modelos estocásticos en la predicción de series temporales.

Los modelos a presentar en esta investigación son de naturaleza univariante debido a que

solo se estudian los datos de la temperatura media del aire. Se presenta el modelo de regresión lineal simple con la evaluación de los supuestos, el modelo de medias móviles el cual es utilizado con la finalidad de capturar la variación o movilidad de la media a través del tiempo, en otras palabras, no cambia el entorno de la serie. Seguidamente, se emplea el modelo de suavización exponencial simple está basada en el análisis de los errores de los pronósticos en el que se asigna un mayor peso al último valor en la serie de tiempo, es decir, en los primeros retardos y después, se asignan progresivamente los pesos más pequeños a los valores más antiguos. (Contreras, et al., 2016)

Adicionalmente se presenta un modelo ARMA "Media móvil auto regresiva (ARMA)" univariante para una sola serie temporal de datos en la que se combinan los modelos de media móvil (AR) y media móvil (MA). Por otro lado, se tiene el modelo de "Media móvil integrada auto regresiva (ARIMA)" univariante es un tipo especial de ARIMA donde la diferenciación se tiene en cuenta en el modelo para obtener series estacionarias. Por último, el modelo SARIMA también forma parte de este estudio debido a su capacidad de integrar el comportamiento estacional de la serie con algunos de sus coeficientes en cero y componentes adicionales. A continuación, se presenta en la sección 2 el objetivo general de este artículo, segundo, se muestran en la sección 3 los materiales y métodos utilizados para el análisis. En la sección 4 se presentan los resultados y las discusiones generadas a raíz de los mismos, junto con los respectivos pronósticos generados con el modelo que muestra un mejor ajuste. Finalmente se presentan las conclusiones finales en la sección 5 de este documento.

2. Objetivo General

Encontrar el mejor ajuste de los datos de Temperatura del aire promedio (°C) en Puerto Ayora, Ecuador con datos desde el año 2012 hasta el año 2019.

3. Materiales y Métodos

Para la selección del mejor modelo para pronosticar las mediciones de temperatura promedio del aire en puerto Ayora se seguirán los siguientes pasos:

- ✓ Análisis de la serie temporal (estacionariedad, estacionalidad, explosividad, ruido).
- ✓ Identificación y selección del modelo (componentes autorregresivos y de medias móviles, diferenciación).
- ✓ Validación del modelo mediante evaluación de los residuos.

El lenguaje de programación utilizado es R en el entorno de RStudio. La base de datos fue obtenida de la página oficial de Puerto Ayora donde recolectaron los datos de la temperatura promedio del aire (°C) con una frecuencia diaria, se escogió el periodo de tiempo desde el año 2012 hasta el último año donde se comparten los registros de temperatura en el 2019.

3.1. Media Móviles

La serie refleja el hecho de que las oscilaciones más lentas son más evidentes y se eliminan algunas de las oscilaciones más rápidas. Es una técnica adecuada cuando se han estabilizado las fuerzas que generan las series a pronosticar (Hanke & Wichern, 2006); en este se utilizan datos recientes para reducir el efecto de las fluctuaciones aleatorias y responder al cambio en el proceso de una manera más rápida.

Así, el promedio móvil está dado por la suma de los últimos N datos, como se muestra en la ecuación 1.

$$V_T = \frac{1}{N} (a_{T-N+1} + a_{T-N+2} + \dots + a_T) \quad (1)$$

donde V_T es el valor del promedio móvil; N se refiere al número de periodos que se quiere considerar en el promedio móvil; a_T es la demanda histórica en el periodo T y T es el proceso en el que se encuentra el periodo.

3.2. Regresión Lineal Simple

Las técnicas de regresión permiten hacer predicciones sobre los valores de cierta variable Y (dependiente), a partir de los de otra variable X (independiente), y se intuye existe una relación. La naturaleza e intensidad de relaciones entre variables cuantitativas pueden ser examinadas por medio de los análisis de correlación y de regresión. El objetivo, es predecir, describir o estimar el valor de una variable Y que corresponde al valor dado de otra variable X para proceder a pronosticar. Los supuestos que debe cumplir los errores son importantes para aplicar el modelo de regresión y son tres: independencia, aleatoriedad y normalmente distribuidos con media 0 y varianza σ^2 . La relación entre las variables se muestra en la siguiente ecuación.

$$Y = \beta_0 + \beta_1 x + e \quad (2)$$

donde β_0 y β_1 son los parámetros de intercepto y pendiente respectivamente; e es una variable aleatoria que se conoce como término del error.

3.3. Suavización Exponencial Simple

Los métodos de suavización exponencial tienen la propiedad que los pronósticos son combinaciones ponderadas de observaciones pasadas, con observaciones recientes dando más peso a las observaciones recientes en comparación a las observaciones anteriores (Snyder, 2002). El nombre "suavizado exponencial" refleja el hecho de que los pesos disminuyen exponencialmente a medida que las observaciones se alejan. Se emplea para series sin tendencia y sin estacionalidad y su ecuación es la siguiente (ecuación 3).

$$\hat{X}_t = \hat{X}_{t-1} + (\alpha(X_{t-1} - \hat{X}_{t-1})) \quad (3)$$

donde \hat{X}_t pronóstico de la serie de tiempo para el período t; X_{t-1} valor real de la serie de tiempo en el período t-1; \hat{X}_{t-1} es el pronóstico de la serie para el tiempo t-1 y α es la constante de suavizamiento, $0 \leq \alpha \leq 1$.

3.4. ARMA

Este modelo es la combinación de las expresiones del orden (p) del modelo autorregresivo (AR) y el orden (q) del modelo de media móvil (MA). Donde el modelo autorregresivo AR expresa el valor actual de la serie como una función de los valores que tomó la misma en las p muestras anteriores ponderadas por un factor y de una perturbación aleatoria presente. Y la parte del modelo de medias móviles MA considera que el valor de la serie estacionaria se desplaza alrededor de un valor medio μ y su desplazamiento en el tiempo t es ocasionado por infinitas perturbaciones pasadas, ponderadas por un factor que mide la influencia de cada una de ellas en el valor presente de la serie (López & Carlos, 2013). En la ecuación 4 se detalla lo mencionado.

$$\hat{X}_t = \sum_{i=1}^p \vartheta_i X_{t-i} + \varepsilon_t + \sum_{i=1}^q \varphi_i \varepsilon_{t-i} \quad (4)$$

donde ϑ_i es el parámetro autorregresivo en el tiempo t; ε_t es el error en el tiempo t y φ_i es el parámetro de medias móviles en el tiempo t.

3.5. ARIMA

Se ha descubierto que los modelos estocásticos atribuidos a Box-Jenkins, conocidos como ARIMA, son más eficientes y fiables incluso para los pronósticos a corto plazo. Además, los modelos estocásticos no tienen distribución, ya que no se requieren suposiciones sobre los datos (López & Carlos, 2013). El modelo ARIMA consta de las expresiones AR orden autorregresivo (p), orden de diferenciación (d) y MA orden de media móvil (q). Los modelos de Box-Jenkin se indican mediante ARIMA (p, d, q). "I" implica que el proceso debe someterse a una diferenciación y cuando se realiza el modelado, los resultados se someten a un proceso de integración para producir pronósticos y estimaciones. El modelo ARIMA se expresa como en la ecuación 5.

$$\vartheta(B)X_t = \phi(B)\nabla^d X_t = \varphi_0 + \varphi(B)w_t \quad (5)$$

donde $\phi(B)$ es el operador autorregresivo estacionario, sus raíces $\phi(B) = 0$, caen fuera del círculo unitario; $\vartheta(B) = \phi(B)\nabla^d$ es el operador autorregresivo generalizado, es no estacionario con d diferenciaciones, las raíces de $\vartheta(B) = 0$ y son igual a la unidad; y $\varphi(B)$ es el operador de medias móviles, se asume que es invertible y sus raíces $\varphi(B) = 0$, fuera del círculo unitario.

3.6. SARIMA

Es autorregresivo e integrado de promedio móvil estacional, se basa en ARIMA, con algunos de sus coeficientes en cero y componentes adicionales para integrar el comportamiento estacional de la serie. La parte estacional indica el comportamiento para un periodo de tiempo está relacionado con los que ocurrió en periodos de tiempo anteriores. En concreto, el modelo SARIMA $(p,d,q) \times (P,D,Q)_s$ multiplicativo captura todos los operadores mencionados anteriormente, es decir la componente autorregresiva, la componente de media móviles (tendencia), la componente de diferenciación (estacionariedad) y el factor multiplicativo (estacionalidad) con el fin de capturar todas las componentes que describen a la serie temporal que se desea modelar.

$$\vartheta_p(B)\phi_p(B^S)\nabla_S^D\nabla^dX_t = \varphi_Q(B)\varphi_Q(B^S)w_t \quad (6)$$

donde se combinan los operadores de la componente estacional y de la componente no estacionaria del modelo ARIMA previamente presentado.

Los modelos multiplicativos no son siempre buenos para ajustar todas las series. Existen ocasiones donde se requiere de una modificación que permite al operador de medias móviles mixto ser no multiplicativo y de manera alternativa o adicionalmente ocurre que el operador autorregresivo puede ser no multiplicativo.

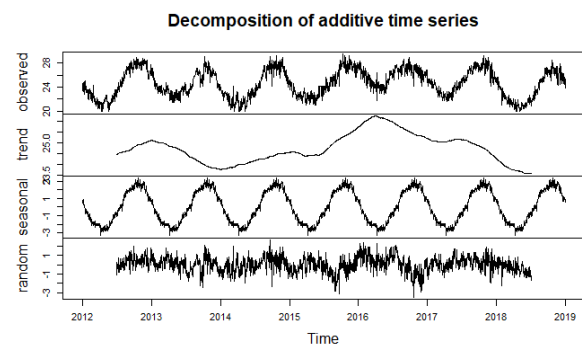
Casi siempre para las series estacionales el orden de diferenciación estacional D por lo general no será mayor que 1, y especialmente

en los casos para las series con periodicidad mensual con $s=12$.

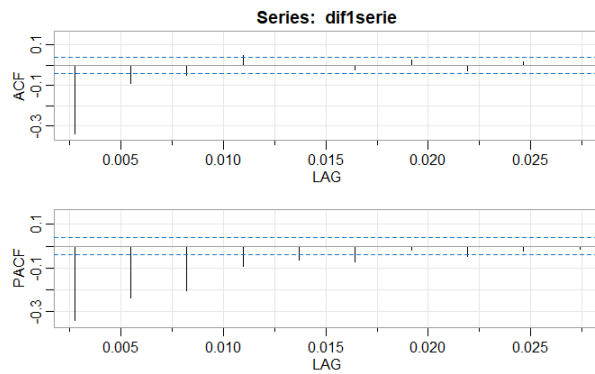
4. Resultados y Discusión

Bajo el criterio AIC y BIC, se escogió el mejor modelo para la parte ARMA, ARIMA y SARIMA para los datos de la temperatura del aire promedio en Puerto Ayora y al final se comparó cuál de los 3 modelos seleccionados previamente se ajustaba más a nuestra serie temporal. Se muestra una tabla con los distintos modelos y sus respectivos parámetros:

Modelos	AIC	BIC
ARMA (2,1)	4788.1	4817.33
ARIMA (4,1,1)	5577.83	5613.71
SARIMA (4,1,1) \times (0,1,0)[365]	5826.41	5860.56



Como notamos en el gráfico de descomposición de la serie de temperatura promedio del aire en puerto Ayora, se observa la presencia de estacionalidad por lo cual es necesaria eliminarla con una diferenciación en la parte estacional a través del modelo SARIMA. Adicionalmente se observa la tendencia de la serie la cual presenta un comportamiento particular y por último se tienen a los errores



Al diferenciarla una sola vez la serie se convierte en estacionaria.

Modelo SARIMA (4,1,1) (0,1,0)

Se aplicó el modelo SARIMA, dado que se escogió como el mejor de los modelos analizados debido a la componente estacional de la serie con frecuencia diaria.

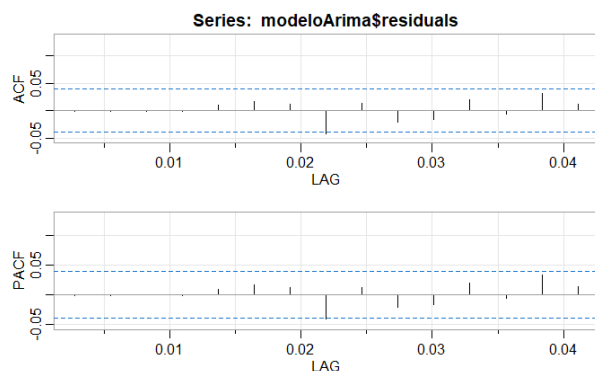
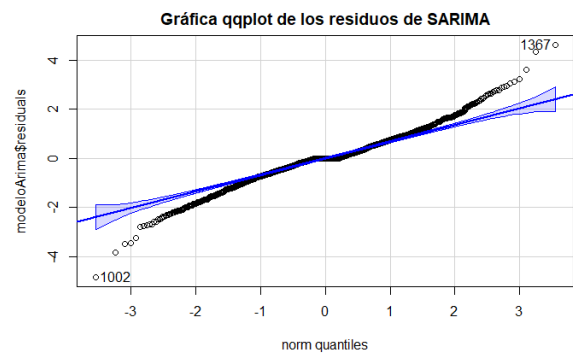


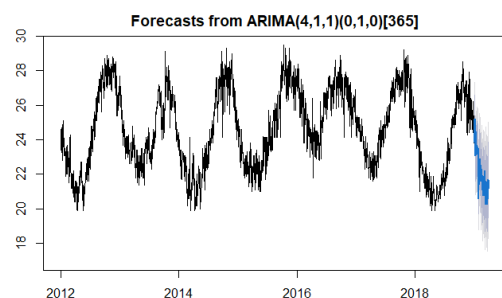
Gráfico ACF y PACF del modelo SARIMA se visualiza que en los primeros retardos los valores de auto correlación y auto correlación parcial están dentro de los intervalos de confianza indicando que el orden de los parámetros p y q es apropiado.

En la gráfica de qqplot de los residuos de modelo escogido se visualizan valores que se salen de los intervalos de confianza, sin embargo, el teorema del límite central es aplicable para estos datos.



Pronósticos

Se procedió a calcular los pronósticos respectivos de los próximos 90 días de la temperatura media del aire en Puerto Ayora, como se visualiza en la gráfica, el ajuste del modelo es bueno ya que se consideran todas las componentes. La componente estacional de la serie es capturada en el modelo, junto con la tendencia que se muestra en el descompose previamente mostrado. Adicionalmente se muestra la tabla de los primeros 24 de los 90 días pronosticados de la temperatura promedio en puerto Ayora, cada uno con sus respectivos intervalos de confianza del 80% y 95%.



Pronósticos del modelo de SARIMA(4,1,1)(0,1,0)					
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2019.027	23.99635	22.82754	25.16516	22.20881	25.78389
2019.055	24.00757	22.73028	25.28486	22.05413	25.96102
2019.082	23.87081	22.54372	25.19790	21.84120	25.90042
2019.110	24.77540	23.41891	26.13190	22.70083	26.84998

2019.0	25.21	23.81	26.61	23.07	27.35
137	627	811	443	797	457
2019.0	25.14	23.71	26.56	22.95	27.32
164	201	492	909	947	455
2019.0	24.25	22.80	25.70	22.03	26.46
192	126	188	063	463	788
2019.0	23.35	21.89	24.82	21.11	25.60
219	816	057	575	368	264
2019.0	23.76	22.28	25.25	21.49	26.03
247	614	156	071	568	659
2019.0	23.07	21.57	24.57	20.77	25.36
274	243	256	230	858	628
2019.0	23.87	22.36	25.39	21.56	26.19
301	648	259	037	119	177
2019.0	24.47	22.95	26.00	22.14	26.81
329	936	241	631	410	462
2019.0	23.78	22.24	25.32	21.42	26.13
356	169	229	110	738	601
2019.0	24.28	22.73	25.83	21.91	26.65
384	351	215	488	091	612
2019.0	23.28	21.72	24.84	20.89	25.67
411	484	191	778	454	515
2019.0	23.48	21.91	25.06	21.07	25.89
438	582	162	003	829	336
2019.0	23.68	22.10	25.27	21.26	26.11
466	657	134	180	217	097
2019.0	23.58	21.99	25.18	21.14	26.02
493	714	108	319	618	810
2019.0	23.58	21.98	25.19	21.13	26.04
521	756	084	429	029	484
2019.0	23.48	21.87	25.10	21.01	25.96
548	788	062	514	450	126
2019.0	23.18	21.56	24.81	20.69	25.67
575	812	045	579	881	743

2019.0	23.38	21.75	25.02	20.88	25.89
603	830	032	628	322	338
2019.0	23.28	21.64	24.93	20.76	25.80
630	844	023	664	773	915
2019.0	23.58	21.93	25.24	21.05	26.12
658	854	020	688	233	475

Se observa como los valores pronosticados capturan el comportamiento cíclico de la temperatura promedio en puerto Ayora debido a que la serie efectivamente tiene un comportamiento estacional a lo largo del tiempo.

5. Conclusiones

En este estudio se consideraron los mejores tres modelos de análisis univariante de series de tiempo: modelos ARMA, ARIMA y SARIMA. El mejor ajuste de los tres modelos utilizados en este estudio se seleccionó con base en el modelo que indicaba el menor AIC y BIC. El SARIMA (4,1,1)x(0,1,0) mostró ser el mejor modelo porque se ajustaba mejor a nuestra serie de datos de la temperatura del aire promedio en Puerto Ayora debido a que captura la componente estacional de la serie.

Referencias bibliográficas

- Arturo, C. J., Catya, A. Z., José, L. M., & Diana, S. P. (2016). Análisis de series de tiempo en el pronóstico de la demanda. *Elsevier*, 387-396.
- Hanke, J. E., & Wichern, D. W. (2006). *Pronósticos en los negocios*. Pearson Educación.
- López, D., & Carlos, M. (2013). Modelado de pérdidas en una transmisión. *Revista Tecnura*, 17(37), 53 - 63.
- Siarni-Namini, S., Tavakoli, N., & Namin, A. S. (Diciembre de 2018). A comparison of ARIMA and LSTM in forecasting time series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* , 1394-1401.
- Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research*, 140(3), 684–699.