

Modelo de pronóstico ARIMA en el monitoreo de parámetros ambientales

Model of ARIMA forecast in the monitoring of environmental parameters

Jose Sosa,¹ Eusebio Roque,² Rubén Palomino³

RESUMEN

El presente proyecto de Investigación menciona la solución a la falta de obtener un modelo matemático que ayude a anticiparnos a sus efectos como las heladas por las noches, exceso de radiación solar al medio día característicos de las anomalías climáticas. Como parámetro ambiental, se dispone de los registros de la temperatura ambiental adquiridos de la estación de monitoreo, ubicada en la Escuela de Formación Profesional de Ingeniería Metalúrgica de la Universidad Nacional Daniel Alcides Carrión, Cerro de Pasco. Donde al ser analizados se determina su periodicidad en el tiempo y una relación de comportamiento con su pasado, facilitando su determinación de su modelo matemático determinístico.

Para la obtención del modelo de pronóstico autorregresivo ARIMA, se emplea la metodología de BOX-JENKINS, cuyos procedimientos consiste en un proceso de cinco pasos para identificar, seleccionar y evaluar el modelo de media condicional por datos de series temporales univariantes, característica de los parámetros ambientales. Como antecedente del estudio, se tiene al proyecto OpenStreetMap CLIMATE-DATA.ORG. Cuyo fin es la disponibilidad de datos del monitoreo de climas a nivel mundial aplicando modelos de pronósticos ARIMA para el pronóstico del mismo ver Gráfico N°10.

Palabra clave: Arima, Autorregresivo, Modelo de Pronóstico, Monitoreo, Parámetros Ambientales.

ABSTRACT

The present research project mentions the solution to the lack of obtaining a mathematical model that helps to anticipate its effects such as frost at night, excess solar radiation at midday characteristic of climatic anomalies. As an environmental parameter, the records of the environmental temperature acquired from the monitoring station are available, located in the School of Vocational Training of Metallurgical Engineering of the Daniel Alcides Carrión National University, Cerro de Pasco. Where, when analyzed, their periodicity is determined in time and a behavioral relationship with their past, facilitating their determination of their deterministic mathematical model.

To obtain the ARIMA autoregressive forecasting model, the BOX-JENKINS methodology is used, whose procedures consist of a five-step process to identify, select and evaluate the conditional mean model by univariate time series data, characteristic of the parameters environmental As background to the study, we have the OpenStreetMap project CLIMATE-DATA.ORG. Whose purpose is the availability of global climate monitoring data by applying ARIMA forecast models for the forecast of the same see Graph N ° 10.

Keywords: Arima, Autoregressive, Prognostic Model, Monitoring, Environmental Parameters.

1 Mg. José Luis Sosa Sánchez. Docente de la Universidad Nacional Daniel Alcides Carrión –Cerro de Pasco. Universidad Alas Peruanas

2 Mg. Eusebio Roque Huamán. Docente en la categoría Principal de la Universidad Nacional Daniel Alcides Carrión –Cerro de Pasco.

3 Mg. Rubén Edgar Palomino Isidro. Docente en la categoría Principal de la Universidad Nacional Daniel Alcides Carrión –Cerro de Pasco.



INTRODUCCIÓN

Realizada la exploración por distintos medios de comunicación concerniente al pronóstico en el tiempo de los parámetros ambientales de la ciudad de Cerro de Pasco, solicitado por los medios periodísticos de la ciudad en mención. Se llegó a encontrar información requerida. Conllevando a la pregunta ¿Cómo brindar información anticipada a los interesados del comportamiento de los parámetros ambientales en la ciudad de Cerro de Pasco?

El equipo de investigación se vio en la necesidad de implementar un centro de monitoreo y desarrollar un modelo que pronostique el comportamiento de diversos parámetros entre ellos la temperatura ambiental.

Entonces se ve como una alternativa de aprovechar las ventajas del modelo estocástico autorregresivo integrado de media móvil ARIMA, que tiene como finalidad modelar el comportamiento del proceso en el tiempo y tratarla en forma determinista (Repositorio de la Facultad de Ingeniería - UNAM, 2013, pág. 41) para pronosticar los valores futuros basados en los comportamientos cíclicos de su pasado.

Se determina el modelo ARIMA(p,d,q) empleando la metodología de Box Jenkins (GEP, GM Jenkins y GC Reinsel, 1994), el que nos facilita los cálculos que define el modelo según la ecuación (Jaume Arnau i Gras , 2001, pág. 55):

$$\Phi_p(L)(1-L)^d Y_t = \delta + \theta_q(L) a_t$$

Los modelos con periodicidad o cíclicos en el tiempo como es el caso de los parámetros climáticos, presenta la forma general ARIMA(p,d,q) (P,D,Q)_s y su ecuación es (Jaume Arnau i Gras , 2001, págs. 92-93).

$$\Phi_p(L)\vartheta(L^s)(1-L)^d(1-L^s)^D Y_t = \delta + \theta_q(L)\sigma(L^s)a_t$$

Donde:

Φ_p = *Coefficiente Autorregresivo*

p = *Orden Autorregresivo*

L = *Operador de Retardo*

ϑ = *Coefficiente de Autoregresivo estacionario*

s = *Periodo Estacional*

d = *Orden de las diferencias*

D = *Orden de las diferencias Estacionales*

δ = *Media del proceso*

θ_q = *Coefficiente de media móvil*

q = *Orden de media móvil*

σ = *Coefficiente de media móvil estacional*

a_t = *Función de inovación o perturbaciones*

Después del trabajo de investigación se logró brindar información anticipada a los interesados del comportamiento de los parámetros ambientales en la ciudad de Cerro de Pasco.

Concluyendo que los modelos numéricos de pronóstico para el comportamiento dinámico de parámetros ambientales (Monitoreo y control de los pronósticos, 2015), relacionados a sus límites máximos permisibles, identifica anomalías como es el caso de estudio, que previo a un análisis se puede evitar efectos negativos de un proceso o también se puede mejorar y optimizar, según (Fred Nichols Kerlinger, 1987).

MATERIALES Y MÉTODOS

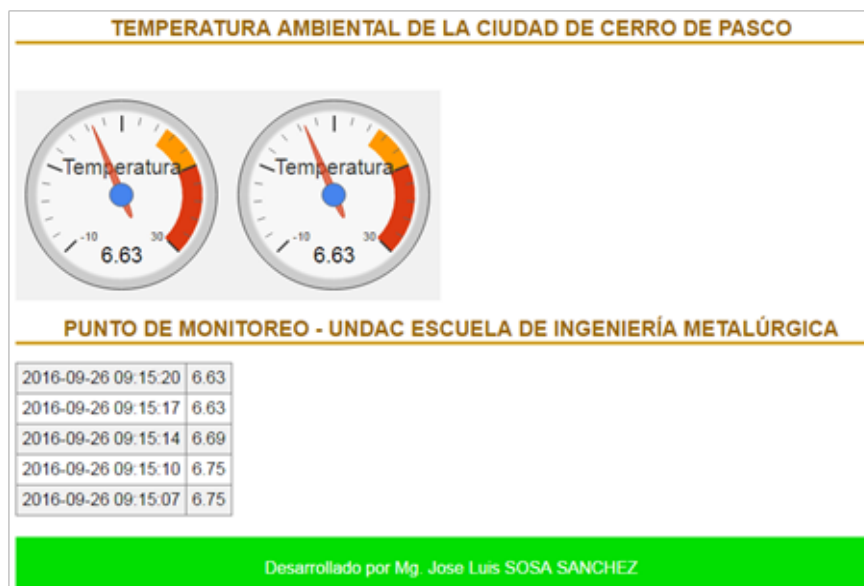
Los materiales empleados son:

1. Software Estadístico: Empleo de las funciones autocorr, parcorr, arima, forecast del lenguaje de programación computacional MATLAB.
2. Sistema de Monitoreo: Interfaz gráfica vía internet disponible en la web <http://www.sosasite>.

com/, monitoreo de la temperatura ambiental de la ciudad de Cerro de Pasco, cuyo dominio dispone de una base de datos en Mysql ver Gráfico N°1.

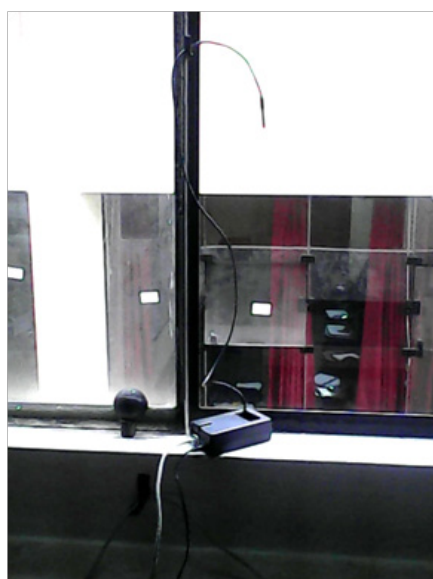
3. Instrumentos Electrónicos: Equipos conformado por una tarjeta de adquisición de datos ARDUINO ETHERNET y un sensor de temperatura modelo DS18B20 ver Gráfico N°2.

Gráfico N°1 Interfaz de monitoreo del parámetro ambiental por medio del sitio web.



Fuente: Elaboracion propia

Gráfico N°2 Equipo de monitoreo de temperatura ambiental de la E.F.P de Ingeniería Metalúrgica UNDAC



Fuente: EFP de Ingenieria Metalúrgica UNDAC

La metodología a seguir se detalla en siete actividades como sigue, con su diagrama de flujo disponible en el Gráfico N°3.

1. Carga de los datos procedentes del dominio www.sosaside.com, con la finalidad de determinar la serie temporal, proveniente de la adquisición de datos del sensor de temperatura DS18B20 ver gráfico N°2.

2. Gráfica de las funciones de autocorrelación, para la identificación del modelo autorregresivo, siguiendo la metodología de Box Jenkins, con el software de tratamiento estadístico MATLAB.

3. Cálculo de la primera diferencial de la serie y sus gráficas de autocorrelación para verificar la estacionariedad de la serie ver gráfico N° 5.

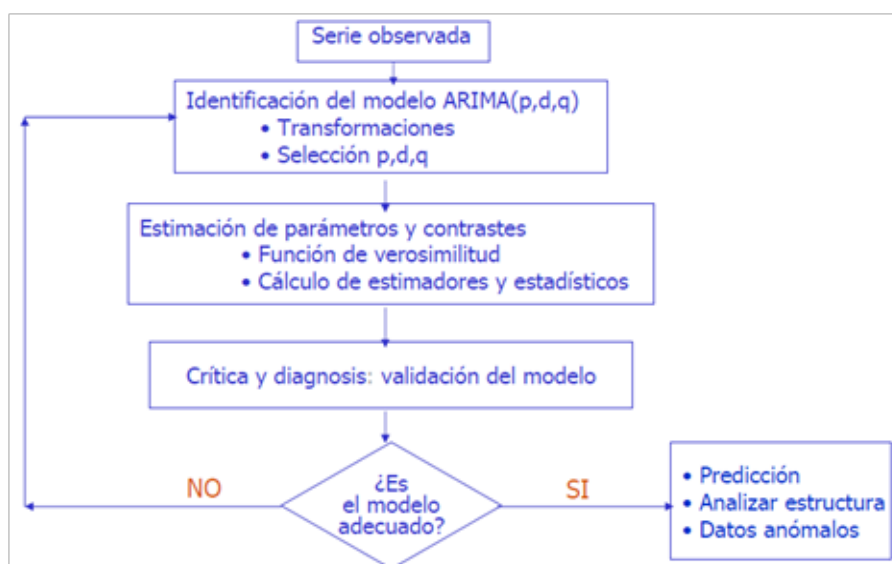
4. Determinación de los coeficientes del modelo autoregresivo para completar el proceso de estimación del modelo ver gráfico N°6

5. Cálculo de la bondad de ajuste del modelo; linealidad, normalidad de los restos de la serie para validar si el modelo cumple la fase de diagnóstico como estipula la metodología Box Jenkins ver gráfico N°7.

6. Simulación del modelo ARIMA con la prueba de sensibilidad de Monte Carlos, para evaluar su desempeño en el pronóstico de datos ver gráfico N° 8.

7. Calculo del pronostico de la serie y su respectiva difusión de la información a los entes interesados ver gráfico N°9.

Gráfico N°3 Diagrama de flujo para la obtención del modelo ARIMA

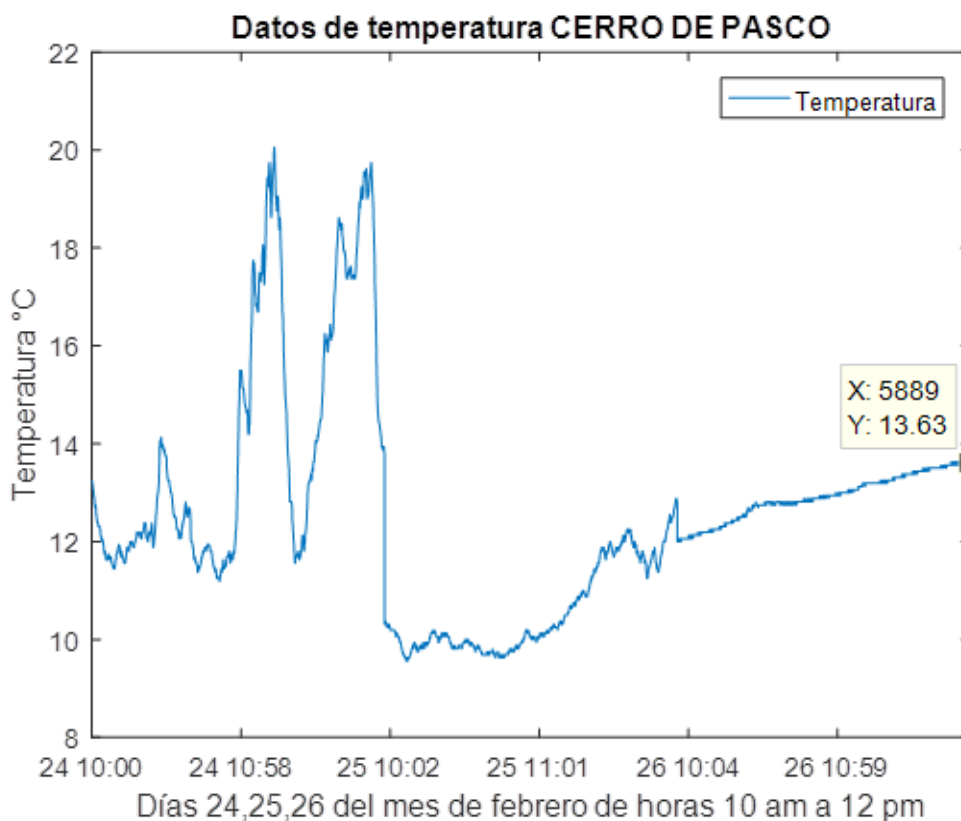


Fuente: Universidad Poletecnica, Maria J. Sanchez Naranjo

RESULTADOS

Planteamos un modelo ARIMA, generado por el registro de un parámetro ambiental en este caso la temperatura. Entonces basándonos en dicho modelo, debemos comprender la dinámica de la serie de tiempo ver gráfico N°4 y efectuar predicciones de valores futuros (Germán Aneiros Pérez, 2015, pág. 3).

Gráfico N°4 Exploración de la base de datos



Fuente: Elaboracion propia

Al proceder a explorar los registros de la temperatura de horas 10:00 a.m. hasta 12:00 p.m. de los días 24, 25 y 26 del mes de febrero del 2016.

Se determina la serie temporal, con una cantidad de 5889 datos, luego obtuvimos el modelo ARIMA, ver gráfico N° 6, cuya ecuación general $ARIMA(5,1,0)(0,0,1)_{s=2}$, (Armando Aguirre Jaime, 1994, pág. 230) se encuentra en función de sus retardos L.

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4 - \phi_5 L^5)(1 - L)Y_t = \delta + (1 + \theta_{s=2} L^2)a_t$$

$$(1 - 0.032L - 0.878L^2 - 0.076L^3 + 0.059L^4 + 0.037L^5) = 1.46 \times 10^{-5} + (1 - 0.733L^2)a_t$$

$$a_t \approx N_{0,0.0037}$$

Donde, a_t es el error aleatorio llamado también ruido blanco, que sigue una distribución normal o gaussiana con media 0 y varianza 0.0037.

Gráfico N° 6 Coeficientes del modelo ARIMA

ARIMA(5,1,2) Model:

 Conditional Probability Distribution: Gaussian

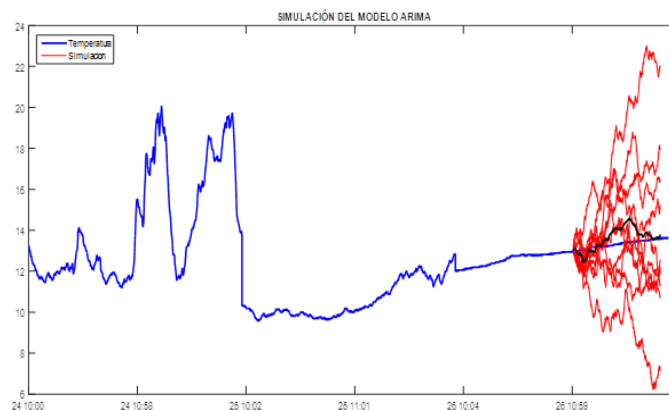
| Parameter | Value | Standard Error | t Statistic |
|-----------|-------------|----------------|-------------|
| Constant | 1.46365e-05 | 0.000328519 | 0.0445532 |
| AR{1} | 0.0326094 | 0.00762636 | 4.27589 |
| AR{2} | 0.878305 | 0.0411511 | 21.3434 |
| AR{3} | 0.076769 | 0.00654231 | 11.7342 |
| AR{4} | -0.0596349 | 0.00985686 | -6.05009 |
| AR{5} | -0.0370219 | 0.00884137 | -4.18735 |
| MA{2} | -0.733744 | 0.041184 | -17.8162 |
| Variance | 0.00373973 | 3.71425e-06 | 1006.86 |

Fuente: Elaboracion propia

La extracción de los últimos 879 datos de la serie valida el modelo, pronosticando la cantidad restante para comparar, si el modelo se ajusta al comportamiento real o no, cuya simulación del Modelo $ARIMA(5,1,0)(0,0,1)$, ver gráfico N° 8, se analiza por el metodo de Monte Carlos con una longitud de 10 partes, ver gráfico N° 9, denominados funciones de predicción (Maria Pilar González Casimiro, Abril, 2009, pág. 99) los mismos visualizados por las líneas de color rojo, como resultados se obtiene a simple vista un pronóstico favorable ya que las líneas de color azul y negras

siguen la misma tendencia, ver gráfico N° 6, así comprobamos que el modelo está en función a su propio pasado y después del análisis nos basamos en la modelación ARIMA que no es encontrar el modelo exacto que represente al proceso generador de las observaciones, es más bien encontrar el modelo que se aproxime al verdadero proceso y que explique el comportamiento de la variable en estudio de forma adecuada y práctica (Repositorio de la Facultad de Ingeniería - UNAM, 2013, pág. 63).

Gráfico N° 8 Verificación de la simulación del modelo con su gráfica histórica



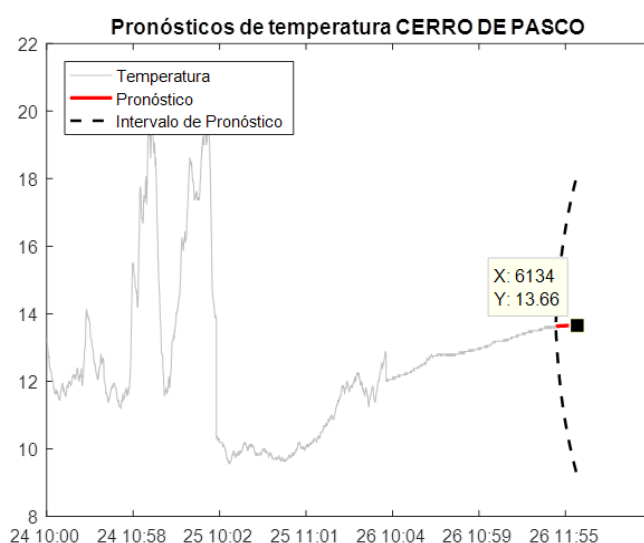
Fuente: Elaboracion propia

Discusión:

Las líneas de rojo son los distintos escenarios generados por el modelo de simulacion de Monte Carlos y se prueba para 10, no se ponen más por que no se puede visualizar las tendencias claramente; Presentan formas alteradas por las anomalías históricas como en el día 24 hora 10:58 a.m. según la gráfica N°8 ; La línea de color negro es el promedio de todos los posibles escenarios y se aprecia su ajuste a los datos reales representado por la línea azul.

Por último, se determina el pronóstico del parámetro de temperatura ver gráfico N° 9.

Gráfico N° 9 Resultados de la simulación del modelo ARIMA



Fuente: Elaboracion propia

Discusión:

Se trabaja con un grado de significancia de 5%, el cual genera el intervalo de pronostico representado por las líneas discontinuas ver cuadro N°1, parte: Determina el pronóstico de la serie; La línea roja presenta una tendencia a subir moderadamente y es la información futura contado del día 26 hora 11:49:35 a.m. en adelante ver gráfico N°11, para lo cual generan 245 datos, entonces el promedio la adquisición de datos es cada 5 segundos en un minuto, por lo tanto 12 lecturas en 245 lecturas equivale a 20 minutos con 25 segundos aproximadamente llegando a obtener el pronóstico hasta la 12:00 p.m. del mismo día.

CONCLUSIONES

El empleo de intervalos de predicción aproximada al 95% de bondad ayuda a la identificación de posibles escenarios del comportamiento de la temperatura consigo en una gestión de prevención, ver gráfico N° 9.

La comprobación de los residuos generado por el modelo, determina la capacidad del pronóstico, por lo que se verifica su normalidad y su estacionalidad en los gráficos de las funciones de autocorrelación simple y parcial, ver gráfico N° 7.

El modelo ETA (EL PRONOSTICO METEOROLOGICO, 2015, pág. 8 Cap 13) empleados por el

SENAMHI (Servicio Nacional de Meteorología e Hidrología del Perú, 2016) requiere de 72 horas para para realizar pronósticos mientras que los modelos ARIMA, se procesa en tiempo real.

RECOMENDACIONES

Emplear el modelo autorregresivo en todos los procesos que tengan la característica de ser cíclica o periódica en el tiempo siempre en cuando se disponga de gran cantidad de datos históricos de la variable en estudio.

Disponer de una base de datos histórica sobre la variable de estudio y que sea de acceso libre para estudios de anomalías por cualquier centro de investigación.

Para el funcionamiento de los modelos ARIMA es necesario tener estaciones de monitoreo con sensores que midan cada parámetro ambiental en tiempo real y que tengan como propósito ayudar a las distintas actividades industriales y académicas con el fin de mejorar los procesos anticipándose al cambio.

BIBLIOGRAFÍA

Armando Aguirre Jaime. (1994). Introducción al tratamiento de series temporales. Madrid: Diaz de Santos S.A.

Bernardi Cabrer Bórras. (2005). Econometría Empresarial II. Valencia: Universidad de Valencia.

Datos climaticos, m. (2016). CLIMATE-DATA.ORG. Obtenido de CLIMATE-DATA.ORG: <http://es.climate-data.org/>

Fernando T. Maestre, Adrián Escudero, Andreu Bonet. (2008). Introducción al Análisis Espacial de Datos en Ecología. Madrid: DYKINSON, SL.

Fred Nichols Kerlinger, H. B. (1987). Foundations of Behavioral Research. USA: South Melbourne, Vic. : Wadsworth/Thomson Learning, ©2000.

GEP, GM Jenkins y GC Reinsel. (1994). El análisis de series temporales: Predicción y control. USA: Prentice Hall.

Germán Aneiros Pérez. (2015). Modelos Box-Jenkins. La coruña: Universidad de Santiago de Compostela.

Gonzalo, Ríos. (2008). Series de Tiempo. Santiago: Universidad de Chile.

Jaume Arnau i Gras . (2001). Diseño de series temporales: Técnicas de Análisis. España: Universidad de Barcelona.

Maria Jesus Sanchez N. (2016). Series Temporales Univer- riantes. España: Universidad Politecnica de Madrid.

Maria Pilar González Casimiro. (Abril, 2009). Analisis de series temporales: Modelos ARIMA. Lejona, España: Universidad del País Vasco.

Mathworks. (2016). Box-Jenkins Model Selection. Obtenido de Box-Jenkins Model Selection: <http://www.mathworks.com/help/econ/box-jenkins-model-selection.html?searchHighlight=BOX%20JENKINS>

Ministerio del Ambiente - MINAM. (Abril, 2014). Protocolo de muestreo por emergencia ambiental. Lima - Peru.

Monitoreo y control de los pronósticos. (12 de 5 de 2015). SlideShare. Obtenido de <http://es.slideshare.net/smantharck/karla-vea-unidad-2-expo-copia>

R. Dario Sbarato, Viviana Sbarato, Jose E. Ortega. (2007). Predicción y Evaluación de impactos ambientales sobre la atmósfera . Brujas.

Repositorio de la facultad de Ingeniería - UNAM. (2013). Modelación ARIMA. Mexico: Universidad Autonoma de Mexico.

SENAMHI. (2015). EL PRONOSTICO METEOROLOGICO. LIMA.

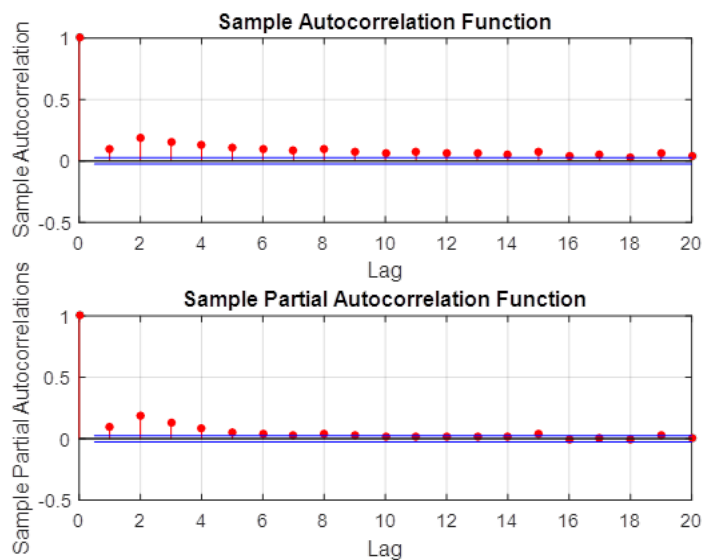
SENAMHI. (2016). Servicio Nacional de Meteorología e Hidrología del Perú. Obtenido de EL PRONOSTICO METEOROLOGICO: www.senamhi.gob.pe/main_down.php?ub=mmt&id=cap13

SOSA SANCHEZ, J. L. (2016). Influencia de los parametros climaticos, soportado en la TIC'S determina un microclima apropiado del invernadero. Ciencia y Desarrollo, 25-37.

SOSA SANCHEZ, J. L. (12 de Febrero de 2016). Monitoreo de Parametros climaticos - Temperatura. Obtenido de Monitoreo de Parametros climaticos - Temperatura: <http://www.sosasite.com/>

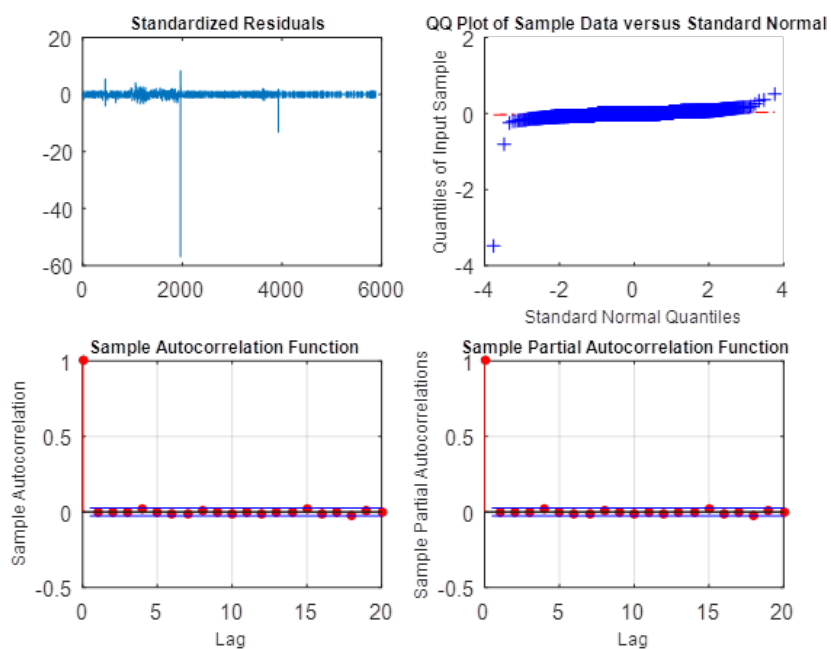
ANEXOS

Gráfico N° 5 Funcion de autocorrelación simple y parcial



Fuente: Elaboracion propia

Gráfico N° 7 Análisis de sensibilidad de los restos de la serie



Fuente: Elaboracion propia

Cuadro N°1 Código del proyecto en el lenguaje de programación Matlab

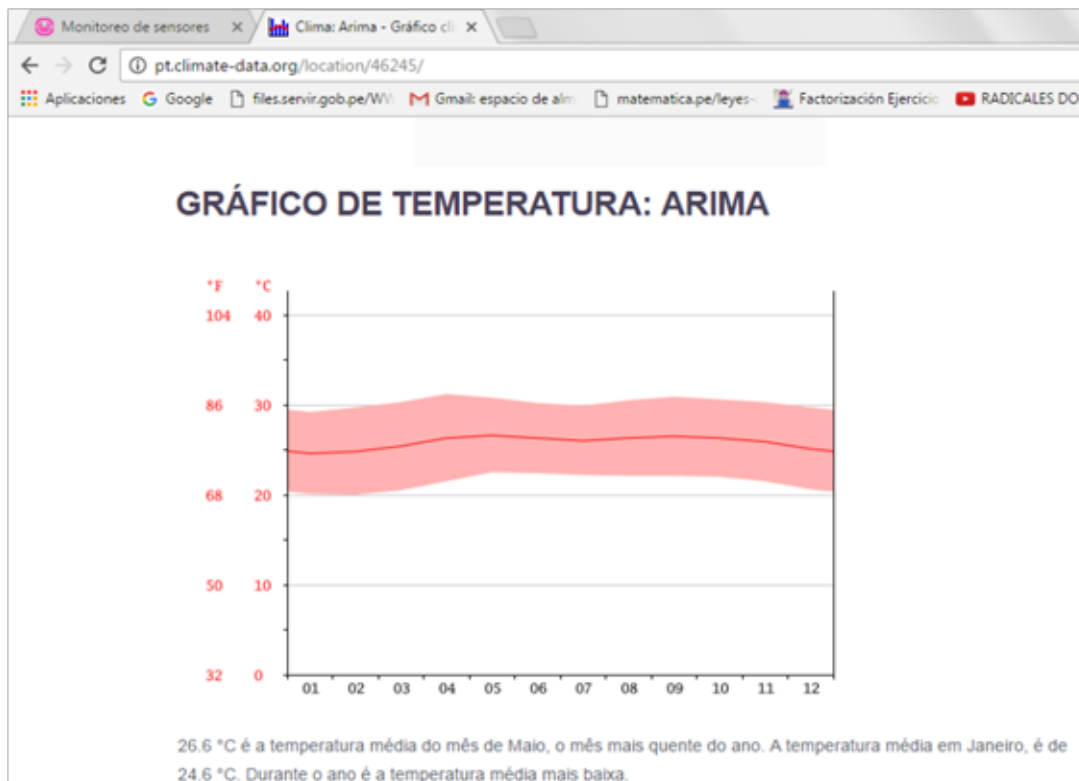
```

%% CARGA DE LOS 5889 DATOS DE TEMPERATURA
load temperatura.mat
y=datostemperatura.T;
dates=datetime(datostemperatura.FECHA); %convierte a numerico
T = length(y);
figure
plot(y)
h1 = gca;
h1.XLim = [0,T];
h1.XTick = 1:1000:T;
h1.XTickLabel = datestr(dates(1:1000:T), 'dd HH:MM');
%dt = datestr(now, 'mmm dd, yyyy HH:MM:SS.FFF AM')
title('Datos de temperatura CERRO DE PASCO')
%% DETERMINA LAS FUNCIONES DE AUTOCORRELACION
figure
subplot(2,1,1), autocorr(y)
subplot(2,1,2), parcorr(y)
%% DETERMINA LA PRIMERA DIFERENCIAL DE LA SERIE
dY = diff(y);
figure
plot(dY)
h2 = gca;
h2.XLim = [0,T];
h2.XTick = 1:1000:T;
h2.XTickLabel = datestr(dates(2:1000:T), 'dd HH:MM');
title('Datos diferenciados de temperatura CERRO DE PASCO')
%% DETERMINA LAS FUNCIONES DE AUTOCORRELACION DE LOS DATOS DIFERENCIADOS
figure
subplot(2,1,1), autocorr(dY)
subplot(2,1,2), parcorr(dY)
%% DETERMINA LOS COEFICIENTES DEL MODELO
Mdl = arima('ARLags', [1 2 3 4 5], 'MALags', 2, 'D', 1);
EstMdl = estimate(Mdl, y);
%% DETERMINA LAS BONDADES DE AJUSTE DEL MODELO
res = infer(EstMdl, y);
figure
subplot(2,2,1)
plot(res./sqrt(EstMdl.Variance))
title('Standardized Residuals')
subplot(2,2,2), qqplot(res)
subplot(2,2,3), autocorr(res)
subplot(2,2,4), parcorr(res)
hvec = findall(gcf, 'Type', 'axes');
set(hvec, 'TitleFontSizeMultiplier', 0.8, ...
    'LabelFontSizeMultiplier', 0.8);
%% DETERMINA EL PRONOSTICO DE LA SERIE
[yF, yMSE] = forecast(EstMdl, 245, 'Y0', y);
UB = yF + 1.96*sqrt(yMSE);
LB = yF - 1.96*sqrt(yMSE);
figure
h4 = plot(y, 'Color', [.75, .75, .75]);
hold on
h5 = plot(5890:6134, yF, 'r', 'LineWidth', 2);
h6 = plot(5890:6134, UB, 'k--', 'LineWidth', 1.5);
plot(5890:6134, LB, 'k--', 'LineWidth', 1.5);
fDates = [dates; dates(T) + cumsum(diff(dates(T-245:T)))]';
h7 = gca;
h7.XTick = 1:1000:(T+245);
h7.XTickLabel = datestr(fDates(1:1000:end), 'dd HH:MM');
legend([h4, h5, h6], 'Temperatura', 'Pronostico', ...
    'Intervalo de Pronostico', 'Location', 'Northwest')
title('Pronosticos de temperatura CERRO DE PASCO')
hold off

```

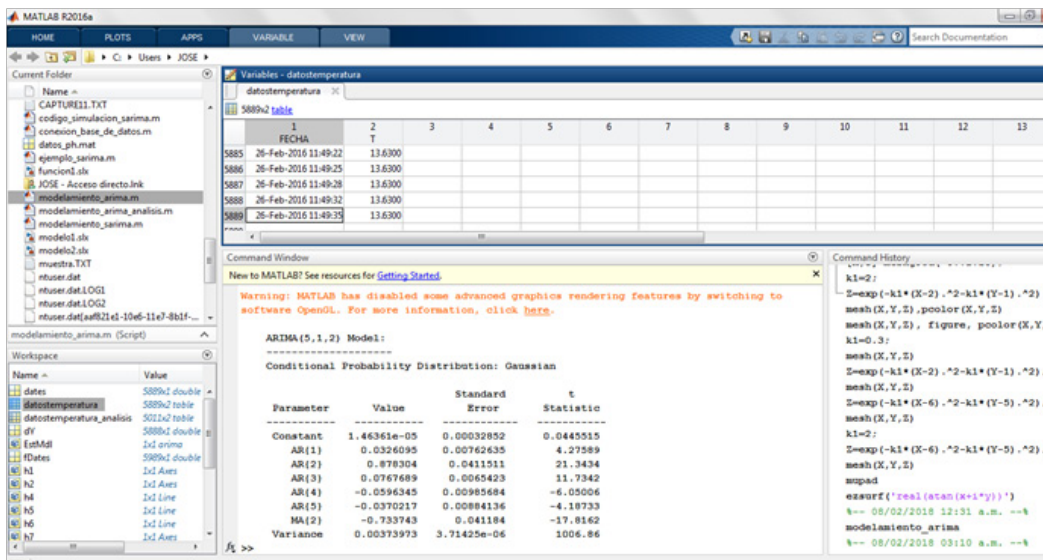
Fuente: Elaboracion propia

Gráfico N° 10 Modelo Arima para el pronóstico del parametro de temperatura



Fuente: <http://pt.climate-data.org/location/46245/>

Gráfico N°11 Dato 5889 corresponde al dia 26 de febrero de 2016 hora 11:49:35



Fuente: Elaboración propia

